AP STATISTICS

2022-23 AP EXAM STUDY GUIDE

by David Hornbeck

Table of Contents

AP STATISTICS EXAM FORMAT

SECTION 1: 40 Multiple Choice, 90 Minutes
Worth 50 points – take # of correct answers and multiply by 1.25.

General scores to aim for:
- If you want a 3: 25 or more correct
- If you want a 4: 28 or more correct
- If you want a 5: 32 or more correct

SECTION 2: 6 Free Response Questions, 90 Minutes
Worth 50 points.

The first 5 FRQs are scored out of 4 points each, and your score on each one gets multiplied by 1.875.

The last FRQ - #6 – is worth 4 points, but your score gets multiplied by 3.125. It is usually significantly <u>longer</u> and *may* have a couple of very difficult portions.

STRATEGY AND HOW NOT TO USE YOUR STUDY TIME

FRQ Strategy
#1 or #2 is usually the easiest question. Find it and knock it out.

Go check out #6. Do the parts you understand and try to get at least a point.

Go back and do the rest of #1-5 in order of what you find easiest.

If you have remaining time, spend it either finishing #6 or polishing up your #1-5.

SCORES TO AIM FOR
3: 40-44
4: 56-60
5: 70-72

HOW *NOT* TO USE YOUR STUDY TIME

I can't tell you how or what to study, but I can recommend how *not* to study.

Are you just trying to pass the exam? **Don't waste your time working on the #6s.** Focus on the fundamentals.

Do you feel weak in a particular unit? **Don't spend <u>all</u> of your time on it – mix it up.**

Do you feel better at MC than FRQs, or vice versa? **Don't spend <u>all</u> of your time on either one – mix it up.**

The AP exam can cover *anything*, so you don't want to put all of your eggs in any one basket. Keep yourself fresh on everything you can.

AP STATISTICS CONCEPTS/FACTS/DEFINITIONS/SKILLS

The units are in the order I taught them in.

UNIT 1: DESIGNING STUDIES

- Carrying out a simulation
- Interpreting results of a simulation (simulated P-value)
- Convenience sample
- Creating a dotplot
- Simple random sample
- How to carry out a simple random sample (strips from hat, using RNG, etc.)
  - randInt
- Definitions: individuals (subjects), variable, categorical, quantitative, distribution of a variable, population (of interest, of inference), representative, bias (over- and underestimates), explanatory variable, response variable
- Observational study
- Scope of inference (causation, correlation, population of inference)
- Stratified sampling (purpose and statistical advantages)
  - How to carry out
- Cluster sampling (purpose and advantage)
  - How to carry out
- Systematic random sample
- Types of bias
  - Undercoverage (selection bias)
  - Nonresponse
  - Response
- Experiments
  - Difference between experiments and observational studies
    - Methodology, inferences
  - How to design & describe
- Confounding
  - What it is, how to explain it well
- Purpose/necessity (or lack thereof) of control groups
- Placebo/placebo effect
- Types of experiments
  - Completely randomized
  - Blocking randomized
    - Matched pairs

UNIT 2: EXPLORING DATA

- Frequency v. relative frequency (w/ tables)
- Bar graphs
  - How to make
  - How to read (NOT SCVU)/determining association or not
- Finding flaws in graphs
- Two-way tables
  - Marginal distributions, using relative frequency over frequency (when?)
  - Conditional distributions
- Graphs of two or more categorical variables:
  - Comparative bar graphs, segmented bar graphs, mosaic plots

- Graphs of quantitative data: creating them, describing them (SCVU), analyzing them
    - Dotplots – show individual values: can calculate mean, find exact median
    - Stemplots – show individual values: can calculate mean, find exact median, LOOK FOR GAPS
    - Histograms – does not show individual values: cannot calculate mean, can only estimate median
        - equivalent to a *condensed* dotplot
    - Boxplots – five-number summary, IQR; cannot only determine rough symmetry or skew, NOT number of peaks (can't say approximately normal, etc.)
        - CANNOT determine number of observations from a boxplot (if one is more spread out than another, it doesn't mean it has more data points)
- SCVU
    - Shape: Skewed, roughly symmetric, approximately normal, bimodal, uniform
    - Center: mean or median; knowing which to choose for each type of graph
    - Variability: range, standard deviation, IQR, concentration of data (most/least); knowing which to choose for each type of graph
    - Unusual features:
        - Outlier: Describe if there
            - $Q3 + 1.5IQR, Q1 - 1.5IQR$
        - Gaps, clusters, etc.
- Comparing distributions: *compare* shape, then center, then spread
- Standard deviation; formula, interpretation
- Calculator skill: STAT -> CALC -> 1: 1VarStats

<u>UNIT 3: MODELING DISTRIBUTIONS OF DATA</u>
- Calculating & interpreting percentiles
- Cumulative frequency distributions: creating, finding median & percentiles, determining skew
- z-scores
    - Computing them ($z = \frac{x-\mu}{\sigma}$)
    - Interpreting them (number of standard deviations above or below mean)
    - No units; unaffected by transformations of data
- Normal curves
    - Standard normal curve $N(0, 1)$ where each value is a z-score
    - Empirical rule: 68-95-99.7 (use for 34, 16, 2.5, etc.)
    - Finding areas under curve: normalcdf
    - Finding percentiles: invNorm
    - STANDARDIZING

<u>UNIT 4: PROBABILITY</u>
- NOT INTUITIVE
- ALWAYS SHOW WORK
- Law of large numbers (probability as long-run proportion)
- Performing a simulation
    - Simulated P-values
- Using random digit table
- Notation: $\cup, \cap, A^C$ (A')
- Mutually exclusive events: $P(A \cap B) = 0, P(A \cup B) = P(A) + P(B)$
- General addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Two-way tables
- Conditional probability formula: $P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B|A) = \frac{P(A \cap B)}{P(A)}$
- Tree diagrams

- General multiplication rule: $P(A \cap B) = P(A) * P(B|A) = P(B) * P(A|B)$
- Independent events: definition & formulas
    - $P(A) = P(A|B)$
    - $P(B) = P(B|A)$
    - $P(A \cap B) = P(A) * P(B)$
- Independent events = events with no association
- False positives & false negatives
- $P(\text{at least one}) = 1 - P(\text{none})$

UNIT 5: RANDOM VARIABLES

- Calculating mean (expected value) of probability distribution
- Calculating standard deviation of probability distribution
- Interpreting mean and standard deviation of probability distribution
- Using normalcdf & invNorm
- Calculating probabilities based on probability distribution
- Binomial random variables & their distribution
    - Binary: every choice can be categorized as success or failure
    - Independent: trials are independent of one another
    - Number of successes is fixed
    - Success: probability of success on each trial is the same
- Binomial distribution of a variable X with n trials and probability of success p
    - Mean: $\mu_X = np$
    - Standard deviation: $\sigma_X = \sqrt{np(1-p)}$
    - Shape: approximately Normal if $np \geq 10$ and $n(1-p) \geq 10$
    - interpreting mean and standard deviation
- Geometric distribution of a variable X with probability of success p
    - Mean: $\mu_X = \frac{1}{p}$
    - Standard deviation: $\sigma_X = \frac{\sqrt{1-p}}{p}$
    - interpreting mean & standard deviation
- Combining random variables
    - $\mu_{X \pm Y} = \mu_X \pm \mu_Y$
    - $\sigma_{X \pm Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$

UNIT 6: SAMPLING DISTRIBUTIONS

- The three distributions:
    - Population distribution: distribution of values for all individuals
    - Distribution of a sample: distribution of values for individuals in the sample
    - Sampling distribution: distribution of a statistic for all possible samples of the same size
- Sampling variability
- Definitions: parameter, statistic, random process, expected value (= mean)
- Increasing sample size makes statistic more *precise*
- Random sampling or assignment ensures statistic is *unbiased* (*accurate*)
- Sampling distribution of $\hat{p}$
    - Population parameter: p
    - $\hat{p}$ = sample proportion
    - Center: Mean of all possible $\hat{p}$'s: $\mu_{\hat{p}} = p$ (unbiased)
    - Spread: Standard deviation of all possible $\hat{p}$'s: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ so long as $n \leq \frac{1}{10}N$

- Shape: Approximately Normal if $np \geq 10, n(1-p) \geq 10$
- Sampling distribution of $\hat{p}_1 - \hat{p}_2$
  - Population parameters: p_1 and p_2 ($p_1 - p_2$)
  - Sample proportions: $\hat{p}_1$ and $\hat{p}_2$ ($\hat{p}_1 - \hat{p}_2$)
  - Center: $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ (unbiased)
  - Spread: So long as $n_1 \leq \frac{1}{10}N_1$ and $n_2 \leq \frac{1}{10}N_2 \ldots\ \sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
  - Shape: Approximately Normal if $n_1p_1 \geq 10, n_1(1-p_1) \geq 10, n_2p_2 \geq 10, n_2(1-p_2) \geq 10$
- Sampling distribution of $\bar{x}$
  - Population parameter: μ
  - $\bar{x} =$ sample mean
  - Center: Mean of all possible $\bar{x}$'s: $\mu_{\bar{x}} = \mu$ (unbiased)
  - Spread: Standard deviation of all possible $\bar{x}$'s: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, SE_{\bar{x}} = \frac{s_x}{\sqrt{n}}$ so long as $n \leq \frac{1}{10}N$ (trials reasonably independent)
  - Shape: Approximately Normal if:
    - Population distribution is approximately Normal, or
    - Sample size is at least 30 (Central Limit Theorem), or
    - Sample data shows no strong skew or outliers
- Sampling distribution of $\bar{x}_1 - \bar{x}_2$
  - Population parameters: μ_1 and μ_2 ($\mu_1 - \mu_2$)
  - Sample means: $\bar{x}_1$ and $\bar{x}_2$ ($\bar{x}_1 - \bar{x}_2$)
  - Center: $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$ (unbiased)
  - Spread: So long as $n_1 \leq \frac{1}{10}N_1$ and $n_2 \leq \frac{1}{10}N_2$, $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
    - Because we don't know σ_1 or σ_2, we use $SE_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
  - Shape: Approximately Normal if:
    - BOTH populations are Normal, or
    - BOTH sample sizes are greater than 30, or
    - BOTH sample distributions are reasonably okay
      - *Rarely is it a *combination* of two of these three.
- EFFECT OF INCREASING SAMPLE SIZE: REDUCES STANDARD DEVIATION BY SQUARE ROOT OF FACTOR OF INCREASE (Sample size times 100 ⇒ standard deviation divided by 10)

UNIT 7: ESTIMATING WITH CONFIDENCE (PROPORTIONS)

- If a distribution is approximately Normal, then the probability that a sample proportion or mean will be within 2 standard deviations of the mean of the sampling distribution is 95%; therefore, if we take our sample proportion or mean and add/subtract 2 standard deviations, there is a 95% chance that we will capture the true mean of the sampling distribution!
  - % changes depending on how many standard deviations you add/subtract to/from the sample proportion or mean
- Interpretation of confidence level
  - percent of all possible intervals that would capture true population parameter
  - percent probability that *new* interval (not yet calculated) would capture the true population parameter
  - percent probability that sample proportion or mean will be within margin of error of true population parameter
- Interpretation of confidence interval
- Interpretation of margin of error

- Confidence interval -> interval of all plausible values
- EFFECT OF INCREASING SAMPLE SIZE ON WIDTH OF CONFIDENCE INTERVALS
- Formula: statistic $\pm$ (critical value) $*$ (standard deviation of statistic)
    - For sample proportion: $\hat{p} \pm z^* \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$
    - For difference of proportions: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
    - HOW TO DO BOTH IN CALCULATOR
- Calculating z^*
    - MEMORIZE
        - 90% -> $z^* = 1.645$
        - 95% -> $z^* = 1.96$
        - 99% -> $z^* = 2.576$
    - Otherwise: $z^* = \text{invNorm}\left(0.C + \frac{1}{2}(1 - 0.C)\right)$
- CALCULATING MINIMUM SAMPLE SIZE GIVEN MAXIMUM MARGIN OF ERROR
    - Formulas:
        - Proportion: $ME = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$… use $\hat{p} = 0.5$ if no $\hat{p}$ available (this will lead to most conservative n)
- Conditions
    - Random, 10% for both proportions, Large Counts
        - Large Counts: $(n\hat{p} \geq 10, n(1-\hat{p}) \geq 10$… ALLOWED TO (MUST) USE $\hat{p}$
        - If two proportions, must check INDEPENDENT random samples

UNIT 8: ESTIMATING WITH CONFIDENCE (MEANS)
- Calculating t^* with $\text{df} = n - 1$
    - t^* always greater than corresponding z^*
    - $\text{invT}\left(0.C + \frac{1}{2}(1 - 0.C), n - 1\right)$
- Necessity of t distribution:
    - If we don't know σ, then we use s_x. Because $s_x < \sigma$, the standard deviation of $\bar{x}$ will be too small ($\frac{s_x}{\sqrt{n}} < \frac{\sigma}{\sqrt{n}}$). To compensate we use a *larger* critical value, called t^*.
- Characteristics of t distribution:
    - More density in the tails, less density near the mean
    - Approaches a Normal distribution as n increases
- Formulas
    - For sample mean: $\bar{x} \pm t^* \left(\frac{s_x}{\sqrt{n}}\right)$
    - For difference of means: $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
    - Paired: $\bar{x}_d \pm t^* \left(\frac{s_{x_d}}{\sqrt{n}}\right)$
- CALCULATING MINIMUM SAMPLE SIZE GIVEN MAXIMUM MARGIN OF ERROR
    - Mean: can only be done if we have an idea what σ is, SO WE USE z^* INSTEAD OF t^*…

$$ME = z^* \left(\frac{\sigma}{\sqrt{n}}\right)$$

UNIT 9: TESTING A CLAIM (PROPORTIONS)
- Null hypothesis H_0: what we assume to be true
- Alternative hypothesis H_a: what we would like to show or that we suspect

- Once you have a sample statistic, you find how many standard deviations above or below the mean of the sampling distribution of the statistic is… this is called the TEST STATISTIC.
- Then, you calculate the probability that the probability that a test statistic this extreme or more (depends on the direction of the alternative hypothesis) would occur *due to chance*. This resulting probability is called the P-VALUE.
- Interpreting P-value
- Significance level α
- Rejecting H_0 or *failing* to reject H_0 (NEVER ACCEPT H_0)
- Error types
    - Type I error: reject H_0 when it is not actually true; being convinced when you shouldn't be
        - Probability of Type I error is exactly the probability of rejecting the null: α!
    - Type II error: *failing to reject* H_0 when it is actually true; *not* being convinced that H_0 is *not* true when you should be
- STATE: Identify parameter p or μ, state hypotheses and significance level
- PLAN: Name test (one-prop z-test or one-sample t-test), check conditions
- DO: Report test statistic and P-value (and df if t-test)
- CONCLUDE: Link P to α, explicitly reject or fail to reject, say that you have or do not have convincing evidence for the alternative hypothesis *in context*
- Two-sided tests and confidence intervals
    - A two-sided test with α will have exact same result (reject/FTR, not plausible/plausible) for a confidence interval with $C = 1 - \alpha$ (%, of course)
- Conditions
    - EXACT SAME *except…*
        - MUST USE p_0 INSTEAD OF $\hat{p}$ FOR STANDARD DEVIATION
- Power
    - Probability of rejecting when you should (given a specific alternative)
    - Power $= 1 -$ probability of a Type II error
    - How to increase power:
        - Increase sample size -> more information (more sensitive test)
        - Increase "effect size" -> the further away the specific alternative is from the null value, the easier it will be to reject the null
        - Increase α -> makes it easier to reject the null
- 2-proportion z-tests
    - MULTIPLE CHOICE: Must use POOLED proportion because you are assuming $p_1 = p_2$. Use $p_c = \frac{x_1+x_2}{n_1+n_2}$. *THIS APPLIES TO LARGE COUNTS, TOO.*
        - Test statistic: $z = \frac{(\hat{p}_1-\hat{p}_2)-0}{\sqrt{\frac{p_c(1-p_c)}{n_1}+\frac{p_c(1-p_c)}{n_2}}} = \frac{(\hat{p}_1-\hat{p}_2)-0}{\sqrt{p_c(1-p_c)(\frac{1}{n_1}+\frac{1}{n_2})}}$

UNIT 10: TESTING A CLAIM (MEANS)

- Same information as for any hypothesis test (error types, power, etc.)
- Conditions:
    - One-sample & paired: Random, 10%, Normality
    - Two-sample: Independent random, 10% for both groups, Normality of both groups
- 1-sample t-tests
    - MULTIPLE CHOICE:
        - Test statistic: $t = \frac{\bar{x}-\mu_0}{\frac{s_x}{\sqrt{n}}}$
- 2-sample t-tests
    - MULTIPLE CHOICE:

        - Test statistic: $t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Paired t-test
    - If you measure the same thing twice for the same individual *or* measure something for two very *similar* individuals, then subtract, you are finding a *mean* of *differences*; this is a *single* mean, despite it looking like two samples. It runs just like a one-sample t-test, only you use each *difference*.
        - Test statistic: $t = \frac{\bar{x}_d - 0}{\frac{s_{x_d}}{\sqrt{n}}}$

UNIT 11: DESCRIBING RELATIONSHIPS

- Describing scatterplots
    - Direction: positive or negative
    - Form: linear or nonlinear
    - Strength: strong, weak, moderately strong, moderately weak
    - Outliers: describe if exist
- Interpreting r: strength and direction, NOT form
- Interpreting r^2: percent of variability in y-variable that can be accounted by for LSRL of y on x
- Definition of LSRL: line that minimizes squared residuals
- Finding equation of LSRL from computer readout $\hat{y} = a + bx$
- Interpret y-intercept (a) of LSRL: the predicted value of y when $x = 0$
- Interpret slope (b) of LSRL: *predicted* increase or decrease in y unit/s for each *additional* one unit of x
- Interpreting s: if we make many, many predictions using this LSRL, we will typically be off by about s
- Residual = Actual y − Predicted $y = y - \hat{y}$
- Interpreting residuals
    - Positive residual = underestimate
    - Negative residual = overestimate
- Residual plots
    - No form ⇒ linear model appropriate
    - Curved form or clear pattern ⇒ linear model not appropriate
- Using residual plot & predicted value to work backwards to compute actual y
- Slope of regression line: $b = r\left(\frac{s_y}{s_x}\right)$
- Transformations of data
    - If $\log \hat{y}$ v. $\log x$ (or $\ln \hat{y}$ v. $\ln x$) is linear, then a POWER model is appropriate.
    - If $\log \hat{y}$ v. x (or $\ln \hat{y}$ v. x) is linear, then an EXPONENTIAL model is appropriate.
- Computing predicted values from transformed data sets
    - If $\log \hat{y} = k$ (whatever k is), remember to raise 10 to the k.
    - If $\ln \hat{y} = k$ (whatever k is), remember to raise e to the k.

UNIT 12: CHI-SQUARED AND T-TESTS FOR SLOPE

- For distributions of categorical data (like color, gender, political preference, age groups, etc.), there are three primary questions we can ask:
    - 1. Is a proposed distribution a good fit (correct)?
    - 2. Is the distribution of a variable the *same* for different values (often via treatments) of a second categorical variable?
    - 3. Are two variables *associated* or not? (Are they *not independent* or *independent*?)
- These correspond to the three kinds of chi-square tests:
    - 1. Chi-square test of goodness of fit
    - 2. Chi-square test of homogeneity

- 3. Chi-square test of association/independence
- 1. Chi-square test of goodness of fit
  - H_0: Proposed distribution of categorical variable is *correct*
  - H_a: Proposed distribution of categorical variable is NOT *correct*
  - Record the *observed* counts
  - Calculate the *expected* counts np_1, np_2, etc.
  - Calculate $\chi^2 = \Sigma \frac{(\text{observed}-\text{expected})^2}{\text{expected}}$
  - df = (number of categories) − 1
  - MULTIPLE CHOICE: Use χ^2cdf(χ^2, 10000, df)
  - FREE RESPONSE: Put data in L1, use D:χ^2GOF-Test
    - Test will *fill in* expected counts
    - RECORD EXPECTED COUNTS WITH SOME WORK SHOWN
- Facts about χ^2 distribution
  - Minimum = 0
  - Always *right* skewed
  - Mean = df
  - Mode (peak) = df − 2
  - χ^2 is a sum of CONTRIBUTIONS; biggest contributions mean biggest relative deviations from proposed distribution
- 2. Chi-square test of homogeneity
  - Two-way table
  - INDIVIDUALS WILL BE DIVIDED INTO SAMPLES OR TREATMENT GROUPS (by researchers or experimenters)
    - *Homogeneity* implies *sameness* which implies *multiple* groups, not just one overall sample
  - H_0: The distribution of *response* categorical variable is the *same* for each value of the *explanatory* categorical variable.
  - H_a: The distribution of *response* categorical variable is NOT the same for each value of the *explanatory* categorical variable.
  - Put observed counts in Matrix [A]
  - Expected counts:
    - MULTIPLE CHOICE: expected $= \frac{(\text{row total})*(\text{column total})}{\text{table total}}$
    - FREE RESPONSE:
      - Calculator: C:χ^2-Test, observed = [A], expected = [B]. Run the test, *then the expected counts will be in* [B].
      - Go back and show work for at least a couple of the expected counts.
  - df = (number of row categories – 1)(number of column categories – 1)
- 3. Chi-square test of independence/association
  - Version 1:
    - H_0: The two categorical variables have no association for the population of ________.
    - H_a: The two categorical variables have AN association for the population of ________.
  - Version 2:
    - H_0: The two categorical variables are independent for the population of ________.
    - H_a: The two categorical variables are NOT independent for the population of ________.
  - RUNS EXACTLY LIKE A TEST FOR HOMOGENEITY
  - df = (number of row categories – 1)(number of column categories – 1)
- Population regression line: $\hat{y} = \alpha + \beta x$
- Sampling distribution of sample slope b
  - Mean: $\mu_b = \beta$ (unbiased)

- Standard deviation: $\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$
  - Because we don't know σ or σ_x, we use $SE_b = \frac{s}{s_x \sqrt{n-1}}$
  - This means we must use the t distribution for critical values

- Interpret SE_b: if we took many, many samples and computed LSRLs, our sample slope would typically be within SE_b of the slope of the population line (true slope)
- Confidence interval for slope:
  - $b \pm t^* SE_b$, where t is computed from C% confidence with df $= n - 2$
- Find test statistic t and P-value from computer readout
  - Test statistic *by hand*: $t = \frac{b-\beta_0}{SE_b}$
    - *Usually*, $\beta_0 = 0$, so this often turns into $t = \frac{b-0}{SE_b} = \frac{b}{SE_b}$
- t-test for slope
  - Conditions:
    - LINEAR: The actual relationship between x and y is linear. For any fixed value of x, the mean response μ_x falls on the population (true) regression line $\mu_y = \alpha + \beta x$.
    - INDEPENDENT: Individual observations are independent of each other. When sampling without replacement, check the *10% condition*.
    - NORMAL: For any one value of x, the response *y* varies according to a Normal distribution.
    - EQUAL SD: The standard deviation of y (call it σ) is the same for all values of x.
    - RANDOM: The data come from a well-designed random sample or randomized experiment.

Lesson 1: Units 1-3

Definitions:
In order to talk about a <u>population</u> – the overall group – and some <u>parameter</u> (number) that describes this group, we typically must take a <u>sample</u> – a subset of the population – and compute a <u>statistic</u> (number that describes the *sample*).

What can go wrong? The sample might not be <u>representative</u> of the population. How can this occur?

Name of bias and general description	… and how it might make the sample non-representative
1. <u>Nonresponse bias</u>: Such a small percentage of individuals respond that those who do respond	likely differ from those who didn't respond (might feel more strongly, for instance) in a way related to what is being measured. .
2. <u>Voluntary response bias</u>: The sample consists of people who *chose* to respond,	and these individuals may *differ* from those who didn't respond in a way related to what is being measured.
3. <u>Selection bias (undercoverage)</u>: Some subgroup of the population *can't* be selected	so, if this subgroup is *different* than those who *can* be selected *with regards to what is being measured*, the sample may be non-representative.
4. <u>Response bias</u>: Individuals may answer questions untruthfully due to the nature of the question or interviewer and	the sample of responses may be *different* than it would've been if the nature of the question or interviewer were changed.

How do we prevent the first three kinds of bias? We sample <u>randomly</u>.

A random sample of individuals from a population should produce a sample that is representative of the population, thus allowing for *inferences* to that population.

What kinds of random samples are there?

Name	Outline	Example	Purpose
1. Simple random (SRS)	Number everyone, use hat or RNG to select n individuals	Number all students 1-300. Use RNG to get 10 unique integers from 1-300. Select these students.	Simplicity; each random sample equally likely to be selected
2. Stratified random	Create groups different WRT what is being measured, then do SRS *within* each group	(If measuring love of math) Get some students from each math period; these students have different teachers, so they may feel *differently* about math.	Reduce variability in estimates by guaranteeing selection of individuals from all different subgroups
3. Cluster random	Create *similar* groups (mini-populations) and take an SRS *of the groups*	(If measuring whether students prefer Zaxby's or Chick-Fil-A) Number the math classes, then randomly select 3. Get *everyone* from these classes to answer the question.	Save time (maybe money)
4. Systematic	Pick a random starting point, then get every nth individual	(If measuring whether students prefer Zaxby's or Chick-Fil-A) Starting on a randomly selected day, select every 10th student walking in the door and ask them the question.	Doable when population or groups can't easily be numbered

<u>AP TIPS</u>

1. If you invoke bias, you must provide a <u>direction</u>: how the bias will lead to an <u>underestimate</u> or <u>overestimate</u> of the <u>parameter</u>.

2. Don't throw around the word variability if you don't understand what it means. If you ever say "variability," ask yourself if you've answered the question: *in what?*

3. For bias to occur, the sample must be <u>different</u> than those not sampled in some way. This difference must be related to <u>what is being measured</u>.

PRACTICE: The manager of the department at a large firm has 15 full-time data clerks that work for her. Each is responsible for inputting 30 detailed medical records into a computer each day. The manager would like to know what percentage of all of the entries made by these clerks on a certain day have typos. The manager is going to randomly select 60 of these records and review each of them for typos.
(a) State the population, parameter, and sample.

(b) The manager would like to not stay at the office all day, so she is considering using the 30 records from the first 2 clerks to finish that day. Explain how this could lead to bias.

(c) Describe how the manager could use a cluster sample to obtain a random sample of 60 records.

(d) Describe how the manager could use a stratified sample to obtain a random sample of 60 records.

(e) Explain the advantage of the method from (d) as opposed to that from (c).

What if you are looking to show that changes in one variable *cause* changes in some other variable? For that, you need an experiment. An experiment randomly assigns individuals to groups. Random assignment means the groups should be as similar as possible with regards to every variable you can't control directly. Assuming other variables have been directly controlled, you should at the end of your study be able to attribute any differences you saw to the treatment itself – that, or random chance!

There are many kinds of experiments. For the table below, suppose the setting is "Does taking a vitamin supplement every morning improve general well-being? 20 men and 20 women are surveyed at the beginning of the study. Then, 20 individuals will be assigned to take the supplement every day for a week, and the other 20 will receive a placebo. At the end of the study, the 40 people will again be surveyed and the average change in the well-being metric will be calculated for each group."

Name	Outline	Example	Purpose
1. Completely randomized	Number everyone, use hat or RNG to assign individuals to each group	Number the people 1-40. Use RNG to get 20 unique integers from 1-20 and assign them to the supplement. Assign the other 20 to the placebo.	Simplicity
2. Randomized block	Create groups that you expect to differ WRT what is being measured, then do completely randomized *within* each group	Maybe men and women respond differently to vitamins. Take the 20 men, number them 1-20, and use an RNG to get 10 unique numbers from 1-10. Assign these 10 men to the supplement and the other 10 to the placebo. Repeat this process for the women.	Reduces "noise" – variability in the response variable – by comparing groups similar to one another; makes it easier to see what changes were due to treatment
3. Matched pairs	Create blocks (pairs) of size 1 or 2 and do completely randomized *within* each *pair*	Pair the individuals based on their well-being metric at the beginning of the week – lowest with next lowest and so on. Flip a coin. If the coin is heads up, give the person with the lower metric the vitamin and the other the placebo. If tails, swap the treatments. Repeat this for all 20 pairs.	Further reduces "noise" – variability in the response variable – by comparing individuals similar to themselves; makes it easier to see what changes were due to treatment

A well-designed experiment allows you to infer that changes in the explanatory variable caused changes in the response variable. Because experiments *don't* typically randomly select individuals, though, you might NOT be able to infer this causation to any larger group.

AP TIPS

1. If you are ever asked to describe how to randomly assign individuals in an experiment, you must name the groups.
Example: "… assign these 20 people to one group and assign the other 20 to the other group." – Partial (P)
"… assign these 20 people to get the vitamin and the other 20 to get the supplement." – Essential (E)

2. Know your definitions: experimental units are the *smallest units that treatments get applied to*. This may be *groups* of individuals. Treatments are whatever is assigned to experimental units: this includes placebos. The explanatory variable is the independent ("x") variable; the response variable is the "dependent" ("y") variable.

3. If asked to describe variables, they must be able to vary.
Example: "The explanatory variable is taking the vitamin, and the response is having good well-being." – P
"The explanatory is whether or not they get the vitamin, and the response is the level of well-being." – E

PRACTICE: Three neighborhoods, which contain 20, 24, and 27 homes, respectively, require homeowners in the neighborhood to utilize a monthly lawn service during the spring and summer. The lawn service company is looking to test out a new weed-killer, and it gets the permission of all of the homes in the three neighborhoods to conduct an experiment. The company is going to randomly assign some homes to receive the regular weed-killer for 3 months and other homes to receive a new weed-killer. All other lawn services will be kept constant for the homes during this time.

(a) Is the study that is going to be conducted an experiment or an observational study?

(b) Describe how the company could use a completely randomized design to randomly assign which homes receive which weed-killer.

(c) The company is considering doing a randomized block design using the neighborhoods as blocks. Under what conditions would this be statistically advantageous?

The company conducted the study and found statistically significant evidence that homes with the new weed-killer had a lower average number of weeds than the homes with the regular weed-killer.
(d) Interpret what statistically significant means in this context.

(e) Can the company conclude that the new weed-killer is more effective at killing weeds than the regular one?

To display data from surveys, observational studies, or experiments, we need graphs.

Graphs exist for both quantitative variables and categorical variables. Quantitative variables have values that it would make sense to take an average of. Categorical variables have values that fit into categories – these may be numerical, like grade level or area code.

For both kinds of variables, it is paramount to determine whether the vertical (or horizontal) axis describes frequency – the *number* of observations – or relative frequency, the fraction or percentage of observations relative to the total.

<table>
<tr><th>Name</th><th>Number/type of variables</th><th>Shape</th><th>Center</th><th>Variability</th><th>Unusual features</th><th>Important notes</th></tr>
<tr><td>Dotplot</td><td rowspan="4">1 or 2 quantitative</td><td>Skew, mound, approx. normal</td><td>Median</td><td>Range, IQR</td><td>Outliers, gaps</td><td>None</td></tr>
<tr><td>Histogram</td><td>Skew, mound, approx. normal</td><td>Find the interval the median is in</td><td>Describe where most data is (clusters), estimate range</td><td>Potential outliers, gaps</td><td>You typically cannot state exact median or range.</td></tr>
<tr><td>Boxplot</td><td>Skew</td><td>Median</td><td>Range, IQR</td><td>Outliers</td><td>Cannot determine # of peaks

Can’t see normality</td></tr>
<tr><td>Stemplot</td><td>Skew, mound, approx. normal</td><td>Median</td><td>Range, IQR, clusters</td><td>Outliers, gaps</td><td>Pay attention to order of stems when determining skew.</td></tr>
<tr><th>Name</th><th>Number/type of variables</th><th colspan="4">How to describe</th><th>Important notes</th></tr>
<tr><td>Bar graph</td><td>1 categorical</td><td colspan="4">Compare relative frequencies in all categories</td><td>All categories!!

Pay attention to relative frequency if relevant</td></tr>
<tr><td>Side-by-side bar graph</td><td rowspan="3">2+ categorical</td><td colspan="4" rowspan="3">Compare values of one category *within* each value of the other category</td><td rowspan="3">All categories!!

Pay attention to relative frequency if relevant</td></tr>
<tr><td>Segmented bar graph</td></tr>
<tr><td>Mosaic plot</td></tr>
</table>

When describing a graph of a *quantitative* variable, you will want to discuss SCVU: shape, center, variability, and unusual features. As always, you will want context.

Much of this is natural to you, so let's discuss those particularities which are easy to forget:

1. The 5-number summary: A list of five numbers in the format #,#,#,#,#. These numbers, in order, are

Min, Q1, Median, Q3, Max

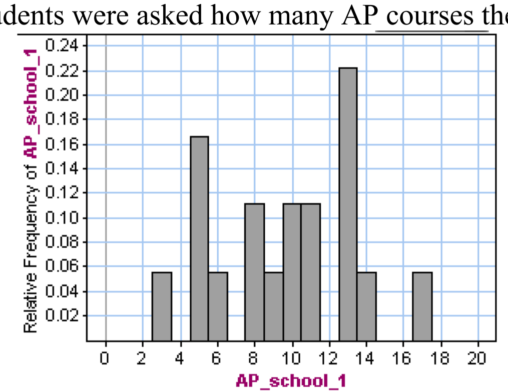
2. Finding outliers
   - Compute IQR = Q3 – Q1
   - Multiply this by 1.5
   - Add this to Q3 and subtract it from Q1; these values are the fences
   - Any value *above* Q3 + 1.5IQR and *below* Q1 – 1.5IQR is an outlier

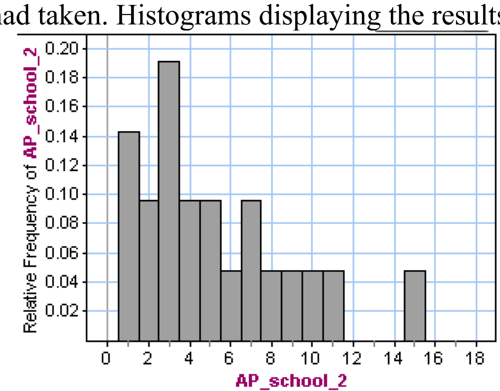
3. Median of a histogram
   - Find the total number of observations (or remember that the total relative frequency is 1)
   - Starting from the left, add up the frequencies (or relative frequencies) until you reach or pass the halfway mark – either half the total (frequency) or 0.5 (relative frequency)
   - The median is in the interval you passed the halfway mark in

4. Resistance
   - The following measures are nonresistant: they will change with the presence of much higher or lower values.
     - Measures of variability: Range, standard deviation
     - Measures of center: Mean
   - The following measures are resistant: they will change by little to none with the presence of much higher or lower values.
     - Measures of variability: IQR
     - Measures of center: Median

PRACTICE:

At School 1, eighteen AP Statistics students were asked how many AP courses they had taken. At School 2, twenty-one AP Statistics students were asked how many AP courses they had taken. Histograms displaying the results are below.





(a) Compare the distributions of the number of AP courses taken by the AP Statistics students at School 1 and School 2.

(b) Each of the students at School 1 was instructed to include the current AP Statistics course in their count of AP courses, while the School 2 students was instructed not to. If each student at School 1 were to remove the current AP Statistics course from their count, what effect would this have on the following summary statistics for School 1?

i. Mean

ii. Median

iii. IQR

iv. Standard deviation

(c) Below are the summary statistics for the students at School 1.

Minimum	Q1	Median	Q3	Maximum	Mean	St. Deviation
3	6	10	13	17	9.67	3.83

Create a boxplot for the number of AP courses taken by the AP Statistics students at School 1. What feature of the histogram is not visible in the boxplot?

(d) Estimate the median number of AP courses taken by all 39 AP Statistics students at School 1 and School 2. Show your work.

AP TIPS

1. When asked to compare distributions, use words like <u>greater than</u> and <u>less than</u>. You *need* comparison words.

2. When checking for outliers, you must <u>explicitly</u> compare any values to the fences. If you simply calculate the fences and then say, "There is a high outlier," you won't get credit: you need to explicitly state that upper value is *greater than* the fence.

3. Don't forget gaps! If a gap is present, it very well may be <u>the</u> most striking feature of the distribution.

When describing location in a distribution, we often use percentiles or z-scores.

The percentile of an observation is the percent of observations less than or equal to it. For instance, Q1 is the 25^{th} percentile, the median is the 50^{th} percentile, and Q3 is the 75^{th} percentile.

We may also use z-scores to identify position. A z-score determines the number of standard deviations above or below the mean that a value is. NOTE: z-scores exist *even in non-normal distributions*.

In a normal distribution, z-scores correspond to well-known areas. The Empirical Rule states:

Approximately 68% of observations are between -1 and 1 standard deviations from the mean.
Approximately 95% of observations are between -2 and 2 standard deviations from the mean.
Approximately 99.7% of observations are between -3 and 3 standard deviations from the mean.

The above rule is helpful for determining if a distribution with an unknown shape *could* be approximately normal: look at its areas compared to what they should be by the Empirical Rule if the distribution is close to normal!

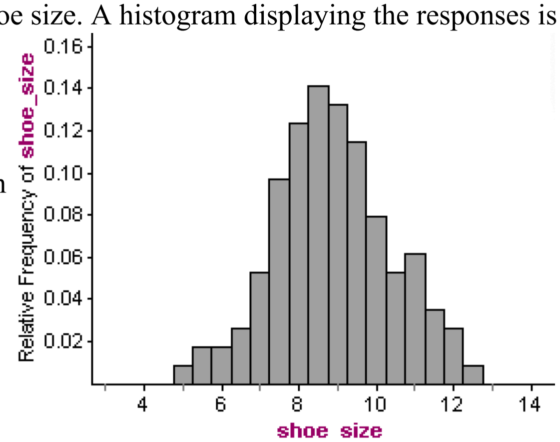
For other areas in a normal distribution, we use the following functions:

Function	Input	Output	Example
`normalcdf(`	Lower and upper value	Area under the curve (proportion of observations)	`normalcdf(1,1ØØØ,Ø,1)` ≈ 0.1587 *About 15.87% of observations have a z greater than 1.*
`invNorm(`	Area under to the curve to the left of the value you're interested in	The value (with the input area to the left of it)	`invNorm(Ø.3, 2Ø, 4)` ≈ 17.9 In a dist. with $\mu = 20$ and $\sigma = 4$, the value with 30% of observations below it is 17.9.

PRACTICE: A sample of 1,350 high school students were asked their shoe size. A histogram displaying the responses is shown below.

The mean of the distribution is 9 and the standard deviation is 1.5.

(a) Compute the proportion of observations within one standard deviation of the mean. Does this reasonably match up with a normal distribution?



Treat the distribution as if it is approximately normal.
(b) Approximately what proportion of students in the sample have a shoe size larger than 8?

(c) Estimate the 80th percentile of shoe sizes in the sample.

(d) In a sample of 1,200 middle school students, the mean was lower than 9, but the standard deviation was still 1.5. If a middle school student with a shoe size of 8 is at the 65th percentile in the sample, what was the mean shoe size of the middle school sample?

AP TIPS

1. ALWAYS DRAW A PICTURE WHEN USING NORMALCDF OR INVNORM.

2. Label your mean and standard deviation in your calculator-speak.

3. Remember that z-score percentiles remain constant in <u>any</u> normal distribution – if you aren't given a mean or standard deviation, just use <u>0</u> and <u>1</u>.

4. Always label your axes. If creating a stemplot, provide a key.

Lesson 2: Units 4-5

Probability is the value that a relative frequency of successes will approach after *many* trials.

In order to compute probabilities for quantities that can't be modeled well with normal, binomial, or geometric distributions, we use simulations: simply *simulate* the scenario, then carry out multiple trials.

To carry out a simulation, there are 4 general steps:
1. Ask (or rephrase) the question in terms of probability.
2. Choose a random process with matching probabilities to the real-world event.
3. Describe what constitutes one trial of the simulation and what to record after the simulation.
4. Perform many trials.
5. Answer the question of interest.

PRACTICE: In the early 2000s, McDonald's ran a Monopoly game where customers would get Monopoly pieces from their cups that could earn them prizes. According to McDonald's, there was a 9 in 40 chance (22.5%) that a customer would win a free food item. Jerome went to McDonald's regularly and felt he wasn't winning much free food: at one point, it took him 8 orders until we won a prize (he won his first prize on his 8^{th} order). He felt the game was rigged.

(a) Express "he felt the game was rigged" in terms of probability in relation to his needing 8 orders to win a prize.

(b) Design a simulation using a random digit table to determine the probability from (a).

(c) Use the portion of a random digit table below to carry out 2 trials of your simulation.

101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335

Simulations are valuable in answering many complex probability questions: we also have probability *rules* when working with known probabilities.

Probability of A or B: Add the probabilities of A and B, then subtract the intersection (what was double counted)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probability of A and B: Multiply the probability of A and the probability of B given that A already happened

$$P(A \cap B) = P(A) * P(B|A)$$

Probability of A given B: Using B as the total, divide the probability of A within B

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Two events are independent if one occurring first doesn't affect the probability of the other

$$P(A|B) = P(A)$$

and if the probability of both occurring is just the product of their probabilities

$$P(A \cap B) = P(A) * P(B)$$

PRACTICE: A random sample of 1,000 high school graduates was asked whether they had taken either of AP Statistics or AP Biology. The results are shown in the table below.

	Took AP Stats	Didn't take AP Stats	Total
Took AP Biology	125	160	285
Didn't take AP Biology	215	500	715
Total	340	660	1,000

(a) Find the probability that a randomly selected student in the sample took AP Biology or AP Statistics.

(b) Determine whether taking AP Biology and taking AP Statistics are independent in the sample.

Suppose we use the sample proportion of students who took AP Stats and infer this to the population of all high school graduates that year. Among all students who took AP Stats, 94% planned on attending a 4-year college; among all students who didn't take AP Stats, 63% planned on attending a 4-year college.

(c) What is the probability that a randomly selected graduate that year planned on attending a 4-year college?

(d) What is the probability that a randomly selected graduate who planned on attending a 4-year college took AP Stats?

When a random process is going to be conducted, we can often classify the outcomes as a probability distribution. A probability distribution outlines (a) the values the process can result in and (b) the probability of each of those values. In a probability distribution, all probabilities should add up to 1.

The results of such a process are called a random variable. There are two kinds of random variables: continuous and discrete. Continuous variables have *no gaps* – there are infinitely many values, like time, height, and weight. Discrete variables have *gaps*: anything that is a "number of successes" is discrete.

The mean of a random variable is called the expected value. The formulas for this and the standard deviation are:

$$\mu_X = E(X) = \sum x_i p_i = \text{sum of each value times its probability (weight)}$$

$$\sigma_X = \sum (x_i - \mu_X)^2 p_i = \text{sum of each squared deviation times its probability}$$

The expected value is the average number of successes in many trials. The standard deviation is typically how far away values will be from the expected value.

PRACTICE: In a small urban neighborhood with six parking meters, historical data shows that the number of working meters on a random selected day follow the random variable W with probability distribution below:

w	0	1	2	3	4	5	6
$P(w)$	0.01	0.05	0.08	0.19	0.3		0.2

(a) Calculate and interpret $P(W = 5)$.

(b) Compute and interpret $E(W)$.

(c) Compute the probability that all 6 meters are working given that at least 4 meters are working.

(d) Randomly select two days in a given year. What is the probability that 4 meters are working on both days?

(e) Randomly select four days in a given year. What is the probability that all 6 meters are working on at least one of those days?

Two predictable kinds of discrete random variables are binomial and geometric.

Distribution	Description	Conditions	Formulas	Shape
Binomial	$X =$ number of successes in fixed number of trials Parameters: n, p	1. Binary (success or failure) 2. Independent trials 3. Number of trials is fixed 4. Probability of success stays the same	$\mu_X = np$ $\sigma_X = \sqrt{np(1-p)}$	Right skewed if $np \leq 10$ Left skewed if $n(1-p) \leq 10$ Approx. normal if $np, n(1-p) \geq 10$
Geometric	$G =$ number of trials until getting first success Parameter: p	1. Binary 2. Independent trials 3. Counting number of trials until one success	$\mu_G = \frac{1}{p}$ $\sigma_G = \frac{\sqrt{1-p}}{p}$	Always right skewed

There are nice formulas and calculator functions for calculating binomial and geometric probabilities.

	Desired probability	Formula by hand	Calculator function
Binomial	Precisely x successes	$\binom{n}{x} p^x (1-p)^{n-x}$ *x successes, n-x failures,* $\binom{n}{x}$ *arrangements*	`binompdf(`
	Less than or equal to x successes	Use formula above, *add up multiple of them*	`binomcdf(`
Geometric	Precisely x trials until success	$(1-p)^{x-1}(p)$ *x-1 failures, then a success*	`geometpdf(`
	Less than or equal to x trials until success	$1-(1-p)^x$ *1 minus the probability of all failures*	`geometcdf(`

PRACTICE: A professional development presenter presented to 973 employees at a business conference. At the end of the presentation, each employee was asked to fill out an anonymous survey where they responded to the statement, "This presentation was helpful." The results of the survey are displayed below.

Answer	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
Frequency	294	108	66	264	241

The presenter would like to randomly sample 20 attendees and ask them for more detailed feedback. Let $S =$ number of employees in the sample who answered, "strongly disagree."

(a) Describe the distribution of S.

(b) The presenter wants to ensure that they get at least 8 people who responded "strongly disagree" in the sample. Calculate the probability the presenter will get at least 8 such people.

(c) After the presentation, the presenter was walking around and talking to employees. Each one seemed like they loved the presentation. At one point, 10 people in a row all told the presenter they found the presentation very helpful.

Upon seeing the survey results, though, the presenter felt that these people must have just been nice to his face. Let $A =$ number of people it takes to find one who *wouldn't* have said "strongly agree."

(d) Describe the distribution of A.

(e) The presenter said he saw 10 people in a row who said they loved the presentation. Find $P(A > 10)$.

(f) Based on your answer to (e), does the presenter have reason to be suspicious of the 10 people in a row who said they loved the presentation?

AP TIPS

1. The less work there is to show, <u>the more important it is to show it</u>.

2. When using any of the pdf or cdf functions, <u>label all of your inputs</u>.

3. Always label n and p for binomial calculations and p for geometric calculations.

4. Your formula sheet has all of the formulas for binomial and geometric distributions.

Lesson 3: Units 6-8

When we want to estimate a population parameter, we take a sample and compute a statistic. Every sample will be different, though, so random sampling will produce an entire distribution of different values of the statistic. The distribution of all possible values of a statistic for a given sample size is called a sampling distribution.

The statistics we worked with in this class are:

- sample proportion $\hat{p}$
- sample mean $\bar{x}$
- difference of sample proportions $\hat{p}_1 - \hat{p}_2$
- difference of sample means $\bar{x}_1 - \bar{x}_2$
- sample mean difference (mean of sample differences) $\bar{x}_d$
- slope of sample regression line b

For *any* of these statistics, we have the following characteristics of the sampling distribution:

1. The mean of all possible statistics is equal to the actual parameter *if random sampling or assignment has been used.*

Definition: When the mean of the sampling distribution of a statistic is equal to the actual parameter, then the statistic is called an unbiased estimator of the parameter.

2. The standard deviation of all possible statistics is the corresponding formula assuming *the individuals are reasonably independent (and samples are reasonably independent if multiple were taken).* This often requires checking that the sample size is less than 10% of the population size (so that the probability of success only changes *negligibly* when sampling without replacement).

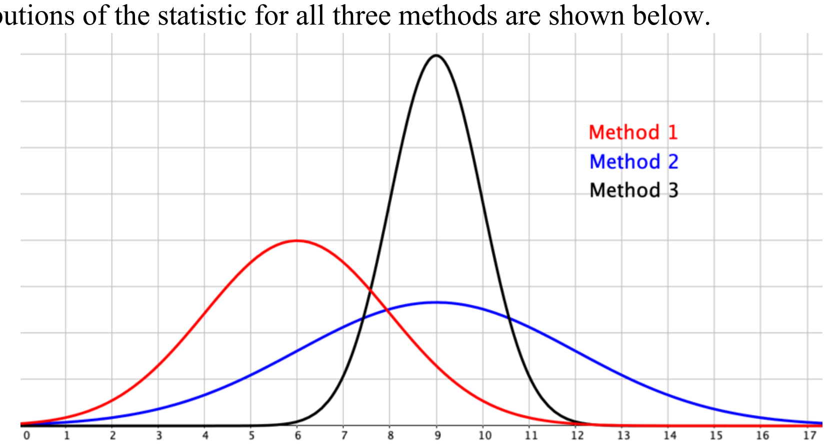
3. The shape of the sampling distribution will depend upon the sample size and perhaps p. For proportions, this amounts to checking the Large Counts condition; for means, it is the Normality condition.

The formulas for all sampling distributions are on the formula sheet and can be found in the Inference Summary.

One thing to pay attention to in *all* of the standard deviation formulas: there is a $\sqrt{n}$ in the denominator. This means that *increasing the sample size by a factor of k will reduce the standard deviation of the sampling distribution by a factor of* $\sqrt{k}$.

PRACTICE: At one school, the average number of cousins per student is 9 and the standard deviation is 6.

Three different sampling methods are going to be used to get a sample of students and calculate a sample mean in order to estimate the population mean. The sampling distributions of the statistic for all three methods are shown below.



(a) What is the likely shape of the population distribution?

(b) Which of the three methods will result in unbiased estimators? Justify.

(c) What was the sample size for method 2?

(d) If the sample size for Method 3 were 100 instead of 4, what effect would this have on the standard deviation of the sampling distribution of the statistics from Method 3?

We use sampling distributions and the shape of normal distributions to construct confidence intervals.

Every confidence interval is of the form:

$$\text{Statistic (Point Estimate)} \pm \underbrace{\text{Critical Value} \times \text{Standard Deviation of Statistic}}_{\text{Margin of Error}}$$

The critical value for a C% confidence interval is the number of standard deviations you need to go on either side of the mean to capture an area of 0.C. We use z^* when we know (or have an estimate) of the population standard deviation; we use t^* when we *don't* know or have a reliable estimate of the population standard deviation (which is usually the case when working with means).

Confidence intervals rely on a confidence level. The level C% means that *if many samples were conducted and an interval were formed for each one, then C% of the intervals would capture the population parameter.*
The C% interval itself is a range of plausible values for the parameter: *any* value in the C% interval is plausible assuming that C is reasonably high (let's say at least 90).

We interpret intervals by simply stating: *We are C% confident that the interval from ___ to ___ captures the population parameter.*

PRACTICE: BLITZ OF MISCELLANEOUS QUESTIONS

1. Suppose we want to construct a two-proportion z-interval to estimate a difference between population proportions. The samples result in $n_1 = 80, \hat{p}_1 = 0.338, n_2 = 112, \hat{p}_2 = 0.446$.

(a) Check the Large Counts condition.

(b) Compute a two-proportion z-interval with 90% confidence with your calculator.

(c) Is the difference between the sample proportions statistically significant?

(d) If a confidence level of 95% were used instead of 90%, would the interval be wider or narrower than the one from (b)?

(e) If the sample size of 112 were instead a sample of size 200, would the 90% interval be wider or narrower than the one from (b)?

2. When estimating a population mean, previous data indicates that $\sigma = 12$. If we were to construct a 99% interval and wanted the margin of error to be no larger than 2, what is the minimal sample size that would be needed?

3. A company board would like a certain proposal to pass at an upcoming shareholder meeting. It would like to estimate what proportion of shareholders will vote in agreement with them, so they are going to take a random sample and compute a 95% confidence interval.

(a) If they want the margin of error of their 95% interval to be at most 2%, what is the smallest sample size the company can select?

(b) A survey from a month prior indicated that 58% of shareholders agreed with the company board. If the company believes this result is still likely close to the truth, then what is the smallest sample size the company can select to keep the margin of error of the 95% interval to at most 2%?

4. When estimating a population mean, a sample of size 20 is selected and no information about the population distribution is known.

(a) What must be done to check the Normality condition?

(b) Find the critical value that should be used to construct a 95% interval about the sample mean.

5. A recent poll asked 1,005 Americans if they approve or disapprove of President Biden. A resulting 95% confidence interval was 0.44 ± 0.038.

(a) Interpret the confidence interval.

(b) Interpret the confidence level.

(c) Which is more plausible: that 44% of Americans approve of President Biden, or that 47% of Americans approve of President Biden?

We'll spend the rest of this lesson discussing the <u>Inference Summary</u> and looking at practice FRQs.

AP TIPS

1. For one- and two-proportion intervals, you can use the <u>observed</u> numbers of successes and failures when checking Large Counts. This may require using $\hat{p}$ (or $\hat{p}_1$ and $\hat{p}_2$).

2. Never leave off your C% when referring to a confidence interval!

3. Don't forget that the standard deviation of the sampling distribution is <u>inversely</u> related to sample size by a factor of $\sqrt{k}$.

4. Don't memorize critical values: pay very close attention to whether a problem is about <u>proportions</u> or <u>means/slopes</u>.

5. If calculating a <u>minimum</u> sample size required for a maximal margin of error, always round <u>up</u>.

6. If a value is <u>inside</u> a C% interval, then it's *plausible* – no matter where it is within the interval.

Lesson 4: Units 9-10

Confidence intervals provide a <u>range of estimates</u> for a parameter. If we simply want to know *whether we have convincing evidence* that a parameter is less than, greater than, or not equal to some value, we conduct a <u>hypothesis test</u>.

The logic of a test is as follows:

LOGIC OF A HYPOTHESIS TEST

1. Form an alternative hypothesis: What are you looking for evidence of?

2. Form a null hypothesis: this is the hypothesis you are testing *against*.

3. Choose a significance level α. This is the threshold for being convinced.

4. Conduct a randomized, controlled study and get a sample statistic.
Does this statistic provide evidence of H_a? Proceed.

5. Your randomized, controlled study should mean that you have evidence for H_a for only <u>one of two reasons</u>:
Reason 1: Your alternative hypothesis is correct – you got your evidence because H_a is correct.
Reason 2: The null hypothesis is true, and you got your evidence just by chance in the randomization process.

6. Compute the <u>test statistic</u>: the number of standard deviations away your sample statistic is from the <u>null</u> parameter. The higher this is, the better the evidence you have that H_0 is incorrect.

7. Compute the <u>P-value</u>, the probability of getting *this extreme* of a test statistic <u>if the null is true</u>.
The <u>lower</u> this probability is, the <u>less likely</u> it is that the statistic would've occurred <u>if the null was true</u>. The lower this is, the *better* the evidence you have that <u>the null couldn't be correct</u>.
The <u>higher</u> this probability is, the <u>more likely</u> it is that the statistic <u>could've occurred by chance even if the null is true</u>. The higher this is, the more plausible it is that <u>the null *could* still be correct</u>.

8. If $P < \alpha$, then <u>reject</u> the null hypothesis. This is because you have <u>convincing</u> (highly unlikely if H_0 is true) <u>evidence</u> of H_a.

If $P > \alpha$, then you <u>fail</u> to reject the null hypothesis. This is because you *don't* have convincing evidence of H_a since it's <u>plausible</u> a statistic as extreme as the one you got could've occurred by chance <u>even if H_0 was true.</u>

PRACTICE: A computer battery manufacturer wanted to test if their newest batteries lasted longer than 8 hours on average.
(a) Define the parameter of interest and state appropriate hypotheses for a significance test.

The company took a random sample of 40 batteries and got a sample mean of $\bar{x}$, which led to a test statistic of 1.87. They are going to use a significance level of $\alpha = 0.05$.

(b) Interpret this test statistic.

(c) Draw a picture and shade the area that corresponds to the P-value. Then, compute this P-value.

(d) Interpret the P-value from (c).

(e) What conclusion should the manufacturer reach?

Of course, some studies may reach an incorrect conclusion. These are called Type I and Type II errors. Here's a comprehensive list of interpretations and examples.

Error Type	Interpretation	Example	Probability
Type I	*Reject H_0 when it was true* *Be convinced of H_a when it was false* *Get a P-value less than α just by chance*	False positive on drug test Convinced of presence of disease when there's no disease Think a proportion is too high when it isn't too high	α – you control it!
Type II	*Fail to reject H_0 when it was false* *Not be convinced of H_a when it was true*	Rule not guilty when person was guilty *Not* convinced that disease is present when it is *Don't think* that an average is too low when it actually is	β – harder to compute

We often try to avoid Type II errors. This leads to the concept of power.

	Interpretation	How to Increase It	<--- Why
Power	Probability of rejecting H_0 if null is false Probability of being convinced of H_a when it's true Must have specified value of parameter that would make H_a true	1. Increase α 2. Increase n 3. Increase how far away proposed parameter is from H_0	1. Larger rejection region -> easier to be convinced (more P-values lead to rejection) 2. Decrease st. dev. of sampling dist. -> larger rejection region *More precise test -> easier to detect* 3. Easier to detect H_0 is wrong if it's *more wrong*

PRACTICE: Two large neighborhoods are adjacent in a certain town. A resident of Neighborhood 1 feels that all of the cars pulling in to Neighborhood 2 seem quite new, and they wonder if the average car is newer in Neighborhood 2 than Neighborhood 1. If that is the case, then they are going to try to "keep up with the Joneses" and buy a new car. They decide to randomly sample 10 cars from each neighborhood and compare the average years of these samples. Sample data is below.

Neighborhood	Years of cars
Neighborhood 1	2006, 2007, 2009, 2012, 2014, 2014, 2015, 2015, 2018, 2019
Neighborhood 2	2012, 2013, 2013, 2015, 2018, 2018, 2019, 2019, 2020, 2020

(a) State the parameter/s of interest and appropriate hypotheses for a significance test.

(b) Name the appropriate test for this data and check the conditions for inference.

(c) Compute the test statistic and P-value.

(d) What conclusion should the resident reach using a significance level of $\alpha = 0.01$?

(e) What kind of error could the resident have made? Explain a potential consequence (lol) of this error.

(f) If the resident had indeed made the kind of error in part (e), what could they do to reduce this probability of error if they were to conduct a similar study again?

We'll spend the rest of this lesson discussing the Inference Summary and looking at practice FRQs.

AP TIPS

1. On one-proportion z-tests, you must use p_0 when calculating standard error AND Large Counts.

2. On two-proportion z-tests, you must use $p_C = \frac{x_1+x_2}{n_1+n_2}$ when calculating standard error AND Large Counts.

3. Stick to the script on conclusions! "Because P … α, we reject/fail to reject… We have/don't have convincing evidence that…"

4. We are NEVER convinced of the null hypothesis. So, on Type II errors, don't say "We think that… null is true". Instead, it's "We *don't think… alternative is true*…"

5. If asked to interpret a P-value, you must begin with "Assuming the null (in context) is true…".

6. On the Do step, use the calculator!! You'll spend much of your time on conditions, so why waste it on Do?

Lesson 5: Solve for why?

The difference between a 3 and 5 on the AP Statistics exam may come down to not just knowing what, but *why*.

Below are a series of "why" questions for you to answer. Note: if a question is in context, the answer should be in context!

Some tips:

- Don't say "the data" or "results." Determine what is being measured and refer to it specifically ("proportion of people in this city who plan on voting" or "population average number of miles walked last week").
- If your answer can be followed by "So what?", then you haven't fully answered. For instance: "This could lead to bias because then the doctor might know what treatment patients received." To which one might say: SO WHAT? This is only a problem if the doctor would treat patients differently! So, we might add "… which might systematically affect how the doctor assesses the patients. For instance, the doctor might treat patients who received the new drug as if they improved more than they actually did."
- Know your conditions!

1. A high school principal wants to estimate what percentage of students at her school think she's a good principal. She decides to obtain a random sample of 50 students and have them complete a quick survey.
(a) What is the purpose of the high school principal <u>randomly</u> selecting the 50 students?

(b) The principal is considering asking each student to come to her office to fill out the survey. Explain why this might lead to bias.

2. A researcher is testing which of two diaper creams is more effective at reducing the severity of diaper rashes. They have 20 babies with diaper rash that have been given permission to participate in the study. Rather than randomly assign 10 of the babies to one cream and the remaining 10 babies to the other cream, the researcher is considering pairing the babies by initial severity of the rash and then randomly assigning one baby in each pair to each cream. What is the purpose of the pairing compared to simply randomly assigning 10 babies to each cream?

3. A company that runs 10 daycares all over metro Atlanta is considering changing their daily hours from closing at 6:00 p.m. to closing at 5:00 p.m. They would like to conduct a survey to estimate what percentage of families feel they would be greatly affected by this change.
(a) The company is considering using a stratified random sample. Identify a good stratifying variable and explain why you chose this variable.

(b) Why might the company want to conduct a stratified random sample over a cluster random sample?

4. The assignment of treatments in an experiment is almost always done *randomly*. Why is this?

5. A study found that 10 middle schools that required students to wear a uniform had a higher percentage of students graduate from college than did 12 middle schools that didn't require students to wear any kind of uniform. Suppose that both sets of middle schools were randomly sampled out of the respective populations of all middle schools that do/do not require a uniform.
(a) Was this an experiment or observational study? Explain.

(b) Why can we <u>not</u> conclude that wearing uniforms makes students more likely to graduate from college?

6. When performing literally any inference procedure, we check a condition called *Random*. Why do we check this?

7. When we perform inference procedures for means or proportions, we check that the sample size is less than 10% of the population size. Why do we check this?

8. When we perform inference procedures for proportions, we check some variation of $np \geq 10$ and $n(1-p) \geq 10$. Why do we check this?

9. When we perform inference procedures for means, we check whether the population distribution is normal, the sample size/s is/are greater than 30, or whether sample data looks reasonably symmetric. Why do we do this?

10. Why does p_C exist, and when do we use it?

11. Why is the median resistant, while the mean is not?

Lesson 6: Simulation Activity

Simulation Activity
To see if fish oil can help reduce blood pressure, 12 males with high blood pressure were recruited and randomly assigned to one of two treatments. The first treatment was a four-week diet that included fish oil and the second was a four-week diet that included regular oil. At the end of the four weeks, each volunteer's blood pressure was measured again and the reduction in diastolic blood pressure was recorded. The results of this study are shown below. Note that a negative value means the subjects blood pressure *increased* [*New England Journal of Medicine* 320 (1989)].

Fish Oil:	8	12	10	14	0	0	$\bar{x}_f = 7.33$
Regular Oil:	-6	0	1	2	-4	6	$\bar{x}_r = -0.17$

Difference $(\bar{x}_f - \bar{x}_r)$ = 7.50

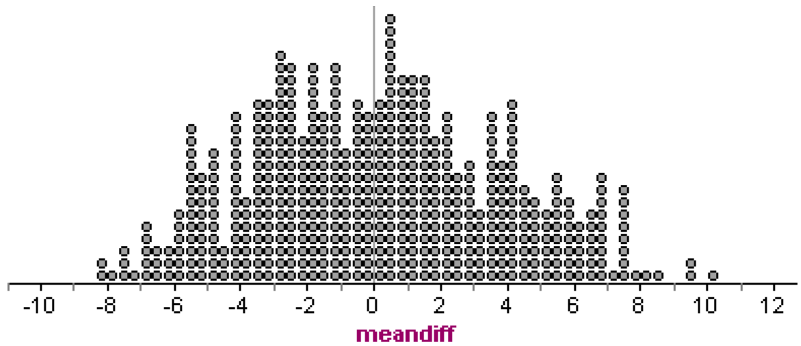
Is this convincing evidence that fish oil is better than regular oil for reducing diastolic blood pressure? In other words, is the difference in mean reduction statistically significant?

(a) Should/can we carry out a two-sample t-test? Why or why not?

(b) Describe a simulation that could be conducted to estimate the likelihood of getting a mean difference as large as 7.50 *assuming that the types of oil had the same effects.*

(c) Carry out 3 trials of your simulation.

The dotplot below shows the mean difference for 500 simulations similar to those you carried out.

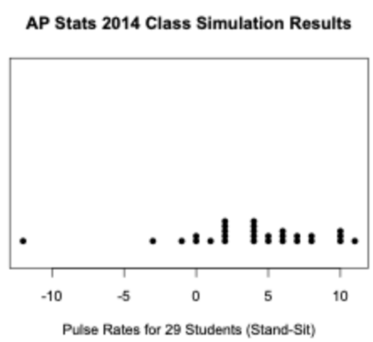


Based on the dotplot, is there convincing evidence that fish oil is more effective at reducing blood pressure than regular oil?

<u>Summary Example</u>:
An AP Statistics class (n = 29) did an activity where each student was randomly assigned a "treatment," standing or sitting, then measured their pulse rate for 30 seconds. Then each student did the other treatment (as a matched pairs design). The graph of the difference for each student and some summary statistics are given below.

Summary Statistics for Distribution of Difference in Pulse Rates (Standing – Sitting)

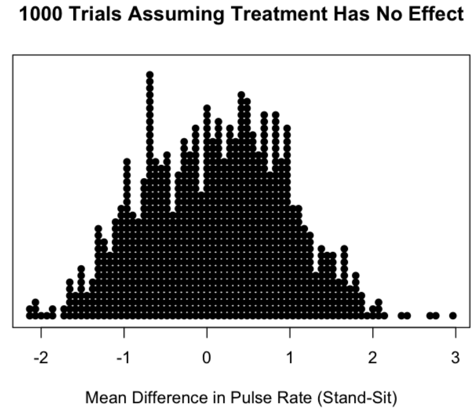
Mean: 3.93	Standard deviation: 4.67	Median: 4.0



(a) Based on the graph of the distribution of differences in pulse rate, do you *think* we have statistically significant evidence that standing pulse rates are higher than sitting pulse rates?

Let's assume that the treatments (standing and sitting) had no effect whatsoever. This would mean that a person's pulse rate was different because of other factors (time of day, cortisol levels, and other uncontrolled factors), *but not from the fact that they were standing or sitting* – had they been randomly assigned the opposite treatment, their pulse rate would have been the same.

The teacher performed a simulation with statistics software under the assumption that the differences in standing or sitting pulse rate occurred only by random chance alone, not because of the treatments themselves. The dotplot with 1000 trials is shown.



(b) What does one dot represent in the dot plot?

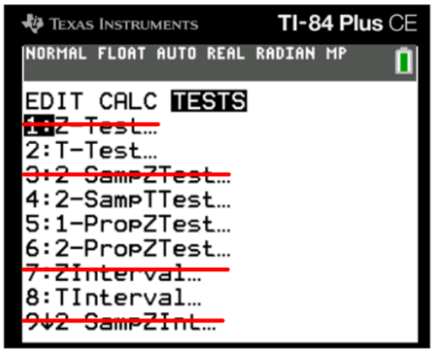
(c) Does this simulation provide evidence that standing heart rates are higher than sitting pulse rates?

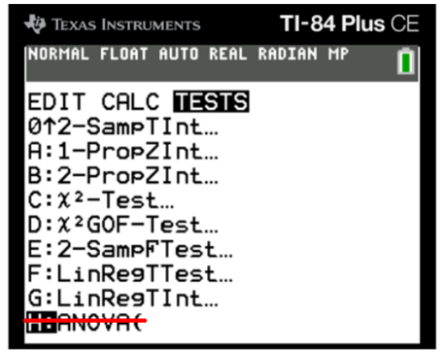
Lesson 7: Wrapping Up

1. Make sure your calculator is charged/has new batteries.

2. Go to `MODE` and make sure your `Stat Wizards` are `ON`.

3. Which tests will you certainly NOT use on the exam?





4. Strategy for attacking FRQs
   - Go knock out #1 or #2 (the easiest question).
   - Go start on #6: go as far as you can.
   - Go back and finish #1-5.
   - Go back to #6.

5. Tricky Fact Blitz: True or False?

   (a) The standard error of $\hat{p}_1 - \hat{p}_2$ for a two-proportion z-test is $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.

   (b) The area to the right of 2 standard deviations above the mean is larger under a t-distribution with 20 df than under a normal distribution.

   (c) We use t because that's what we always use for means.

   (d) If an experiment randomly assigned treatments, then you can infer the results of the experiment to a larger population.

   (e) A segmented bar graph is used to show the relationship between two quantitative variables.

   (f) An interval for a population slope uses a t-critical value with $df = n - 2$.

   (g) A value of r close to 1 or -1 indicates a linear relationship is appropriate.

6. Visiting the Cram Sheet

7. Miscellaneous: Any questions or topics you'd like to discuss? If not, we can work on a practice exam.

Inference Summary

	The Four-Step Process & how to get a 4	
	CONFIDENCE INTERVALS	**SIGNIFICANCE TESTS**
STATE	State the parameter of interest. State the confidence level.	Define the parameter/s. State your hypotheses and α.
PLAN	Name the method or give the formula. Check conditions.	Name the test. Check conditions.
DO	Give correct interval.	Report the test statistic and P-value. Report df if relevant.
CONCLUDE	We are ___% confident that the interval from ____ to _____ captures the true parameter in context.	ONE OF TWO OPTIONS: 1. Because $P =$___<___$= \alpha$, we reject H_0. We have convincing evidence that alternative hypothesis in context. OR 2. Because $P =$___>___$= \alpha$, we fail to reject H_0. We do not have convincing evidence that alternative hypothesis in context.

Confidence Interval: statistic $\pm$ (critical value)(standard deviation of statistic) … (for inference about slope, this is just SE_b)

Standardized test statistic $= \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$ ..... this is the z or t we use to calculate the P-value.

Below is a comprehensive list of the inference methods in AP Statistics. They are largely separated into two categories: PROPORTIONS and MEANS.

How do you know when to use **PROPORTIONS**? Key words: **PERCENTAGE**, **PROPORTION**, **RATIO**

How do you know when to use **MEANS**? Key words: **MEAN**, **AVERAGE**, **NUMBER OF**

How do you know when to do an **INTERVAL**? Key words: **CONSTRUCT**, **CONFIDENCE LEVEL**, **ESTIMATE**

How do you know when to do a **HYPOTHESIS TEST**? Key words: **CARRY OUT**, **CONVINCING EVIDENCE, STATISTICALLY SIGNIFICANT**

WHY DO WE HAVE TO USE t, AND HOW DOES IT DIFFER FROM z?

- We use t when we don't know the population standard deviation – this exclusively applies to means and slopes.
- Any critical value t is larger than its corresponding z critical value for any finite n.
- The shape of a t distribution (for means) is based on the degrees of freedom, calculated by $n-1$ for means and $n-2$ for slopes. The larger the sample size, the larger the degrees of freedom, and the more the t distribution resembles a Normal distribution.

CALCULATING CRITICAL VALUES FOR CONFIDENCE INTERVALS

Let $C\%$ be the confidence level. Then, convert this into a decimal $(0.C)$ and use one of the following commands:

- For z, enter: **invNorm**$(\mathbf{0.C + \frac{1}{2}(1 - C), 0, 1})$.
- For t, enter: **invT**$(\mathbf{0.C + \frac{1}{2}(1 - C)}, \textbf{df})$

All this says is that you must remember to add the left tail, since both calculator commands only do areas to the left.

TABLE I: INFERENCE ABOUT A PROPORTION

Name	Formula/s	TI83/84 Function	Conditions	Additional Information
One-proportion z-interval for p	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $z^* = \text{invNorm}(\text{low}, .C + \frac{1}{2}(1 - .C))$	**A: 1-PropZInt**	1. Random: Data from a random sample or randomized experiment 2. Independence: Sample size is less than 10% of population size 3. Large Counts: At least 10 successes and failures: $n\hat{p} \geq 10$, $n(1-\hat{p}) \geq 10$ MUST USE p_0 for LC if a test.	Must show specific numbers for condition #2 Interval contains all plausible values of parameter p
One-proportion z-test for p	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ P-value: Use normalcdf	**5: 1-PropZTest**		Reject if $P < \alpha$ Fail to reject if $P > \alpha$

TABLE 2: INFERENCE ABOUT A MEAN

Name	Formula/s	TI83/84 Function	Conditions	Additional Information
One-sample t-interval for μ	$\bar{x} \pm t^* \left(\frac{s_x}{\sqrt{n}}\right)$ with df $= n - 1$ $t^* = \text{invT}(\text{low}, .C + \frac{1}{2}(1 - .C), \text{df})$	**8: TInterval**	1. Random: Data from a random sample or randomized experiment 1a. $n \leq \frac{1}{10}N$ if sampling without replacement 2. Independence: Reasonable to believe that individuals in sample are independent of one another (based on context of problem – usually randomness!) (if *paired*, order of treatments was randomized) 3. Normality: Population distribution is approx. Normal OR $n \geq 30$ (CLT) OR Sample data has no strong skew or outliers	Must draw boxplot if $n < 30$ Interval contains all plausible values of parameter μ
One-sample t-test for μ (could be PAIRED)	$z = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$ with df $= n - 1$ P-value: Use tcdf if PAIRED $z = \frac{\bar{x}_D - 0}{\frac{s_{x_D}}{\sqrt{n}}}$	**2: TTest**		Reject if $P < \alpha$ Fail to reject if $P > \alpha$

TABLE 3: INFERENCE ABOUT TWO PROPORTIONS

Name	Formula/s	TI83/84 Function	Conditions	Additional Information
2-proportion z-interval for $p_1 - p_2$	$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	**B: 2-PropZInt**	1. <u>Random</u>: Data from *independent* random samples or randomized experiment 1a. $n_1 \leq \frac{1}{10}N_1, n_2 \leq \frac{1}{10}N_2$ if sampling without replacement 2. <u>Independence</u>: Both sample sizes are less than 10% of their respective population sizes 3. <u>Large Counts</u>: At least 10 successes and 10 failures for both samples MUST USE $p_c = \frac{x_1+x_2}{n_1+n_2}$ for 2-prop test	Large Counts – use ACTUAL for interval Large counts – MUST CALCULATE (use H_0) Interval contains all plausible values of the parameter (the *difference*) $p_1 - p_2$
2-proportion z-test for $p_1 - p_2$	$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$ where p_c is overall proportion of successes (still must be weighted by respective sample sizes)	**6: 2-PropZTest**		Must show specific numbers for condition #2 Reject if $P < \alpha$ Fail to rej. if $P > \alpha$

TABLE 4: INFERENCE ABOUT TWO MEANS

Name	Formula/s	TI83/84 Function	Conditions	Additional Information
2-sample t-interval for $\mu_1 - \mu_2$	$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ Use whichever df is <u>smaller</u>.	**0: 2-SampTInt**	1. <u>Random</u>: Data from random samples or randomized experiment 2. <u>Independent</u>: Reasonable to believe that samples are independent of one another (based on context of problem – usually randomness!) 3. <u>Large Counts</u>: Population distribution is approx. Normal OR $n \geq 30$ (CLT) OR Sample data has no strong skew or outliers	Must draw boxplot if $n < 30$ Interval contains all plausible values of the parameter (the *difference*) $p_1 - p_2$
2-sample t-test for $\mu_1 - \mu_2$	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ where p_c is overall proportion of	**4: 2-SampTTest**		Must draw boxplot if $n < 30$ Reject if $P < \alpha$ Fail to reject if $P > \alpha$

TABLE 4: INFERENCE ABOUT DISTRIBUTION OF CATEGORICAL VARIABLES

# of Samples/Groups	Name	Formula/s	TI83/84 Function	Conditions	Additional Information
1	Chi-square test for goodness of fit	$\chi^2 = \Sigma \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ $\text{df} = \text{\# of categories} - 1$	**D: χ^2GOF-Test**	1. <u>Random</u>: Data from random samples or randomized experiment 2. <u>Large Counts</u>: All *expected* counts are at least 5	Enter observed counts into L1 and expected counts into L2 Expected counts = Multiply each category size by the assumed proportion
2 or more	Chi-square test for homogeneity	$\chi^2 = \Sigma \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ df = (# rows – 1)(# columns – 1)	**C: χ^2-Test**		Enter observed counts into Matrix A, run test, then go back to Matrix B and you will have your expected counts Expected counts = $\frac{\text{(row total)(column total)}}{\text{overall total}}$

How to tell apart test of HOMOGENEITY and test of INDEPENDENCE/ASSOCIATION
A test of homogeneity will have either *multiple* samples or an overall group split into *multiple* treatment groups.
A test of independence/association will have <u>one</u> sample of individuals that are just *categorized* by two variables.

TABLE 5: RELATIONSHIP BETWEEN TWO CATEGORICAL VARIABLES

# of Samples	Name	Formula/s	TI83/84 Function	Conditions	Additional Information
2	Chi-square test for independence	$\chi^2 = \Sigma \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ df = (# rows – 1)(# columns – 1)	**C: χ^2-Test**	1. <u>Random</u>: Data from random samples or randomized experiment 2. <u>Large Counts</u>: All *expected* counts are at least 5	Enter observed counts into Matrix A, run test, then go back to Matrix B and you will have your expected counts Expected counts = $\frac{\text{(row total)(column total)}}{\text{overall total}}$

TABLE 6: INFERENCE ABOUT A SLOPE (RELATIONSHIP BETWEEN 2 QUANTITATIVE VARIABLES)

Name	Formula/s	TI83/84 Function	Conditions	Additional Information
One-sample t-interval for β	$b \pm t^*(SE_b)$ $\text{df} = n - 2$ SE_b will usually be given in computer readout; no need to divide by $\sqrt{n}$.	**G: LinRegTInt**	1. Linear: - scatterplot is linear, residual plot has no form 2. Independent observations; check 10% condition if sampling without replacement. 3. Normal: Responses are scattered Normally about residual = 0 line for all x. 4. Equal SD: Responses scattered evenly about residual = 0 line. 5. Random: Data comes from a random sample or randomized experiment.	Interval contains all plausible values of the parameter (the *true slope*) β
One-sample t-test for β	$t = \frac{b - \beta_0}{SE_b}$ (usually $t = \frac{b-0}{SE_b}$) $\text{df} = n - 2$	**F: LinRegTTest**	1. Linear: - scatterplot is linear, residual plot has no form 2. Independent observations; check 10% condition if sampling without replacement. 3. Normal: Responses are scattered Normally about residual = 0 line for all x. 4. Equal SD: Responses scattered evenly about residual = 0 line. 5. Random: Data comes from a random sample or randomized experiment.	Reject if $P < \alpha$ Fail to reject if $P > \alpha$

Correct Interpretations of Intervals Every interval starts with this: "We are C% confident that the interval from ____ to ____ captures…"		
Type of Interval	**Interpretation (continued from script above)**	**Notes**
One proportion z-interval	the true proportion of … in (population).	Context is your best friend.
One sample t-interval	the true mean … in (population).	Context is your best friend.
Paired t-interval	the true mean difference of … (order of subtraction) in (population).	Mean difference = singular! Subtraction occurs in pairs *before* mean is calculated.
Two-proportion z-interval	the true difference in proportions of … (order of subtraction) in (population).	Order of subtraction IS important.
Two-sample t-interval	the true difference in mean … (order of subtraction) in (population).	Order of subtraction IS important, and "true difference in means" is *plural* (aligns with *two*-sample). Subtraction occurs *after* the two means are calculated.

Correct Forms of Hypotheses

Test	Correct H_0	Correct H_a (only one in each row, of course)
One-proportion z-test	$H_0: p = p_0$(some value)	$H_a: p >, <, \neq p_0$
One-sample t-test	$H_0: \mu = \mu_0$(some value)	$H_a: \mu >, <, \neq \mu_0$
Paired t-test	$H_0: \mu_D = 0$	$H_a: \mu_D >, <, \neq 0$
Two-proportion z-test	$H_0: p_1 - p_2 = 0$ OR $H_0: p_1 = p_2$	$H_a: p_1 - p_2 >, <, \neq 0$ $H_a: p_1 >, <, \neq p_2$
Two-sample t-test	$H_0: \mu_1 - \mu_2 = 0$ OR $H_0: \mu_1 = \mu_2$	$H_a: \mu_1 - \mu_2 >, <, \neq 0$ $H_a: \mu_1 >, <, \neq \mu_2$
Chi-squared GOF test	H_0: The distribution of proportions of _____ are equal to those stated.	H_a: The stated distribution of proportions of ______ are not equal to those stated.
Chi-squared homogeneity	H_0: The distributions of _______ are the same for each _________. OR Often, just use the context of the problem.	H_a: The distributions of _______ are not the same for each _________. OR Often, just use the context of the problem.
Chi-squared association/ independence	H_0: _____ and _____ are independent in the population of _____. OR H_0: There is not an association between _____ and _____ in the population of _____.	H_a: _____ and _____ are not independent in the population of _____. H_a: There is an association between _____ and _____ in the population of _____.

What information can be gleaned from tests and intervals?

What if you reject the null hypothesis and have statistically significant evidence of the alternative hypothesis?

- IF the study is observational, you cannot conclude cause-and-effect; you can only conclude association ("a relationship"). However, if the random condition is met, the results can be inferred to the population that the sample came from.
- IF the study is an experiment that meets the conditions for inference, you can conclude cause-and-effect. However, unless the individuals were randomly *selected* for the experiment (unlikely), the results cannot necessarily be inferred to a larger population.
- Any value within a confidence interval (of a high level, say 90%+) is *plausible*. So, for example, if every value in an interval is positive, then it is *not* plausible that the value is 0, so you have statistically significant evidence that the value is positive.

Name That Test

Often, it's not running a hypothesis test or constructing a confidence interval that is difficult for students: it's *choosing the correct one*.

As such, the following questions are parts of AP FRQs. For each, simply <u>name the correct inference procedure</u>.

You may have seen some of these before, but it's still great practice.

The solutions are on page 54.

1.

The anterior cruciate ligament (ACL) is one of the ligaments that help stabilize the knee. Surgery is often recommended if the ACL is completely torn, and recovery time from the surgery can be lengthy. A medical center developed a new surgical procedure designed to reduce the average recovery time from the surgery. To test the effectiveness of the new procedure, a study was conducted in which 210 patients needing surgery to repair a torn ACL were randomly assigned to receive either the standard procedure or the new procedure.

(a) Based on the design of the study, would a statistically significant result allow the medical center to conclude that the new procedure causes a reduction in recovery time compared to the standard procedure, for patients similar to those in the study? Explain your answer.

(b) Summary statistics on the recovery times from the surgery are shown in the table.

Type of Procedure	Sample Size	Mean Recovery Time (days)	Standard Deviation Recovery Time (days)
Standard	110	217	34
New	100	186	29

Do the data provide convincing statistical evidence that those who receive the new procedure will have less recovery time from the surgery, on average, than those who receive the standard procedure, for patients similar to those in the study?

2.

The manager of a local fast-food restaurant is concerned about customers who ask for a water cup when placing an order but fill the cup with a soft drink from the beverage fountain instead of filling the cup with water. The manager selected a random sample of 80 customers who asked for a water cup when placing an order and found that 23 of those customers filled the cup with a soft drink from the beverage fountain.

(a) Construct and interpret a 95 percent confidence interval for the proportion of all customers who, having asked for a water cup when placing an order, will fill the cup with a soft drink from the beverage fountain.

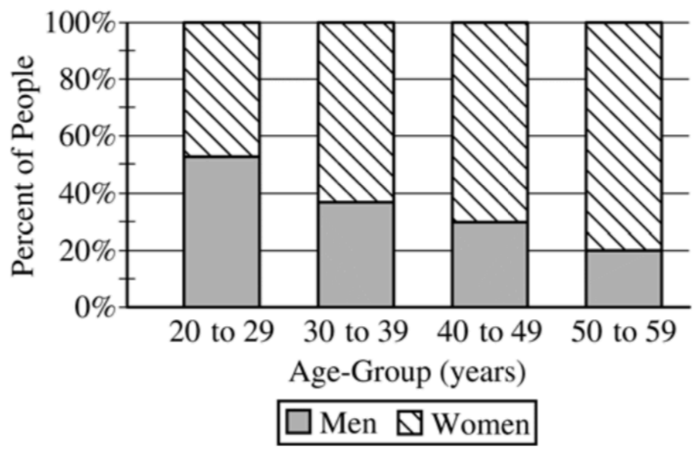
3.

A researcher conducted a medical study to investigate whether taking a low-dose aspirin reduces the chance of developing colon cancer. As part of the study, 1,000 adult volunteers were randomly assigned to one of two groups. Half of the volunteers were assigned to the experimental group that took a low-dose aspirin each day, and the other half were assigned to the control group that took a placebo each day. At the end of six years, 15 of the people who took the low-dose aspirin had developed colon cancer and 26 of the people who took the placebo had developed colon cancer. At the significance level $\alpha = 0.05$, do the data provide convincing statistical evidence that taking a low-dose aspirin each day would reduce the chance of developing colon cancer among all people similar to the volunteers?

4.

The table and the bar chart below summarize the age at diagnosis, in years, for a random sample of 207 men and women currently being treated for schizophrenia.

	Age-Group (years)				
	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
Men	53	23	9	3	88
Total	99	63	30	15	207



Do the data provide convincing statistical evidence of an association between age-group and gender in the diagnosis of schizophrenia?

5.

A polling agency showed the following two statements to a random sample of 1,048 adults in the United States.

<u>Environment statement:</u> Protection of the environment should be given priority over economic growth.

<u>Economy statement:</u> Economic growth should be given priority over protection of the environment.

The order in which the statements were shown was randomly selected for each person in the sample. After reading the statements, each person was asked to choose the statement that was most consistent with his or her opinion. The results are shown in the table.

	Environment Statement	Economy Statement	No Preference
Percent of sample	58%	37%	5%

(a) Assume the conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the proportion of all adults in the United States who would have chosen the <u>economy statement</u>.

6.

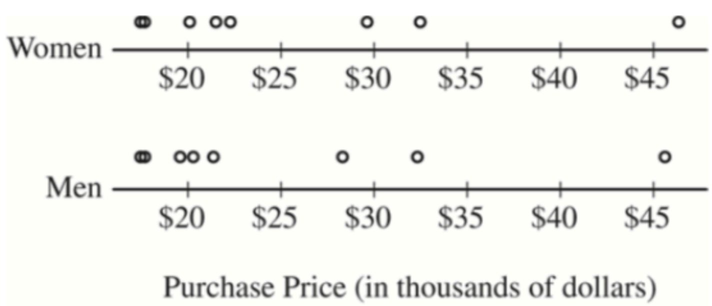
A researcher conducted a study to investigate whether local car dealers tend to charge women more than men for the same car model. Using information from the county tax collector's records, the researcher randomly selected one man and one woman from among everyone who had purchased the same model of an identically equipped car from the same dealer. The process was repeated for a total of 8 randomly selected car models.

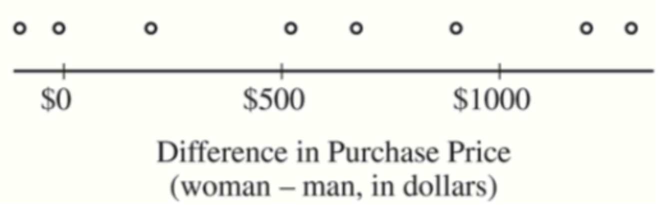
The purchase prices and the differences (woman – man) are shown in the table below. Summary statistics are also shown.

Car model	1	2	3	4	5	6	7	8
Women	$20,100	$17,400	$22,300	$32,500	$17,710	$21,500	$29,600	$46,300
Men	$19,580	$17,500	$21,400	$32,300	$17,720	$20,300	$28,300	$45,630
Difference	$520	–$100	$900	$200	–$10	$1,200	$1,300	$670

	Mean	Standard Deviation
Women	$25,926.25	$9,846.61
Men	$25,341.25	$9,728.60
Difference	$585.00	$530.71

Dotplots of the data and the differences are shown below.





Do the data provide convincing evidence that, on average, women pay more than men in the county for the same car model?

7.

A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded “yes.” In December 2007, 622 of 1,020 randomly selected adults in the United States had responded “yes” to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond “yes” to the question changed from December 2007 to December 2008 ?

8.

An environmental group conducted a study to determine whether crows in a certain region were ingesting food containing unhealthy levels of lead. A biologist classified lead levels greater than 6.0 parts per million (ppm) as unhealthy. The lead levels of a random sample of 23 crows in the region were measured and recorded. The data are shown in the stemplot below.

Lead Levels

2	8
3	0
3	5 8 8
4	1 1 2
4	6 8 8
5	0 1 2 2 3 4
5	9 9
6	3 4
6	6 8

Key: $2|8 = 2.8$ ppm

(a) What proportion of crows in the sample had lead levels that are classified by the biologist as unhealthy?

(b) The mean lead level of the 23 crows in the sample was 4.90 ppm and the standard deviation was 1.12 ppm. Construct and interpret a 95 percent confidence interval for the mean lead level of crows in the region.

9.

During a flu vaccine shortage in the United States, it was believed that 45 percent of vaccine-eligible people received flu vaccine. The results of a survey given to a random sample of 2,350 vaccine-eligible people indicated that 978 of the 2,350 people had received flu vaccine.

(a) Construct a 99 percent confidence interval for the proportion of vaccine-eligible people who had received flu vaccine. Use your confidence interval to comment on the belief that 45 percent of the vaccine-eligible people had received flu vaccine.

10.

Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

11.

Two treatments, A and B, showed promise for treating a potentially fatal disease. A randomized experiment was conducted to determine whether there is a significant difference in the survival rate between patients who receive treatment A and those who receive treatment B. Of 154 patients who received treatment A, 38 survived for at least 15 years, whereas 16 of the 164 patients who received treatment B survived at least 15 years.

(b) The conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the difference between the proportion of the population who would survive at least 15 years if given treatment A and the proportion of the population who would survive at least 15 years if given treatment B.

12.

A French study was conducted in the 1990s to compare the effectiveness of using an instrument called a cardiopump with the effectiveness of using traditional cardiopulmonary resuscitation (CPR) in saving lives of heart attack victims. Heart attack patients in participating cities were treated with either a cardiopump or CPR, depending on whether the individual's heart attack occurred on an even-numbered or an odd-numbered day of the month. Before the start of the study, a coin was tossed to determine which treatment, a cardiopump or CPR, was given on the even-numbered days. The other treatment was given on the odd-numbered days. In total, 754 patients were treated with a cardiopump, and 37 survived at least one year; while 746 patients were treated with CPR, and 15 survived at least one year.

(a) The conditions for inference are satisfied in the study. State the conditions and indicate how they are satisfied.

(b) Perform a statistical test to determine whether the survival rate for patients treated with a cardiopump is significantly higher than the survival rate for patients treated with CPR.

13.

High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20 Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40 Standard Deviation of Reductions: 9.40

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

14.

Product advertisers studied the effects of television ads on children's choices for two new snacks. The advertisers used two 30-second television ads in an experiment. One ad was for a new sugary snack called Choco-Zuties, and the other ad was for a new healthy snack called Apple-Zuties.

For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad nor the Apple-Zuties ad.

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices.

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties
A	Choco-Zuties only	21	4
B	Apple-Zuties only	13	12
C	Neither	22	3

(a) Do the data provide convincing statistical evidence that there is an association between type of ad and children's choice of snack among all children similar to those who participated in the experiment?

15.

A bottle-filling machine is set to dispense 12.1 fluid ounces into juice bottles. To ensure that the machine is filling accurately, every hour a worker randomly selects four bottles filled by the machine during the past hour and measures the contents. If there is convincing evidence that the mean amount of juice dispensed is different from 12.1 ounces or if there is convincing evidence that the standard deviation is greater than 0.05 ounce, the machine is shut down for recalibration. It can be assumed that the amount of juice that is dispensed into bottles is normally distributed.

During one hour, the mean number of fluid ounces of four randomly selected bottles was 12.05 and the standard deviation was 0.085 ounce.

(a) Perform a test of significance to determine whether the mean amount of juice dispensed is different from 12.1 fluid ounces. Assume the conditions for inference are met.

16.

One of the two fire stations in a certain town responds to calls in the northern half of the town, and the other fire station responds to calls in the southern half of the town. One of the town council members believes that the two fire stations have different mean response times. Response time is measured by the difference between the time an emergency call comes into the fire station and the time the first fire truck arrives at the scene of the fire.

Data were collected to investigate whether the council member's belief is correct. A random sample of 50 calls selected from the northern fire station had a mean response time of 4.3 minutes with a standard deviation of 3.7 minutes. A random sample of 50 calls selected from the southern fire station had a mean response time of 5.3 minutes with a standard deviation of 3.2 minutes.

(a) Construct and interpret a 95 percent confidence interval for the difference in mean response times between the two fire stations.

(b) Does the confidence interval in part (a) support the council member's belief that the two fire stations have different mean response times? Explain.

17.

Investigators at the U.S. Department of Agriculture wished to compare methods of determining the level of *E. coli* bacteria contamination in beef. Two different methods (A and B) of determining the level of contamination were used on each of ten randomly selected specimens of a certain type of beef. The data obtained, in millimicrobes/liter of ground beef, for each of the methods are shown in the table below.

		Specimen									
		1	2	3	4	5	6	7	8	9	10
Method	A	22.7	23.6	24.0	27.1	27.4	27.8	34.4	35.2	40.4	46.8
	B	23.0	23.1	23.7	26.5	26.6	27.1	33.2	35.0	40.5	47.8

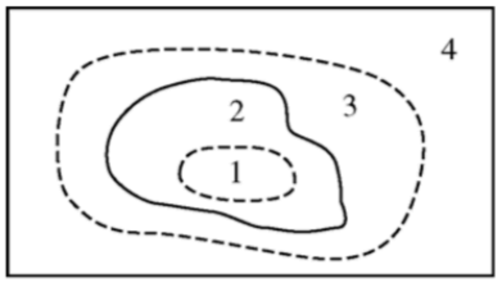
Is there a significant difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef? Provide a statistical justification to support your answer.

18.

A study was conducted to determine where moose are found in a region containing a large burned area. A map of the study area was partitioned into the following four habitat types.

(1) Inside the burned area, not near the edge of the burned area,
(2) Inside the burned area, near the edge,
(3) Outside the burned area, near the edge, and
(4) Outside the burned area, not near the edge.

The figure below shows these four habitat types.



Note: Figure not drawn to scale.

The proportion of total acreage in each of the habitat types was determined for the study area. Using an aerial survey, moose locations were observed and classified into one of the four habitat types. The results are given in the table below.

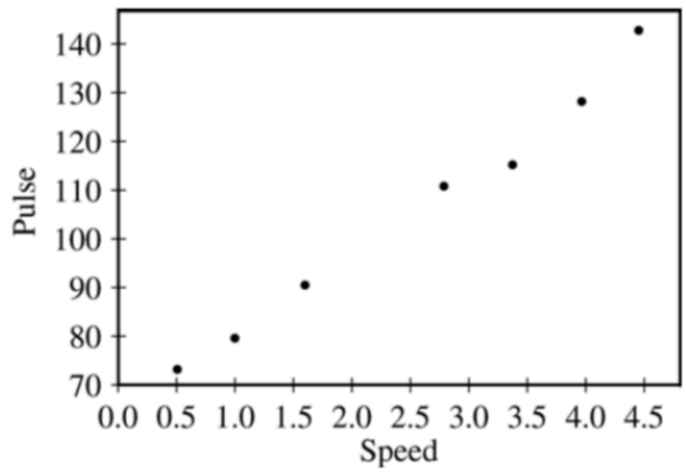
Habitat Type	Proportion of Total Acreage	Number of Moose Observed
1	0.340	25
2	0.101	22
3	0.104	30
4	0.455	40
Total	1.000	117

(a) The researchers who are conducting the study expect the number of moose observed in a habitat type to be proportional to the amount of acreage of that type of habitat. Are the data consistent with this expectation? Conduct an appropriate statistical test to support your conclusion. Assume the conditions for inference are met.

(b) Relative to the proportion of total acreage, which habitat types did the moose seem to prefer? Explain.

19.

John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087 R-Sq = 98.7% R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(c) John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.

20.

A large university provides housing for 10 percent of its graduate students to live on campus. The university's housing office thinks that the percentage of graduate students looking for housing on campus may be more than 10 percent. The housing office decides to survey a random sample of graduate students, and 62 of the 481 respondents say that they are looking for housing on campus.

(a) On the basis of the survey data, would you recommend that the housing office consider increasing the amount of housing on campus available to graduate students? Give appropriate evidence to support your recommendation.

21.

A researcher believes that treating seeds with certain additives before planting can enhance the growth of plants. An experiment to investigate this is conducted in a greenhouse. From a large number of Roma tomato seeds, 24 seeds are randomly chosen and 2 are assigned to each of 12 containers. One of the 2 seeds is randomly selected and treated with the additive. The other seed serves as a control. Both seeds are then planted in the same container. The growth, in centimeters, of each of the 24 plants is measured after 30 days. These data were used to generate the partial computer output shown below. Graphical displays indicate that the assumption of normality is not unreasonable.

	N	Mean	StDev	SE Mean
Control	12	15.989	1.098	0.317
Treatment	12	18.004	1.175	0.339
Difference	12	– 2.015	1.163	0.336

(a) Construct a confidence interval for the mean difference in growth, in centimeters, of the plants from the untreated and treated seeds. Be sure to interpret this interval.

(b) Based only on the confidence interval in part (a), is there sufficient evidence to conclude that there is a significant mean difference in growth of the plants from untreated seeds and the plants from treated seeds? Justify your conclusion.

22.

The Colorado Rocky Mountain Rescue Service wishes to study the behavior of lost hikers. If more were known about the direction in which lost hikers tend to walk, then more effective search strategies could be devised. Two hundred hikers selected at random from those applying for hiking permits are asked whether they would head uphill, downhill, or remain in the same place if they became lost while hiking. Each hiker in the sample was also classified according to whether he or she was an experienced or novice hiker. The resulting data are summarized in the following table.

	Direction		
	Uphill	Downhill	Remain in Same Place
Novice	20	50	50
Experienced	10	30	40

Do these data provide convincing evidence of an association between the level of hiking expertise and the direction the hiker would head if lost?

Give appropriate statistical evidence to support your conclusion.

23.

Baby walkers are seats hanging from frames that allow babies to sit upright with their legs dangling and feet touching the floor. Walkers have wheels on their legs that allow the infant to propel the walker around the house long before he or she can walk or even crawl. Typically, babies use walkers between the ages of 4 months and 11 months.

Because most walkers have tray tables in front that block babies' views of their feet, child psychologists have begun to question whether walkers affect infants' cognitive development. One study compared mental skills of a random sample of those who used walkers with a random sample of those who never used walkers. Mental skill scores averaged 113 for 54 babies who used walkers (standard deviation of 12) and 123 for 55 babies who did not use walkers (standard deviation of 15).

(a) Is there evidence that the mean mental skill score of babies who use walkers is different from the mean mental skill score of babies who do not use walkers? Explain your answer.

24.

A growing number of employers are trying to hold down the costs that they pay for medical insurance for their employees. As part of this effort, many medical insurance companies are now requiring clients to use generic brand medicines when filling prescriptions. An independent consumer advocacy group wanted to determine if there was a difference, in milligrams, in the amount of active ingredient between a certain "name" brand drug and its generic counterpart. Pharmacies may store drugs under different conditions. Therefore, the consumer group randomly selected ten different pharmacies in a large city and filled two prescriptions at each of these pharmacies, one for the "name" brand and the other for the generic brand of the drug. The consumer group's laboratory then tested a randomly selected pill from each prescription to determine the amount of active ingredient in the pill. The results are given in the following table.

ACTIVE INGREDIENT
(in milligrams)

Pharmacy	1	2	3	4	5	6	7	8	9	10
Name brand	245	244	240	250	243	246	246	246	247	250
Generic brand	246	240	235	237	243	239	241	238	238	234

Based on these results, what should the consumer group's laboratory report about the difference in the active ingredient in the two brands of pills? Give appropriate statistical evidence to support your response.

25.

Successfully Completed Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	23.514	1.684	13.95	0.000
GPA	−2.7555	0.4668	−5.90	0.000
S = 0.5658	R-Sq = 76.0%			

(b) For the students who successfully completed the Ph.D. program, is there a significant relationship between GPA and mean number of credit hours per semester?

Give a statistical justification to support your response.

26.

Each person in a random sample of 1,026 adults in the United States was asked the following question.

"Based on what you know about the Social Security system today, what would you like Congress and the President to do during this next year?"

The response choices and the percentages selecting them are shown below.

Completely overhaul the system	19%
Make some major changes	39%
Make some minor adjustments	30%
Leave the system the way it is now	11%
No opinion	1%

(a) Find a 95% confidence interval for the proportion of all United States adults who would respond "Make some major changes" to the question. Give an interpretation of the confidence interval and give an interpretation of the confidence level.

27.

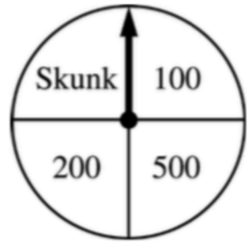
The principal at Crest Middle School, which enrolls only sixth-grade students and seventh-grade students, is interested in determining how much time students at that school spend on homework each night. The table below shows the mean and standard deviation of the amount of time spent on homework each night (in minutes) for a random sample of 20 sixth-grade students and a separate random sample of 20 seventh-grade students at this school.

	Mean	Standard Deviation
Sixth-grade students	27.3	10.8
Seventh-grade students	47.0	12.4

Based on dotplots of these data, it is not unreasonable to assume that the distribution of times for each grade were approximately normally distributed.

(a) Estimate the difference in mean times spent on homework for all sixth- and seventh-grade students in this school using an interval. Be sure to interpret your interval.

28.



Contestants on a game show spin a wheel like the one shown in the figure above. Each of the four outcomes on this wheel is equally likely and outcomes are independent from one spin to the next.

- The contestant spins the wheel.
- If the result is a skunk, no money is won and the contestant's turn is finished.
- If the result is a number, the corresponding amount in dollars is won. The contestant can then stop with those winnings or can choose to spin again, and his or her turn continues.
- If the contestant spins again and the result is a skunk, all of the money earned on that turn is lost and the turn ends.
- The contestant may continue adding to his or her winnings until he or she chooses to stop or until a spin results in a skunk.

(c) A contestant who lost at this game alleges that the wheel is not fair. In order to check on the fairness of the wheel, the data in the table below were collected for 100 spins of this wheel.

Result	Skunk	\$100	\$200	\$500
Frequency	33	21	20	26

Based on these data, can you conclude that the four outcomes on this wheel are not equally likely? Give appropriate statistical evidence to support your answer.

29.

Some boxes of a certain brand of breakfast cereal include a voucher for a free video rental inside the box. The company that makes the cereal claims that a voucher can be found in 20 percent of the boxes. However, based on their experiences eating this cereal at home, a group of students believes that the proportion of boxes with vouchers is less than 0.2. This group of students purchased 65 boxes of the cereal to investigate the company's claim. The students found a total of 11 vouchers for free video rentals in the 65 boxes.

Suppose it is reasonable to assume that the 65 boxes purchased by the students are a random sample of all boxes of this cereal. Based on this sample, is there support for the students' belief that the proportion of boxes with vouchers is less than 0.2 ? Provide statistical evidence to support your answer.

30.

In order to monitor the populations of birds of a particular species on two islands, the following procedure was implemented.

Researchers captured an initial sample of 200 birds of the species on Island A; they attached leg bands to each of the birds, and then released the birds. Similarly, a sample of 250 birds of the same species on Island B was captured, banded, and released. Sufficient time was allowed for the birds to return to their normal routine and location.

Subsequent samples of birds of the species of interest were then taken from each island. The number of birds captured and the number of birds with leg bands were recorded. The results are summarized in the following table.

	Island A	Island B
Number Captured in Subsequent Sample	180	220
Number with Leg Bands in Subsequent Sample	12	35

Assume that both the initial sample and the subsequent samples that were taken on each island can be regarded as random samples from the population of birds of this species.

(a) Do the data from the subsequent samples indicate that there is a difference in proportions of the banded birds on these two islands? Give statistical evidence to support your answer.

ANSWERS

1. 2-sample t-test
2. 1-proportion z-interval
3. 2-proportion z-test
4. χ^2 independence
5. 1-proportion z-test
6. Paired t-test
7. 2-proportion z-test
8. 1-sample t-interval
9. 1-proportion z-interval
10. t-test for slope
11. 2-proportion z-interval
12. 2-proportion z-test
13. Two-sample t-test
14. χ^2 homogeneity
15. 1-sample t-test
16. 2-sample t-interval
17. Paired t-test
18. χ^2 goodness of fit
19. t-interval for slope
20. 1-proportion z-test
21. Paired t-interval
22. χ^2 independence
23. 2-sample t-test
24. Paired t-test
25. t-test for slope
26. 1-proportion z-interval
27. 2-sample t-interval
28. χ^2 goodness of fit
29. 1-proportion z-test
30. 2-proportion z-test

AP Exam Review – 54 Multiple Choice Questions

Answers are on page 64.

1. Which of the following is true?
   I. A simple random sample is any sampling technique where each individual in the population has the same chance of being selected.
   II. A simple random sample is a sample where every sample of the same size has the same chance of being selected.
   III. From a simple random sample of a population of size 10, there are 90 equally likely possible samples of size 2 if sampling without replacement.

(A) I only (B) II only (C) III only (D) II and III (E) I, II, and III

2. A statistic student wishes to test the strength of various brands of paper towel. He chooses 5 brands and selects 6 towels from each brand. He randomly selects a towel and places it in an embroidery loop. Exactly 10 ml of water and a large weight are placed in the center of the towel. The time it takes for the towel to break is recorded. He repeats this for the remaining towels. In this study, the explanatory variable is the
(A) amount of time it takes for the towel to break
(B) 10 ml of water and the large weight
(C) brand of paper towel
(D) large weight
(E) thirty paper towels used in the experiment

3. A student is interested in the effects of different walking styles on heart rate. She decides to use 30 volunteers from her school for her experiment. All 30 participants find their at-rest pulse rates. Each participant will walk twice for 10 minutes, once using a fast pace but with no arm movement and again using a fast pace, but with an exaggerated arm movement style. The experimenter throws a coin to determine which style each participant will walk first. All participants get sufficient rest between walks to let their pulse rates return to normal. The student then compares the increased pulse rate based on the walk with no arm movement to the increased pulse rate based on the walk with exaggerated arm movement for each student. Which of the following statements is true?
(A) This is an observational study and not an experiment because the 30 individuals were not randomly selected.
(B) Observations in this study are independent of one another.
(C) Blocking is used in this study to reduce difference in increased pulse rates among individual students.
(D) Because subjects were not randomly assigned to a control group, the design of the experiment is flawed.
(E) This is an example of a completely randomized experiment.

4. A sample of 99 distances has a mean of 24 feet and a median of 24.5 feet. Unfortunately, it has just been discovered that the maximum value in the distribution, which was erroneously recorded as 40, actually had a value of 50. If we make this correction to the data, then
(A) the mean remains the same, but the median is increased.
(B) the mean and median remain the same.
(C) the median remains the same, but the mean is increased.
(D) the mean and median are both increased.
(E) we do not know how the mean and median are affected without further calculations.

5. Mr. Yates picked up a dozen items in the grocery store with a mean cost of $3.25. Then he added an apple pie for $6.50. The new mean for all 13 items is

(A) $3.00	(B) $3.50	(C) $3.75	(D) $4.88	(E) None of the above

6. The five-number summary for scores on a statistics exam is 11, 35, 61, 70, 79. In all, 380 students took the test. About how many had scores between 35 and 61?

(A) 26	(B) 76	(C) 95	(D) 190	(E) None of these

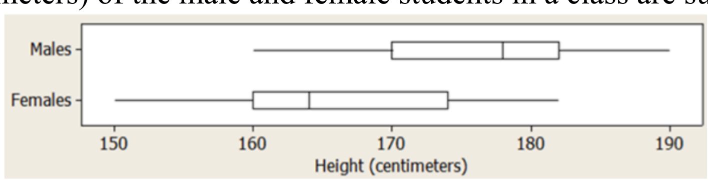
7. According to the U.S. Bureau of Labor Statistics, the monthly percentage change in the number of jobs in a certain state for the twelve months of 2011 had a mean of 0.08% and a standard deviation of 1.70%. From this information we can conclude that
(A) the largest monthly change was 1.78%.
(B) the distribution of monthly changes is strongly skewed to the right.
(C) most of the monthly changes were negative.
(D) the magnitude of the monthly deviations from the mean change averaged about 1.70%.
(E) a mistake has been made. It makes no sense for the standard deviation to be greater than the mean.

8. When testing water for chemical impurities, results are often reported as bdl, that is, below detection limit. The following are the measurements of the amount of lead in a series of water samples taken from inner-city households (in parts per million)

5, 7, 12, bdl, 10, 8, bdl, 20, 6

Which of the following statements can we be sure is true?
(A) The mean lead level in the water is about 10 ppm.
(B) The mean lead level in the water is about 9 ppm.
(C) The median lead level in the water is 7 ppm.
(D) The median lead level in the water is 8 ppm.
(E) Neither the mean nor the median can be computed because some values are unknown.

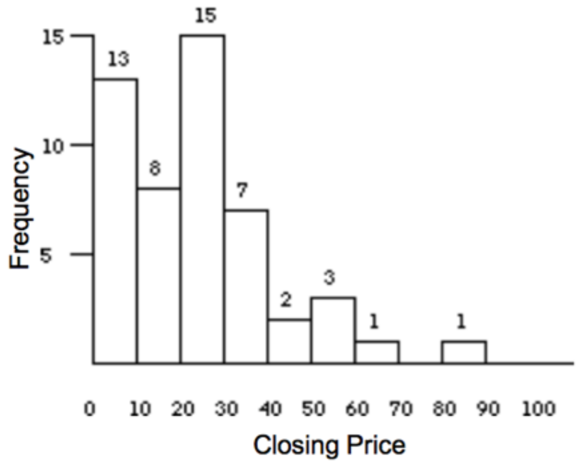
9. The heights (in centimeters) of the male and female students in a class are summarized in the following boxplots:



Which of the following conclusions can be drawn from this graph?
(A) About 50% of the male students have heights between 170 and 178 centimeters.
(B) The median height of male students is about 163 centimeters.
(C) The mean height of male students is about 178 centimeters.
(D) For female students, the mean height is lower than the median height.
(E) About 25% of the male students are taller than the tallest female student.

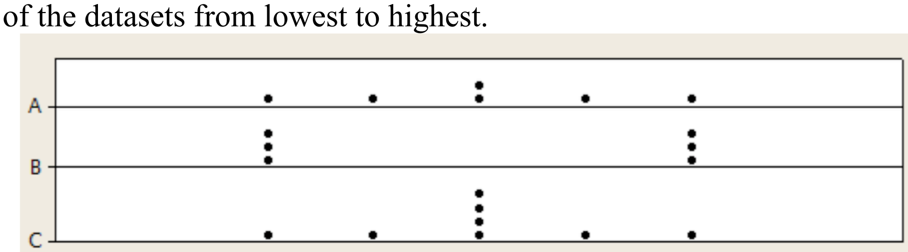
10. A small company that prints custom t-shirts has 6 employees, one of whom is the owner and manager. Suppose the owner makes $120,000 per year and the other employees make between $40,000 and $50,000 per year. One day, the owner decides to give himself a $30,000 raise. Which of the following describes how the company's mean and median salaries would change?
(A) The mean and median would both increase by $5,000.
(B) The mean would increase by $5,000 and the median would not change.
(C) The mean would increase by $6,000 and the median would not change.
(D) The median would increase by $6,000 and the mean would not change.
(E) The mean would increase by $6,000, but we cannot determine the change in the median without more information.

11. The following is a histogram showing the actual frequency of the closing prices of a particular stock on the New York Stock Exchange over a 50-day period. The class that contains the third quartile is
(A) 10-20
(B) 20-30
(C) 30-40
(D) 40-50
(E) 50-60



12. For the data in the previous problem, which measures of center and spread would be most appropriate to use?
(A) Mean and standard deviation
(B) Mean and interquartile range
(C) Mean and range
(D) Median and interquartile range
(E) Median and standard deviation

13. Below are dotplots for three small data sets: A, B, and C. Without performing any calculations, rank the standard deviations of the datasets from lowest to highest.



14. If $P(A) = 0.6, P(B) = 0.2,$ and $P(A \cap B) = 0.10,$ which of the following must be true?
(A) A and B are independent and mutually exclusive.
(B) A and B are not independent, but they are mutually exclusive.
(C) A and B are independent, but they are not mutually exclusive.
(D) A and B are neither independent nor mutually exclusive.
(E) A and B are mutually exclusive, but there is not enough information to determine if they are independent.

15. Suppose the heights of 10-year-old boys is normally distributed with a mean of 55 inches and a standard deviation of 2.4 inches. How tall would a 10-year-old boy have to be so that only 20% of 10-year-old boys are taller than him?
(A) 52.98 (B) 54.16 (C) 55.84 (D) 57.02 (E) 57.44

16. Eric has to be at school by 9:00 a.m. every day. On any given day, the amount of time it takes Eric to get to school is approximately normally distributed with a mean of 15 minutes and a standard deviation of 3 minutes. If Eric leaves at 8:40 on a day, what is the probability he will be late?
(A) 0.0228 (B) 0.0478 (C) 0.0685 (D) 0.9522 (E) 0.9773

17. A convenience store sells one type of drink with either "You win a free drink!" or "Try again!" printed on each drink's bottle cap. The drink advertises that 1 in 5 win a free drink. Six friends walk into the store and each buy a drink. What is the probability that 2 of them win a free drink?
(A) 0.0164 (B) 0.2458 (C) 0.5316 (D) 0.6554 (E) 0.9011

18. Suppose the number of candy bars sold by a machine in a hospital on a randomly selected day is normally distributed with a mean of 27 and a standard deviation of 5.5. The machine costs \$7 to operate per day and each candy bar costs \$1.75 (yeah, expensive). What is the probability that the profits from the machine will be more than \$50?
(A) 0.0001 (B) 0.0778 (C) 0.1024 (D) 0.1225 (E) 0.1555

19. The distribution of diameters of Granny Smith apples is normal with a standard deviation of 1.1 centimeters. If a Granny Smith apple has a diameter that puts it at the 80th percentile of all Granny Smith apples, how much larger is its diameter than average?
(A) 0.765 cm (B) 0.842 cm (C) 0.926 cm (D) 1.1 cm (E) 1.28 cm

20. On the SAT prior to 2016, each question had five answer choices, only one of which is correct. A correct answer was worth 1 point, an incorrect answer was worth $-1/4$ of a point, and a blank answer was worth 0 points. If a student randomly guessed on 10 questions on the old SAT, what is the probability they would earn at least 1 point?
(A) 0.033 (B) 0.088 (C) 0.121 (D) 0.244 (E) 0.322

21. A hypothesis test of $H_0: \mu = 2$ versus $H_a: \mu \neq 2$ for a sample of size 15 results in a test statistic of -1.538. What is the P-value for this test?
(A) 0.062 (B) 0.0703 (C) 0.124 (D) 0.146 (E) 0.154

22. A zoo wanted to estimate the difference in the average weight of its crocodiles and the average weight of its alligators. The zoo has 8 crocodiles and 14 alligators. Which of the following is the most appropriate critical value to use for a 98% confidence interval for the difference in average weights?
(A) 2.28 (B) 2.52 (C) 2.65 (D) 3.00 (E) 3.24

23. Which of the following is true of Q1 in *any* normal distribution?
(A) It is half the mean.
(B) It is half the median.
(C) It is 0.5 standard deviations below the mean.
(D) It is 0.674 standard deviations below the mean.
(E) It is 25% smaller than the median.

24. A certain distributor of brake pads has 3 plants, A, B, and C. 20% of brake pads are made at Plant A, 50% are made at Plant B, and 30% are made at Plant C. At plants A, B, and C, the percentage of parts that are defective are 5%, 7%, and 8%, respectively. If a part is found to be defective, what is the probability that it came from Plant C?
(A) 0.024 (B) 0.069 (C) 0.348 (D) 0.507 (E)

25. Suppose that A and B are events such that $P(A) = 0.3$ and $P(B) = 0.5$. If $P(A \cup B) = 0.8$, which of the following is true?
(A) A and B are independent and mutually exclusive.
(B) A and B are independent, but not mutually exclusive.
(C) A and B are not independent, but they are mutually exclusive.
(D) A and B are neither independent nor mutually exclusive.
(E) It is impossible to determine without additional information.

26. The distribution of test scores on a calculus departmental final at a certain university is normally distributed with a mean of 70 and a standard deviation of 7.5 points. Which of the following is equal to the 90th percentile of scores on this final?
(A) $70 + 0.9(7.5)$ (B) $70 + 1.1(7.5)$ (C) $70 + 1.282(7.5)$ (D) $70 + 1.645(7.5)$ (E) 70.9

27. Suppose a certain company's fruit basket is made of 4 green apples and 5 oranges. The distribution of weights of green apples is normal with a mean of 180 grams (g) and a standard deviation of 14 g, while the distribution of weights of oranges is normal with a mean of 155 g and a standard deviation of 10 g. An empty basket weighs 240 grams. Assuming a basket is a random collection of 4 apples and 5 oranges, what is the standard deviation of the distribution of weights of the company's fruit baskets?
(A) 10.3 grams (B) 14.8 grams (C) 31.6 grams (D) 35.8 grams (E) 36.9 grams

28. The distribution of yearly salaries at a certain large corporation is heavily right skewed with a mean of \$40,000 and a standard deviation of \$4,000. A random sample of 16 employees will be selected and the mean salary of the employees will be calculated. Which of the following most correctly describes the sampling distribution of the sample mean of the 16 employees?
(A) Slightly right skewed, mean = \$40,000, standard deviation = \$250
(B) Slightly right skewed, mean = \$40,000, standard deviation = \$1,000
(C) Approximately normal, mean = \$40,000, standard deviation = \$250
(D) Approximately normal, mean = \$40,000, standard deviation = \$1,000
(E) Approximately normal, mean = \$40,000, standard deviation = \$4,000

29. A government agency in charge of state-provided scholarships is looking to estimate the mean amount of money freshman spend on textbooks at a certain university. They randomly sample 100 freshmen and calculate the mean amount of money the 100 students spend on textbooks. Which of the following is NOT a reason for *randomly* selecting the 100 students?
(A) So that Central Limit Theorem applies
(B) So that the sample mean can be inferred to the population of all students at school
(C) So that the amount spent by each student is independent of one another
(D) So that sample mean is unbiased estimator
(E) So that the sample is reasonably representative of the population

30. Suppose that 4% of all parts produced at a certain factory are defective, and the rest are working properly. A quality control manager will randomly sample 100 parts and examine what proportion of the parts are working properly. Which of the following is the shape of the sampling distribution of the sample proportion of parts that are working properly?
(A) Left skewed (B) Right skewed (C) Uniform (D) Bimodal (E) Approx. normal

31. A breeder of rabbits has a new litter of 30 newborn rabbits. The average weight of the rabbits is 5 ounces and the standard deviation is 3.1 ounces. Which of the following is the shape of the distribution of the weights of the 30 rabbits?
(A) Left skewed (B) Right skewed (C) Approx. normal (D) Uniform (E) Bimodal

32. When randomly sampling with sample size n and calculating a sample mean to estimate a population mean, which of the following correctly describes the statement of the Central Limit Theorem?
(A) As sample size increases, sample means get closer to true mean.
(B) When random sampling, the sample mean is an unbiased estimator.
(C) For large sample sizes, the sampling distribution of the sample mean is approximately normal.
(D) For large n, individuals are reasonably independent.
(E) For large n, a t-critical value is not necessary.

33. Twenty-five students at Morris Knolls High School and their SAT scores are recorded. A 95% confidence interval for the mean SAT score is 900 to 1100. Which of the following statement gives a valid interpretation of this interval?
(A) 95% of the 25 students have a mean score between 900 and 1100.
(B) 95% of the population of all students at Morris Knolls have a score between 900 and 1100.
(C) If this procedure were repeated many times, 95% of the resulting confidence intervals would contain the true mean SAT score at Morris Knolls.
(D) If this procedure were repeated many times ,95% of the sample means would be between 900 and 1100.
(E) If 1000 samples were taken and a 95% confidence interval was computed, approximately 50 of the intervals will capture the true mean SAT score at Morris Knolls.

34. A survey was conducted to determine the percentage of college freshmen that planned to take a math course while in college. The results were given as 75% with a margin of error of 4%. What does the margin of error of 4% mean?
(A) The percentage of the population that was surveyed was between 71% and 79%.
(B) 4% of the population was not included in the survey.
(C) There is a 4% chance that a randomly selected student would say they are 75% confident they will take a math class in college.
(D) There is a 96% chance that 75% is the true percentage of college freshmen that plan to take a math course while in college.
(E) The difference between the sample proportion and the population proportion is likely to be less than 4%.

35. Consider a significance test of the hypotheses $H_0: \mu = 0$ versus $H_0: \mu \neq 0$ at significance level α. Gather a sample mean from a sample of size n and construct a confidence interval about the sample mean that will reach the same conclusion as the significance test. Which of the following will would most likely result in the widest confidence interval for estimating μ?
(A) $n = 100, \alpha = .01$
(B) $n = 50, \alpha = .01$
(C) $n = 100, \alpha = .05$
(D) $n = 50, \alpha = 0.05$
(E) $n = 100, \alpha = 0.10$

36. A health fitness research group wishes to estimate the mean amount of time (in hours) that members of a fitness center spend each week exercising at the center. They want to estimate the mean within a margin of error of 0.5 hours with a 95% level of confidence. Previous data suggests that the standard deviation of the time of hours of members at this fitness center is 2.2 hours. Which of the following is the smallest sample size that meets these criteria?
(A) 60 (B) 75 (C) 90 (D) 180 (E) 190

37. A researcher wishes to use a 99% confidence interval to estimate the proportion of Americans who have visited an entertainment theme park near Orlando within the last five years. The researcher wishes to choose a size that will ensure a margin of error not to exceed 5 percent. Which of the following is the smallest sample size that will guarantee the margin of error will be less than 5 percent?
(A) 40 (B) 200 (C) 400 (D) 600 (E) 700

38. Research has shown that the standard deviation of the washers produced by a certain company is approximately 0.035 mm. Researchers wish to create a 90% confidence interval to estimate the average diameter of the washers such that the interval has a margin of error of no more than 0.005 mm. What is the minimum sample size that company can use?
(A) 11 (B) 12 (C) 122 (D) 133 (E) 145

39. An association wishes to design and conduct a poll to determine the proportion of Americans who oppose a law limiting the sale of handguns. If they wish to be 98% confident that their sample results differ from the true population proportion by no more than 0.07, which of the following choices is the smallest sample sizes they should use?

(A) 156 (B) 215 (C) 250 (D) 300 (E) 500

40. A marine biology professor would like to assess if there is a relationship between how long her students study for her final exam and their scores on the exam. She has all of her students record how long they spend studying for the final exam, and once the exam is over, she randomly samples 20 students and records how long they reported studying and their final score. She calculates a least-squares regression line of $\hat{y} = -12.73 + 4.61x$. When she runs a t-test for the sample slope 4.61, she gets a P-value of 0.054. At a significance level of $\alpha = 0.05$, what conclusion should the professor make?

(A) Find convincing evidence of a relationship between hours and score
(B) Find convincing evidence of no relationship between hours and score
(C) Not find convincing evidence of a relationship between hours and score
(D) Not find convincing evidence of no relationship between hours and score
(E) No conclusion can be made because the sample size was only 20.

41. A quality control manager randomly sampled 100 bags of rice from the supply line to assess the average number of "brokens," or pieces of rice that are broken, in the bags. The manager then creates a 90% confidence interval to estimate the mean amount of brokens in the entire supply line. If the manager were to have selected 25 bags of rice instead of 100 and then created a 90% confidence interval, approximately how would the width of the new interval compare to the original interval?

(A) One fourth of the width
(B) One half of the width
(C) Twice the width
(D) Four times the width
(E) Sixteen times the width

42. A two-sample t-test was conducted with the following hypotheses $H_0: \mu = 4$ versus $H_a: \mu \neq 4$. A sample size of 100 was used with a significance level of $\alpha = 0.05$. The power of the test against the alternative mean $\mu = 5$ was 0.31. Which of the following would increase the power of the test <u>the most</u>?

(A) Change the sample size to 50.
(B) Change the significance level to 0.01.
(C) Use an alternative mean of 1.
(D) Use an alternative mean of 6.
(E) Use an alternative mean of 7.

43. Researchers at a zoo believe that male okapis (relatives of the giraffe) are more aggressive than male zebras, and therefore that there are a higher percentage of okapis in the wild are male than percentage of zebras in the wild that are male. They randomly sample okapis and zebras in the same geographic region and their data is presented below:

	Number of animals	Number of male animals
Okapi	85	50
Zebra	125	58

At a significance level of 0.01, what conclusion should the researchers make?

(A) Because $P < 0.01$, reject the null hypothesis.
(B) Because $0.01 < P < 0.05$, reject the null hypothesis.
(C) Because $0.01 < P < 0.05$, fail to reject the null hypothesis.
(D) Because $P > 0.05$, fail to reject the null hypothesis.
(E) Because $P > 0.05$, reject the null hypothesis.

44. A school system is concerned about the amount of bacteria of students have on their hands with they come in from recess. They take a random sample of 10 students from their magnet elementary school and 10 students from one of their non-magnet elementary schools and measure the amount of bacteria on the students' hands. The amounts, in parts per million, are shown in the table below:

Magnet	3.2	5.4	7.1	4.6	3.8	5.1	4.0	4.2	6.1	5.4
Non-magnet	7.1	8.2	5.1	3.2	5.6	6.6	5.8	4.3	7.0	9.4

What type of test should be used to compare the average amount of bacteria between students at the schools?
(A) Two-proportion z-test
(B) One-sample t-test
(C) Paired t-test
(D) Chi-squared test of independence
(E) Two-sample t-test

45. Four hundred students at a university are selected to participate in a study assessing the percentage of time they spend on social media. Students are to record the percentage of their time spent on social media for one week, after which they view a twenty-minute video on the effects of screen time on the brain and eyes. After the video, students will then again record the percentage of their time spent on social media for another week. To estimate the difference in the percentage of time spent using social media before and after viewing the video, which type of statistical test should be used?
(A) Two-proportion z-test for a difference of proportions
(B) Two-sample t-test for a difference of means
(C) Paired t-test for a mean difference
(D) Two-sample z-test for a difference of means
(E) One-proportion z-test

46. A customer service representative at a telecommunications provider is curious as to what proportions of customers leave each of four types of reviews: 1 – Highly unlikely to recommend, 2 – Somewhat unlikely to recommend, 3 – Somewhat likely to recommend, and 4 – Highly likely to recommend. She randomly samples 100 customer reviews from the previous month and calculates the proportions that leave each review. What type of statistical test should the representative use to test if customers are more likely to leave certain reviews than others?
(A) Chi-squared test of independence
(B) Two-proportion z-test
(C) One-proportion z-test
(D) Chi-squared goodness of fit test
(E) Chi-squared test of homogeneity

47. Suppose that a one-sample significance test of $H_0\colon \mu = 0$ versus $H_a\colon \mu \neq 0$ is going to be conducted. Before it is, two confidence intervals – one with 95% confidence and one with 90% confidence – are created about the sample mean $\bar{x}$. If the 95% interval contains the value 0, but the 90% interval does not contain 0, at which of the following significance levels would you reject the null hypothesis in favor of the alternative?

I. $\alpha = 0.01$
II. $\alpha = 0.05$
III. $\alpha = 0.10$

(A) I, II, and III (B) I and II only (C) II and III only (D) II only (E) III only

48. A track coach believes that a new style of running shoes will help his 800-meter (m) runners run more quickly. He takes his 14 runners and pairs the fastest two runners, the next fastest two, and so on, and randomly assigns one student in each pair to wear the new shoes and the other to wear the current team shoes. He has each student run the 800 m and records their times, as well as the difference in time, as shown below.

	Mean (in minutes)	Standard Deviation (in minutes)
Current shoes	2.115	0.035
New shoes	2.081	0.028
Difference (current – new)	0.034	0.045

Assuming that the distribution of the differences in times is reasonably normally distributed, what is the P-value for the correct test to determine if the new shoes help the students run faster?
(A) 0.007 (B) 0.014 (C) 0.016 (D) 0.034 (E) 0.068

49. A polling group wants to determine if there is an association between gender and political party in a certain city. It randomly samples 1,000 people and anonymously asks them their political preference. The results are shown in the table below.

	Democrat	Independent	Republican	Unsure	Total
Male	166	94	205	33	498
Female	206	91	149	56	502
Total	372	185	354	89	1000

The test statistic for a chi-squared test of association was $\chi^2 = 19.14$. What is the P-value for this test?
(A) 0.000 (B) 0.007 (C) 0.008 (D) 0.014 (E) 0.016

Use the following for questions #50-54: An ambitious reporter for a large university newspaper suspects that Mr. Hazzard, a new statistics teacher, is grading his introductory statistics students too harshly. From school records the reporter determines that over the past 2 years the proportions of students in *all* sections of introductory statistics (taught by many different teachers) received grades of A, B, C, D, or F in the following proportions: A: 0.20, B: 0.30, C: 0.30, D: 0.10, and F: 0.10. The reporter than takes a sample of 90 students who took introductory statistics with Mr. Hazzard in the past 2 years and gathers the following information:

Grade	A	B	C	D	F
Number of students	12	26	28	15	9

The reporter performs the appropriate χ^2 procedure to test the hypothesis that the teacher's grade distribution is different from other teachers of introductory statistics.

50. Assuming the χ^2 statistic has approximately a χ^2 distribution, how many degrees of freedom does the distribution have?
(A) 90 (B) 89 (C) 9 (D) 5 (E) 4

51. Which of the following conditions must be met before the reporter can use the χ^2 procedure in this situation?
(A) The distribution of grades in all introductory statistics courses must be approximately normal.
(B) The number of categories is small relative to the number of observations.
(C) All the observed counts are greater than 5.
(D) Each observation was randomly selected from the population of all grades given by the new teacher.
(E) All expected counts are approximately equal.

52. Which of the following represents the expected count of the grade category D?

(A) $\frac{90}{5}$ (B) (0.10)(90) (C) (0.10)(15) (D) $\frac{15^2}{90}$ (E) $\frac{(15-9)^2}{9}$

53. The grade category that contributes the largest component to the χ^2 statistic is

(A) A (B) B (C) C (D) D (E) F

54. The computed value of the χ^2 statistic for the reporter's test is 6.074. Which of the following is an appropriate conclusion at the $\alpha = 0.05$ significance level?

(A) Reject H_0: there is convincing evidence that the grade distribution of the new teacher is different from that of other teachers.

(B) Accept H_a: there is convincing evidence that the grade distribution of the new teacher is different from that of other teachers.

(C) Fail to reject H_0: there is insufficient evidence that the grade distribution of the new teacher is different from that of other teachers.

(D) Fail to reject H_0: there is convincing evidence that the grade distribution of the new teacher is different from that of other teachers.

(E) Accept H_0: there is insufficient evidence that the grade distribution of the new teacher is different from that of other teachers.

ANSWERS

1. B
2. C
3. C
4. C
5. B
6. C
7. D
8. C
9. E
10. B
11. C
12. D
13. C,A,B
14. D
15. D
16. B
17. B
18. E
19. C
20. E
21. D
22. D
23. D
24. C
25. C
26. C
27. D
28. B
29. A
30. A
31. B
32. C
33. C
34. E
35. B
36. B
37. E
38. D
39. D
40. C
41. C
42. E
43. B
44. E
45. C
46. D
47. E
48. A
49. A
50. E
51. D
52. B
53. D
54. C

2023 AP Exam Practice FRQs

These are mostly FRQs you have never seen or done before.

Some are newer (and therefore more reflective of current trends in FRQs), while some are older: it's good to get practice with both.

I have put the year and question number for each FRQ – if you'd like to look up the scoring guidelines for any questions, go to https://apstudents.collegeboard.org/courses/ap-statistics/free-response-questions-by-year .

2010 #1

Agricultural experts are trying to develop a bird deterrent to reduce costly damage to crops in the United States. An experiment is to be conducted using garlic oil to study its effectiveness as a nontoxic, environmentally safe bird repellant. The experiment will use European starlings, a bird species that causes considerable damage annually to the corn crop in the United States. Food granules made from corn are to be infused with garlic oil in each of five concentrations of garlic —0 percent, 2 percent, 10 percent, 25 percent, and 50 percent. The researchers will determine the adverse reaction of the birds to the repellant by measuring the number of food granules consumed during a two-hour period following overnight food deprivation. There are forty birds available for the experiment, and the researchers will use eight birds for each concentration of garlic. Each bird will be kept in a separate cage and provided with the same number of food granules.

(a) For the experiment, identify

i. the treatments

ii. the experimental units

iii. the response that will be measured

(b) After performing the experiment, the researchers recorded the data shown in the table below.

Garlic oil concentration	0%	2%	10%	25%	50%
Mean number of food granules consumed	58	48	29	24	20
Number of birds	8	8	8	8	8

i. Construct a graph of the data that could be used to investigate the appropriateness of a linear regression model for analyzing the results of the experiment.

ii. Based on your graph, do you think a linear regression model is appropriate? Explain.

1999 #3

The dentists in a dental clinic would like to determine if there is a difference between the number of new cavities in people who eat an apple a day and in people who eat less than one apple a week. They are going to conduct a study with 50 people in each group.

Fifty clinic patients who report that they routinely eat an apple a day and 50 clinic patients who report that they eat less than one apple a week will be identified. The dentists will examine the patients and their records to determine the number of new cavities the patients have had over the past two years. They will then compare the number of new cavities in the two groups.

a. Why is this an observational study and not an experiment?

b. Explain the concept of confounding in the context of this study. Include an example of a possible confounding variable.

c. If the mean number of new cavities for those who ate an apple a day was statistically significantly smaller than the mean number of new cavities for those who ate less than one apple a week, could one conclude that the lower number of new cavities can be attributed to eating an apple a day? Explain.

2006 Form B #5

When a tractor pulls a plow through an agricultural field, the energy needed to pull that plow is called the draft. The draft is affected by environmental conditions such as soil type, terrain, and moisture.

A study was conducted to determine whether a newly developed hitch would be able to reduce draft compared to the standard hitch. (A hitch is used to connect the plow to the tractor.) Two large plots of land were used in this study. It was randomly determined which plot was to be plowed using the standard hitch. As the tractor plowed that plot, a measurement device on the tractor automatically recorded the draft at 25 randomly selected points in the plot.

After the plot was plowed, the hitch was changed from the standard one to the new one, a process that takes a substantial amount of time. Then the second plot was plowed using the new hitch. Twenty-five measurements of draft were also recorded at randomly selected points in this plot.

(a) What was the response variable in this study?

Identify the treatments.

What were the experimental units?

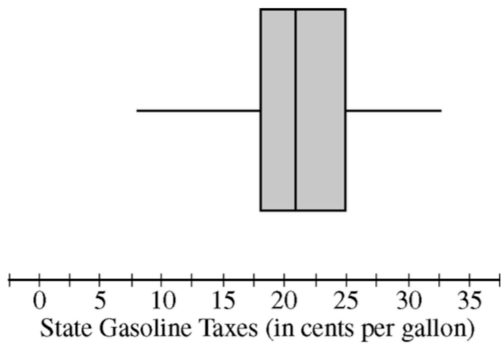
(b) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was randomization used properly in this study? Justify your answer.

(c) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was replication used properly in this study? Justify your answer.

(d) Plot of land is a confounding variable in this experiment. Explain why.

2009 Form B #1

As gasoline prices have increased in recent years, many drivers have expressed concern about the taxes they pay on gasoline for their cars. In the United States, gasoline taxes are imposed by both the federal government and by individual states. The boxplot below shows the distribution of the state gasoline taxes, in cents per gallon, for all 50 states on January 1, 2006.



(a) Based on the boxplot, what are the approximate values of the median and the interquartile range of the distribution of state gasoline taxes, in cents per gallon? Mark and label the boxplot to indicate how you found the approximated values.

(b) The federal tax imposed on gasoline was 18.4 cents per gallon at the time the state taxes were in effect. The federal gasoline tax was added to the state gasoline tax for each state to create a new distribution of combined gasoline taxes. What are approximate values, in cents per gallon, of the median and interquartile range of the new distribution of combined gasoline taxes? Justify your answer.

2005 Form B #1

The graph below displays the scores of 32 students on a recent exam. Scores on this exam ranged from 64 to 95 points.

6	* *
6	* *
7	* * *
7	* * * *
8	* * * *
8	* * * * * *
9	* * * * * * *
9	* * * *

(a) Describe the shape of this distribution.

(b) In order to motivate her students, the instructor of the class wants to report that, overall, the class's performance on the exam was high. Which summary statistic, the mean or the median, should the instructor use to report that overall exam performance was high? Explain.

(c) The midrange is defined as $\frac{\text{maximum} + \text{minimum}}{2}$. Compute this value using the data on the preceding page.

Is the midrange considered a measure of center or a measure of spread? Explain.

1997 #3

A laboratory test for the detection of a certain disease gives a positive result 5 percent of the time for people who do not have the disease. The test gives a negative result 0.3 percent of the time for people who have the disease. Large-scale studies have shown that the disease occurs in about 2 percent of the population.

(a) What is the probability that a person selected at random would test positive for this disease? Show your work.

(b) What is the probability that a person selected at random who tests positive for the disease does not have the disease? Show your work.

2014 #4

As part of its twenty-fifth reunion celebration, the class of 1988 (students who graduated in 1988) at a state university held a reception on campus. In an informal survey, the director of alumni development asked 50 of the attendees about their incomes. The director computed the mean income of the 50 attendees to be \$189,952. In a news release, the director announced, "The members of our class of 1988 enjoyed resounding success. Last year's mean income of its members was \$189,952!"

(a) What would be a statistical advantage of using the median of the reported incomes, rather than the mean, as the estimate of the typical income?

(b) The director felt the members who attended the reception may be different from the class as a whole. A more detailed survey of the class was planned to find a better estimate of the income as well as other facts about the alumni. The staff developed two methods based on the available funds to carry out the survey.

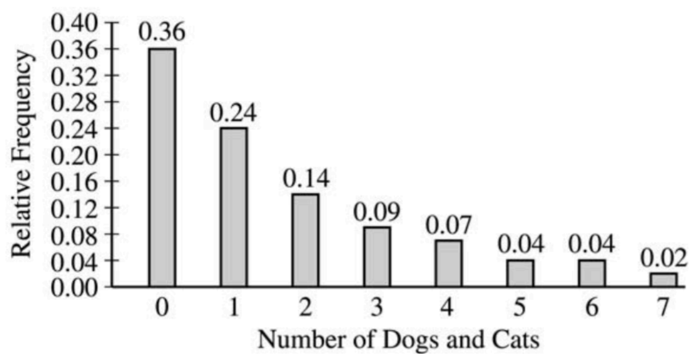
Method 1: Send out an e-mail to all 6,826 members of the class asking them to complete an online form. The staff estimates that at least 600 members will respond.

Method 2: Select a simple random sample of members of the class and contact the selected members directly by phone. Follow up to ensure that all responses are obtained. Because method 2 will require more time than method 1, the staff estimates that only 100 members of the class could be contacted using method 2.

Which of the two methods would you select for estimating the average yearly income of all 6,826 members of the class of 1988 ? Explain your reasoning by comparing the two methods and the effect of each method on the estimate.

2007 Form B #2

The graph below displays the relative frequency distribution for *X*, the total number of dogs and cats owned per household, for the households in a large suburban area. For instance, 14 percent of the households own 2 of these pets.



(a) According to a local law, each household in this area is prohibited from owning more than 3 of these pets. If a household in this area is selected at random, what is the probability that the selected household will be in violation of this law? Show your work.

(b) If 10 households in this area are selected at random, what is the probability that exactly 2 of them will be in violation of this law? Show your work.

(c) The mean and standard deviation of *X* are 1.65 and 1.851, respectively. Suppose 150 households in this area are to be selected at random and $\overline{X}$, the mean number of dogs and cats per household, is to be computed. Describe the sampling distribution of $\overline{X}$, including its shape, center, and spread.

2016 #4

A company manufactures model rockets that require igniters to launch. Once an igniter is used to launch a rocket, the igniter cannot be reused. Sometimes an igniter fails to operate correctly, and the rocket does not launch. The company estimates that the overall failure rate, defined as the percent of all igniters that fail to operate correctly, is 15 percent.

A company engineer develops a new igniter, called the super igniter, with the intent of lowering the failure rate. To test the performance of the super igniters, the engineer uses the following process.

Step 1: One super igniter is selected at random and used in a rocket.

Step 2: If the rocket launches, another super igniter is selected at random and used in a rocket.

Step 2 is repeated until the process stops. The process stops when a super igniter fails to operate correctly or 32 super igniters have successfully launched rockets, whichever comes first. Assume that super igniter failures are independent.

(a) If the failure rate of the super igniters is 15 percent, what is the probability that the first 30 super igniters selected using the testing process successfully launch rockets?

(b) Given that the first 30 super igniters successfully launch rockets, what is the probability that the first failure occurs on the thirty-first or the thirty-second super igniter tested if the failure rate of the super igniters is 15 percent?

(c) Given that the first 30 super igniters successfully launch rockets, is it reasonable to believe that the failure rate of the super igniters is less than 15 percent? Explain.

2021 #3

To increase morale among employees, a company began a program in which one employee is randomly selected each week to receive a gift card. Each of the company's 200 employees is equally likely to be selected each week, and the same employee could be selected more than once. Each week's selection is independent from every other week.

(a) Consider the probability that a particular employee receives at least one gift card in a 52-week year.

(i) Define the random variable of interest and state how the random variable is distributed.

(ii) Determine the probability that a particular employee receives at least one gift card in a 52-week year. Show your work.

(b) Calculate and interpret the expected value for the number of gift cards a particular employee will receive in a 52-week year. Show your work.

(c) Suppose that Agatha, an employee at the company, never receives a gift card for an entire 52-week year. Based on her experience, does Agatha have a strong argument that the selection process was not truly random? Explain your answer.

2010 #4

An automobile company wants to learn about customer satisfaction among the owners of five specific car models. Large sales volumes have been recorded for three of the models, but the other two models were recently introduced so their sales volumes are smaller. The number of new cars sold in the last six months for each of the models is shown in the table below.

Car Model	A	B	C	D	E	Total
Number of new cars sold in the last six months	112,338	96,174	83,241	3,278	2,323	297,354

The company can obtain a list of all individuals who purchased new cars in the last six months for each of the five models shown in the table. The company wants to sample 2,000 of these owners.

(a) For simple random samples of 2,000 new car owners, what is the expected number of owners of model E <u>and</u> the standard deviation of the number of owners of model E?

(b) When selecting a simple random sample of 2,000 new car owners, how likely is it that fewer than 12 owners of model E would be included in the sample? Justify your answer.

(c) The company is concerned that a simple random sample of 2,000 owners would include fewer than 12 owners of model D or fewer than 12 owners of model E. Briefly describe a sampling method for randomly selecting 2,000 owners that will ensure at least 12 owners will be selected for each of the 5 car models.

2011 Form B #3

An airline claims that there is a 0.10 probability that a coach-class ticket holder who flies frequently will be upgraded to first class on any flight. This outcome is independent from flight to flight. Sam is a frequent flier who always purchases coach-class tickets.

(a) What is the probability that Sam's first upgrade will occur after the third flight?

(b) What is the probability that Sam will be upgraded exactly 2 times in his next 20 flights?

(c) Sam will take 104 flights next year. Would you be surprised if Sam receives more than 20 upgrades to first class during the year? Justify your answer.

2009 #2

A tire manufacturer designed a new tread pattern for its all-weather tires. Repeated tests were conducted on cars of approximately the same weight traveling at 60 miles per hour. The tests showed that the new tread pattern enables the cars to stop completely in an average distance of 125 feet with a standard deviation of 6.5 feet and that the stopping distances are approximately normally distributed.

(a) What is the 70th percentile of the distribution of stopping distances?

(b) What is the probability that at least 2 cars out of 5 randomly selected cars in the study will stop in a distance that is greater than the distance calculated in part (a) ?

(c) What is the probability that a randomly selected sample of 5 cars in the study will have a mean stopping distance of at least 130 feet?

2008 Form B #3

A car manufacturer is interested in conducting a study to estimate the mean stopping distance for a new type of brakes when used in a car that is traveling at 60 miles per hour. These new brakes will be installed on cars of the same model and the stopping distance will be observed. The cost of each observation is $100. A budget of $12,000 is available to conduct the study and the goal is to carry it out in the most economical way possible. Preliminary studies indicate that $\sigma = 12$ feet for stopping distances.

(a) Are sufficient funds available to estimate the mean stopping distance to within 2 feet of the true mean stopping distance with 95% confidence?

Explain your answer.

(b) A regulatory agency requires a 95% level of confidence for an estimate of mean stopping distance that is within 2 feet of the true mean stopping distance. The car manufacturer cannot exceed the budget of $12,000 for the study. Discuss the consequences of these constraints.

2002 Form B #4

Each person in a random sample of 1,026 adults in the United States was asked the following question.

"Based on what you know about the Social Security system today, what would you like Congress and the President to do during this next year?"

The response choices and the percentages selecting them are shown below.

Completely overhaul the system	19%
Make some major changes	39%
Make some minor adjustments	30%
Leave the system the way it is now	11%
No opinion	1%

(a) Find a 95% confidence interval for the proportion of all United States adults who would respond "Make some major changes" to the question. Give an interpretation of the confidence interval and give an interpretation of the confidence level.

(b) An advocate for leaving the system as it is now commented, "Based on this poll, only 39% of adults in the sample responded that they want some major changes made to the system, while 41% responded that they want only minor changes or no changes at all. Therefore, we should not change the system." Explain why this statement, while technically correct, is misleading.

2011 Form B #5

During a flu vaccine shortage in the United States, it was believed that 45 percent of vaccine-eligible people received flu vaccine. The results of a survey given to a random sample of 2,350 vaccine-eligible people indicated that 978 of the 2,350 people had received flu vaccine.

(a) Construct a 99 percent confidence interval for the proportion of vaccine-eligible people who had received flu vaccine. Use your confidence interval to comment on the belief that 45 percent of the vaccine-eligible people had received flu vaccine.

(b) Suppose a similar survey will be given to vaccine-eligible people in Canada by Canadian health officials. A 99 percent confidence interval for the proportion of people who will have received flu vaccine is to be constructed. What is the smallest sample size that can be used to guarantee that the margin of error will be less than or equal to 0.02 ?

2005 #5

A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:

- Does not have a high school diploma
- Has a high school diploma

The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

(a) For this survey, state one potential source of bias and describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.

(b) A pilot survey indicated that about 22 percent of the population of adult heads of households do not have a high school diploma. Using this information, how many respondents should be obtained if the goal of the survey is to estimate the proportion of the population who do not have a high school diploma to within 0.03 with 95 percent confidence? Justify your answer.

(c) Since education is largely the responsibility of each state, the agency wants to be sure that estimates are available for each state as well as for the nation. Identify a sampling method that will achieve this additional goal and briefly describe a way to select the survey sample using this method.

2005 Form B #4

A researcher believes that treating seeds with certain additives before planting can enhance the growth of plants. An experiment to investigate this is conducted in a greenhouse. From a large number of Roma tomato seeds, 24 seeds are randomly chosen and 2 are assigned to each of 12 containers. One of the 2 seeds is randomly selected and treated with the additive. The other seed serves as a control. Both seeds are then planted in the same container. The growth, in centimeters, of each of the 24 plants is measured after 30 days. These data were used to generate the partial computer output shown below. Graphical displays indicate that the assumption of normality is not unreasonable.

	N	Mean	StDev	SE Mean
Control	12	15.989	1.098	0.317
Treatment	12	18.004	1.175	0.339
Difference	12	−2.015	1.163	0.336

(a) Construct a confidence interval for the mean difference in growth, in centimeters, of the plants from the untreated and treated seeds. Be sure to interpret this interval.

(b) Based only on the confidence interval in part (a), is there sufficient evidence to conclude that there is a significant mean difference in growth of the plants from untreated seeds and the plants from treated seeds? Justify your conclusion.

2003 #1

Since Hill Valley High School eliminated the use of bells between classes, teachers have noticed that more students seem to be arriving to class a few minutes late. One teacher decided to collect data to determine whether the students' and teachers' watches are displaying the correct time. At exactly 12:00 noon, the teacher asked 9 randomly selected students and 9 randomly selected teachers to record the times on their watches to the nearest half minute. The ordered data showing minutes after 12:00 as positive values and minutes before 12:00 as negative values are shown in the table below.

Students	-4.5	-3.0	-0.5	0	0	0.5	0.5	1.5	5.0
Teachers	-2.0	-1.5	-1.5	-1.0	-1.0	-0.5	0	0	0.5

(a) Construct parallel boxplots using these data.

(b) Based on the boxplots in part (a), which of the two groups, students or teachers, tends to have watch times that are closer to the true time? Explain your choice.

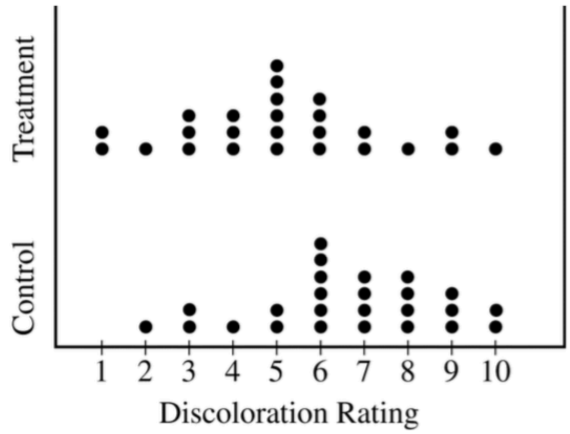
(c) The teacher wants to know whether individual student's watches tend to be set correctly. She proposes to test H_0: $\mu = 0$ versus H_a: $\mu \neq 0$, where μ represents the mean amount by which all student watches differ from the correct time. Is this an appropriate pair of hypotheses to test to answer the teacher's question? Explain why or why not. Do not carry out the test.

2007 #1

The department of agriculture at a university was interested in determining whether a preservative was effective in reducing discoloration in frozen strawberries. A sample of 50 ripe strawberries was prepared for freezing. Then the sample was randomly divided into two groups of 25 strawberries each. Each strawberry was placed into a small plastic bag.

The 25 bags in the control group were sealed. The preservative was added to the 25 bags containing strawberries in the treatment group, and then those bags were sealed. All bags were stored at $0°C$ for a period of 6 months. At the end of this time, after the strawberries were thawed, a technician rated each strawberry's discoloration from 1 to 10, with a low score indicating little discoloration.

The dotplots below show the distributions of discoloration rating for the control and treatment groups.



(a) The standard deviation of ratings for the control group is 2.141. Explain how this value summarizes variability in the control group.

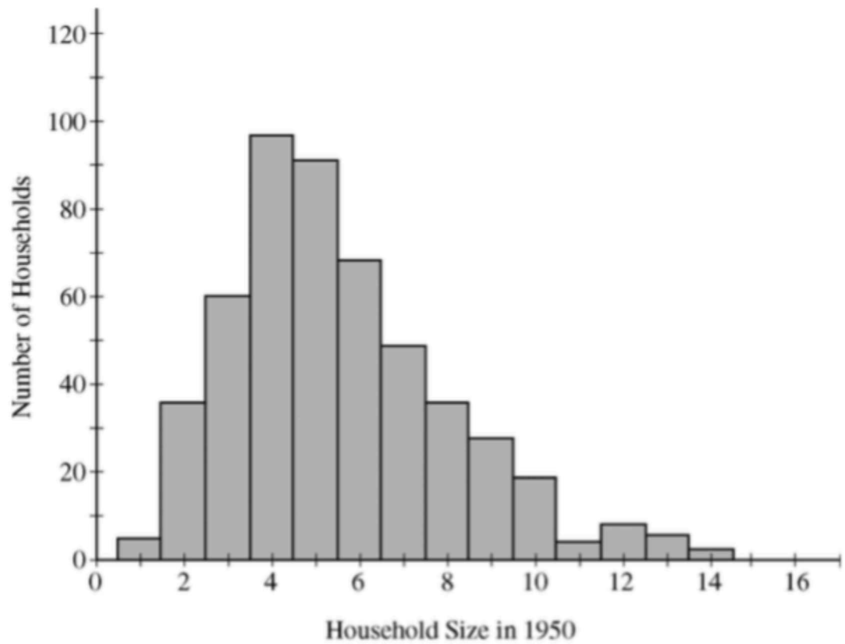
(b) Based on the dotplots, comment on the effectiveness of the preservative in lowering the amount of discoloration in strawberries. (No calculations are necessary.)

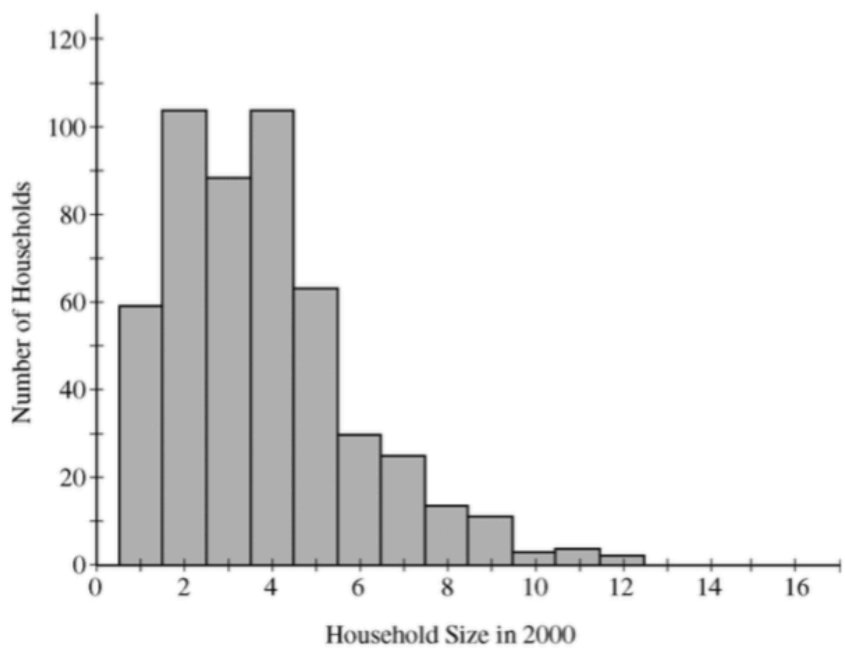
(c) Researchers at the university decided to calculate a 95 percent confidence interval for the difference in mean discoloration rating between strawberries that were not treated with preservative and those that were treated with preservative. The confidence interval they obtained was (0.16, 2.72). Assume that the conditions necessary for the *t*-confidence interval are met.

Based on the confidence interval, comment on whether there would be a difference in the population mean discoloration ratings for the treated and untreated strawberries.

2012 #3

Independent random samples of 500 households were taken from a large metropolitan area in the United States for the years 1950 and 2000. Histograms of household size (number of people in a household) for the years are shown below.





(a) Compare the distributions of household size in the metropolitan area for the years 1950 and 2000.

(b) A researcher wants to use these data to construct a confidence interval to estimate the change in mean household size in the metropolitan area from the year 1950 to the year 2000. State the conditions for using a two-sample *t*-procedure, and explain whether the conditions for inference are met.

2008 Form B #1

A certain state's education commissioner released a new report card for all the public schools in that state. This report card provides a new tool for comparing schools across the state. One of the key measures that can be computed from the report card is the student-to-teacher ratio, which is the number of students enrolled in a given school divided by the number of teachers at that school.

The data below give the student-to-teacher ratio at the 10 schools with the highest proportion of students meeting the state reading standards in the third grade and at the 10 schools with the lowest proportion of students meeting the state reading standards in the third grade.

Ratios in the 10 Schools
with Highest Proportion of Students Meeting Standards

7	21	18	22	9	16	12	17	17	16

Ratios in the 10 Schools
with Lowest Proportion of Students Meeting Standards

14	16	18	20	12	14	16	12	20	19

(a) Display a dotplot for each group to compare the distribution of student-to-teacher ratios in the top 10 schools with the distribution in the bottom 10 schools. Comment on the similarities and differences between the two distributions.

(b) Any statistical test that is used to determine whether the mean student-to-teacher ratio is the same for the top 10 schools as it is for the bottom 10 schools would be inappropriate. Explain why in a few sentences.

2005 Form B #3

In search of a mosquito repellent that is safer than the ones that are currently on the market, scientists have developed a new compound that is rated as less toxic than the current compound, thus making a repellent that contains this new compound safer for human use. Scientists also believe that a repellent containing the new compound will be more effective than the ones that contain the current compound. To test the effectiveness of the new compound versus that of the current compound, scientists have randomly selected 100 people from a state.

Up to 100 bins, with an equal number of mosquitoes in each bin, are available for use in the study. After a compound is applied to a participant's forearm, the participant will insert his or her forearm into a bin for 1 minute, and the number of mosquito bites on the arm at the end of that time will be determined.

(a) Suppose this study is to be conducted using a completely randomized design. Describe a randomization process and identify an inference procedure for the study.

(b) Suppose this study is to be conducted using a matched-pairs design. Describe a randomization process and identify an inference procedure for the study.

(c) Which of the designs, the one in part (a) or the one in part (b), is better for testing the effectiveness of the new compound versus that of the current compound? Justify your answer.

2003 Form B #4

There have been many studies recently concerning coffee drinking and cholesterol level. While it is known that several coffee-bean components can elevate blood cholesterol level, it is thought that a new type of paper coffee filter may reduce the presence of some of these components in coffee.

The effect of the new filter on cholesterol level will be studied over a 10-week period using 300 nonsmokers who each drink 4 cups of caffeinated coffee per day. Each of these 300 participants will be assigned to one of two groups: the experimental group, who will only drink coffee that has been made with the new filter, or the control group, who will only drink coffee that has been made with the standard filter. Each participant's cholesterol level will be measured at the beginning and at the end of the study.

(a) Describe an appropriate method for assigning the subjects to the two groups so that each group will have an equal number of subjects.

(b) In this study, the researchers chose to include a group who only drank coffee that was made with the standard filter. Why is it important to include a control group in this study even though cholesterol levels will be measured at the beginning and at the end of the study?

(c) Which test would you conduct to determine whether the change in cholesterol level would be greater if people used the new filter rather than using the standard filter?

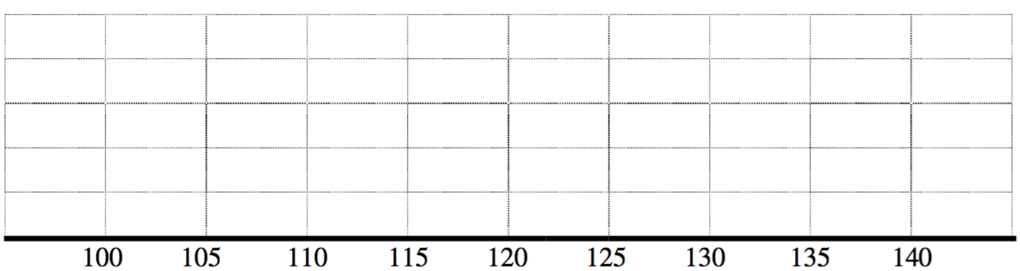
(d) Why would the researchers choose to use only nonsmokers in the study?

2004 Form B #5

A researcher thinks that modern Thai dogs may be descendants of golden jackals. A random sample of 16 animals was collected from each of the two populations. The length (in millimeters) of the mandible (jawbone) was measured for each animal. The lower quartile, median, and upper quartile for each sample are shown in the table below, along with all values below the lower quartile and all values above the upper quartile.

Sample	Values Below Q_1	Q_1	Median	Q_3	Values Above Q_3
Modern Thai dog	114, 116, 116, 120	121	125	128	129, 130, 130, 132
Golden jackal	104, 104, 105, 106	107	108	112	114, 122, 124, 125

(a) Display parallel boxplots of mandible lengths (showing outliers, if any) for the modern Thai dogs and the golden jackals on the grid below.



Based on the boxplots, write a few sentences comparing the distributions of mandible lengths for the two types of dogs.

(b) Is it reasonable to use the sample of mandible lengths of modern Thai dogs to construct an interval estimate of the mean mandible length for the population of modern Thai dogs? Justify your answer. (Note: You do not have to compute the interval.)

(c) Is it reasonable to use the sample data of mandible lengths of modern Thai dogs and the sample data of mandible lengths of golden jackals to perform a two-sample t-test for the difference in mean mandible lengths for the two types of dogs? Justify your answer. (Note: You do not have to conduct the test.)

1997 #4

A random sample of 415 potential voters was interviewed 3 weeks before the start of a state-wide campaign for governor; 223 of the 415 said they favored the new candidate over the incumbent. However, the new candidate made several unfortunate remarks one week before the election. Subsequently, a new random sample of 630 potential voters showed that 317 voters favored the new candidate.

Do these data support the conclusion that there was a decrease in voter support for the new candidate after the unfortunate remarks were made? Give appropriate statistical evidence to support your answer.

2013 #5

Psychologists interested in the relationship between meditation and health conducted a study with a random sample of 28 men who live in a large retirement community. Of the men in the sample, 11 reported that they participate in daily meditation and 17 reported that they do not participate in daily meditation.

The researchers wanted to perform a hypothesis test of

$$H_0 : p_m - p_c = 0$$
$$H_a : p_m - p_c < 0,$$

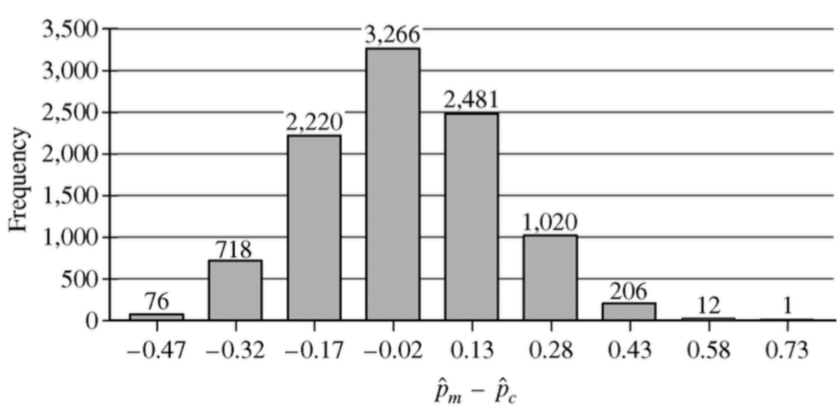
where p_m is the proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation and p_c is the proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily meditation.

(a) If the study were to provide significant evidence against H_0 in favor of H_a, would it be reasonable for the psychologists to conclude that daily meditation causes a reduction in blood pressure for men in the retirement community? Explain why or why not.

The psychologists found that of the 11 men in the study who participate in daily meditation, 0 had high blood pressure. Of the 17 men who do not participate in daily meditation, 8 had high blood pressure.

(b) Let $\hat{p}_m$ represent the proportion of men with high blood pressure among those in a random sample of 11 who meditate daily, and let $\hat{p}_c$ represent the proportion of men with high blood pressure among those in a random sample of 17 who do not meditate daily. Why is it not reasonable to use a normal approximation for the sampling distribution of $\hat{p}_m - \hat{p}_c$?

Although a normal approximation cannot be used, it is possible to simulate the distribution of $\hat{p}_m - \hat{p}_c$. Under the assumption that the null hypothesis is true, 10,000 values of $\hat{p}_m - \hat{p}_c$ were simulated. The histogram below shows the results of the simulation.



(c) Based on the results of the simulation, what can be concluded about the relationship between blood pressure and meditation among men in the retirement community?

2009 #5

For many years, the medically accepted practice of giving aid to a person experiencing a heart attack was to have the person who placed the emergency call administer chest compression (CC) plus standard mouth-to-mouth resuscitation (MMR) to the heart attack patient until the emergency response team arrived. However, some researchers believed that CC alone would be a more effective approach.

In the 1990s a study was conducted in Seattle in which 518 cases were randomly assigned to treatments: 278 to CC plus standard MMR and 240 to CC alone. A total of 64 patients survived the heart attack: 29 in the group receiving CC plus standard MMR, and 35 in the group receiving CC alone. A test of significance was conducted on the following hypotheses.

H_0: The survival rates for the two treatments are equal.

H_a: The treatment that uses CC alone produces a higher survival rate.

This test resulted in a *p*-value of 0.0761.

(a) Interpret what this *p*-value measures in the context of this study.

(b) Based on this *p*-value and study design, what conclusion should be drawn in the context of this study? Use a significance level of $\alpha = 0.05$.

(c) Based on your conclusion in part (b), which type of error, Type I or Type II, could have been made? What is one potential consequence of this error?

2021 #4

The manager of a large company that sells pet supplies online wants to increase sales by encouraging repeat purchases. The manager believes that if past customers are offered $10 off their next purchase, more than 40 percent of them will place an order. To investigate the belief, 90 customers who placed an order in the past year are selected at random. Each of the selected customers is sent an e-mail with a coupon for $10 off the next purchase if the order is placed within 30 days. Of those who receive the coupon, 38 place an order.

(a) Is there convincing statistical evidence, at the significance level of $\alpha = 0.05$, that the manager's belief is correct? Complete the appropriate inference procedure to support your answer.

(b) Based on your conclusion from part (a), which of the two errors, Type I or Type II, could have been made? Interpret the consequence of the error in context.

2010 Form B #5

An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

	HIGHEST LEVEL OF EDUCATIONAL ACHIEVEMENT			
Primary Source for News	Not High School Graduate	High School Graduate But Not College Graduate	College Graduate	Total
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1,437	693	2,500

(a) If an adult is to be selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the internet?

(b) If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

(c) When selecting an adult at random from the sample of 2,500 adults, are the events "is a college graduate" and "obtains news primarily from the internet" independent? Justify your answer.

(d) The company wants to conduct a statistical test to investigate whether there is an association between educational achievement and primary source for news for adults in the city. What is the name of the statistical test that should be used?

What are the appropriate degrees of freedom for this test?

2011 Form B #4

A parent advisory board for a certain university was concerned about the effect of part-time jobs on the academic achievement of students attending the university. To obtain some information, the advisory board surveyed a simple random sample of 200 of the more than 20,000 students attending the university. Each student reported the average number of hours spent working part-time each week and his or her perception of the effect of part-time work on academic achievement. The data in the table below summarize the students' responses by average number of hours worked per week (less than 11, 11 to 20, more than 20) and perception of the effect of part-time work on academic achievement (positive, no effect, negative).

		Average Time Spent on Part-Time Jobs		
		Less Than 11 Hours per Week	11 to 20 Hours per Week	More Than 20 Hours per Week
Perception of the Effect of Part-Time Work on Academic Achievement	Positive Effect	21	9	5
	No Effect	58	32	15
	Negative Effect	18	23	19

A chi-square test was used to determine if there is an association between the effect of part-time work on academic achievement and the average number of hours per week that students work. Computer output that resulted from performing this test is shown below.

CHI-SQUARE TEST

Expected counts are printed below observed counts

	<11	11–20	>20	Total
Positive	21 16.975	9 11.200	5 6.825	35
No effect	58 50.925	32 33.600	15 20.475	105
Negative	18 29.100	23 19.200	19 11.700	60
Total	97	64	39	200

Chi-Sq = 13.938, DF = 4, P-Value = 0.007

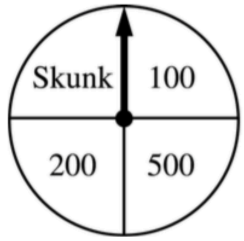
(a) State the null and alternative hypotheses for this test.

(b) Discuss whether the conditions for a chi-square inference procedure are met for these data.

(c) Given the results from the chi-square test, what should the advisory board conclude?

(d) Based on your conclusion in part (c), which type of error (Type I or Type II) might the advisory board have made? Describe this error in the context of the question.

2003 Form B #5



Contestants on a game show spin a wheel like the one shown in the figure above. Each of the four outcomes on this wheel is equally likely and outcomes are independent from one spin to the next.

- The contestant spins the wheel.
- If the result is a skunk, no money is won and the contestant's turn is finished.
- If the result is a number, the corresponding amount in dollars is won. The contestant can then stop with those winnings or can choose to spin again, and his or her turn continues.
- If the contestant spins again and the result is a skunk, all of the money earned on that turn is lost and the turn ends.
- The contestant may continue adding to his or her winnings until he or she chooses to stop or until a spin results in a skunk.

(a) What is the probability that the result will be a number on all of the first three spins of the wheel?

(b) Suppose a contestant has earned \$800 on his or her first three spins and chooses to spin the wheel again. What is the expected value of his or her total winnings for the four spins?

(c) A contestant who lost at this game alleges that the wheel is not fair. In order to check on the fairness of the wheel, the data in the table below were collected for 100 spins of this wheel.

Result	Skunk	\$100	\$200	\$500
Frequency	33	21	20	26

Based on these data, can you conclude that the four outcomes on this wheel are not equally likely? Give appropriate statistical evidence to support your answer.

2014 #1

An administrator at a large university is interested in determining whether the residential status of a student is associated with level of participation in extracurricular activities. Residential status is categorized as on campus for students living in university housing and off campus otherwise. A simple random sample of 100 students in the university was taken, and each student was asked the following two questions.

- Are you an on campus student or an off campus student?
- In how many extracurricular activities do you participate?

The responses of the 100 students are summarized in the frequency table shown.

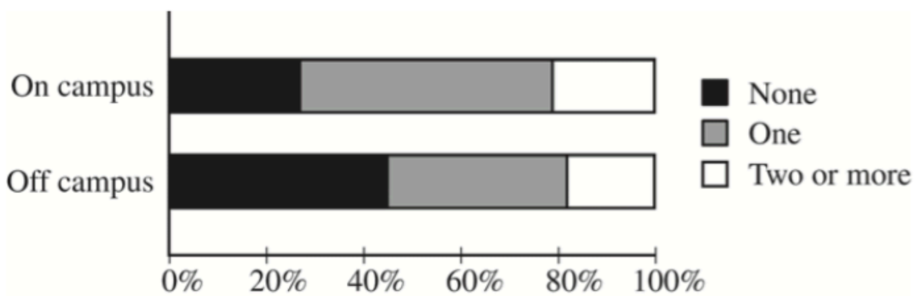
	Residential Status		
Level of Participation in Extracurricular Activities	On campus	Off campus	Total
No activities	9	30	39
One activity	17	25	42
Two or more activities	7	12	19
Total	33	67	100

(a) Calculate the proportion of on campus students in the sample who participate in at least one extracurricular activity and the proportion of off campus students in the sample who participate in at least one extracurricular activity.

On campus proportion:

Off campus proportion:

The responses of the 100 students are summarized in the segmented bar graph shown.



(b) Write a few sentences summarizing what the graph reveals about the association between residential status and level of participation in extracurricular activities among the 100 students in the sample.

After verifying that the conditions for inference were satisfied, the administrator performed a chi-square test of the following hypotheses.

H_0 : There is no association between residential status and level of participation in extracurricular activities among the students at the university.

H_a : There is an association between residential status and level of participation in extracurricular activities among the students at the university.

The test resulted in a p-value of 0.23. Based on the p-value, what conclusion should the administrator make?

1999 #2

The Colorado Rocky Mountain Rescue Service wishes to study the behavior of lost hikers. If more were known about the direction in which lost hikers tend to walk, then more effective search strategies could be devised. Two hundred hikers selected at random from those applying for hiking permits are asked whether they would head uphill, downhill, or remain in the same place if they became lost while hiking. Each hiker in the sample was also classified according to whether he or she was an experienced or novice hiker. The resulting data are summarized in the following table.

	Direction		
	Uphill	Downhill	Remain in Same Place
Novice	20	50	50
Experienced	10	30	40

Do these data provide convincing evidence of an association between the level of hiking expertise and the direction the hiker would head if lost?

Give appropriate statistical evidence to support your conclusion.

2004 #5

A rural county hospital offers several health services. The hospital administrators conducted a poll to determine whether the residents' satisfaction with the available services depends on their gender. A random sample of 1,000 adult county residents was selected. The gender of each respondent was recorded and each was asked whether he or she was satisfied with the services offered by the hospital. The resulting data are shown in the table below.

	Male	Female	Total
Satisfied	384	416	800
Not Satisfied	80	120	200
Total	464	536	1,000

(a) Using a significance level of 0.05, conduct an appropriate test to determine if, for adult residents of this county, there is an association between gender and whether or not they were satisfied with services offered by the hospital.

(b) Is $\frac{800}{1,000}$ a reasonable estimate for the proportion of all adult county residents who are satisfied with the services offered by this hospital? Explain why or why not.

2003 #5

A random sample of 200 students was selected from a large college in the United States. Each selected student was asked to give his or her opinion about the following statement.

> "The most important quality of a person who aspires to be the President of the United States is a knowledge of foreign affairs."

Each response was recorded in one of five categories. The gender of each selected student was noted. The data are summarized in the table below.

	Response Category				
	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Male	10	15	15	25	25
Female	20	25	25	25	15

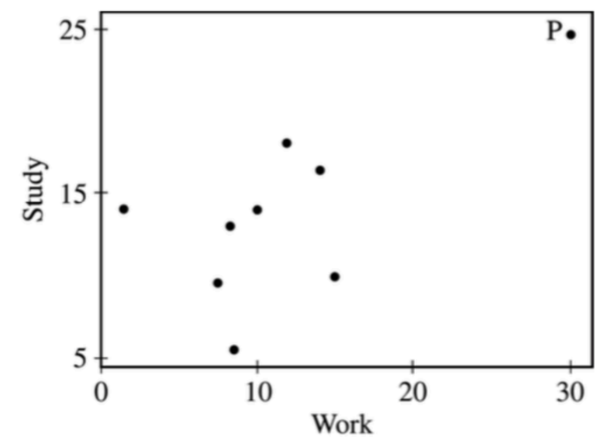
Is there sufficient evidence to indicate that the response is dependent on gender? Provide statistical evidence to support your conclusion.

2003 Form B #1

A simple random sample of 9 students was selected from a large university. Each of these students reported the number of hours he or she had allocated to studying and the number of hours allocated to work each week. A least squares linear regression was performed and part of the resulting computer output is shown below.

Predictor	Coef	StDev	T	P
Constant	8.107	2.731	2.97	0.021
Work	0.4919	0.1950	2.52	0.040
S = 4.349	R-Sq = 47.6%	R-Sq (adj) = 40.1%		

The scatterplot below displays the data that were collected from the 9 students.

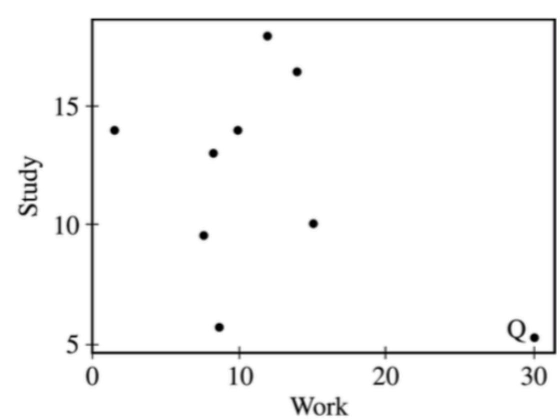


(a) After point P, labeled on the graph on the previous page, was removed from the data, a second linear regression was performed and the computer output is shown below.

Predictor	Coef	StDev	T	P
Constant	11.123	3.986	2.79	0.032
Work	0.1500	0.3834	0.39	0.709
S = 4.327	R-Sq = 2.5%	R-Sq (adj) = 0.0%		

Does point P exercise a large influence on the regression line? Explain.

(b) The researcher who conducted the study discovered that the number of hours spent studying reported by the student represented by P was recorded incorrectly. The corrected data point for this student is represented by the letter Q in the scatterplot below.

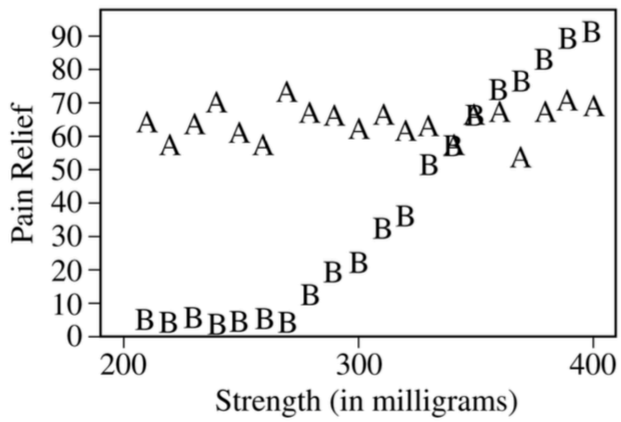


Explain how the least squares regression line for the corrected data (in this part) would differ from the least squares regression line for the original data.

2000 #1

Two pain relievers, A and B, are being compared for relief of postsurgical pain. Twenty different strengths (doses in milligrams) of each drug were tested. Eight hundred postsurgical patients were randomly divided into 40 different groups. Twenty groups were given drug A. Each group was given a different strength. Similarly, the other twenty groups were given different strengths of drug B. Strengths used ranged from 210 to 400 milligrams. Thirty minutes after receiving the drug, each patient was asked to describe his or her pain relief on a scale of 0 (no decrease in pain) to 100 (pain totally gone).

The strength of the drug given in milligrams and the average pain rating for each group are shown in the scatterplot below. Drug A is indicated with A's and drug B with B's.



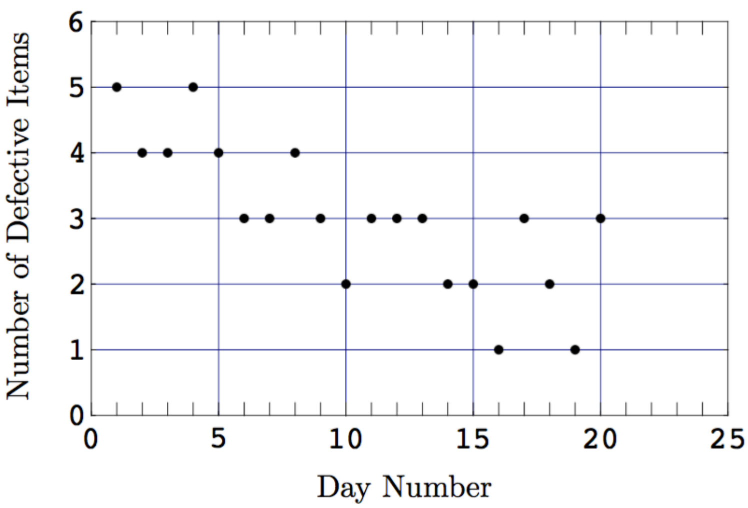
(a) Based on the scatterplot, describe the effect of drug A <u>and</u> how it is related to strength in milligrams.

(b) Based on the scatterplot, describe the effect of drug B <u>and</u> how it is related to strength in milligrams.

(c) Which drug would you give <u>and</u> at what strength, if the goal is to get pain relief of at least 50 at the lowest possible strength? Justify your answer based on the scatterplot.

1998 #2

A plot of the number of defective items produced during 20 consecutive days at a factory is shown below.



(a) Draw a histogram that shows the frequencies of the number of defective items.

(b) Give one fact that is obvious from the histogram but is not obvious from the scatterplot.

(c) Give one fact that is obvious from the scatterplot but is not obvious from the histogram.

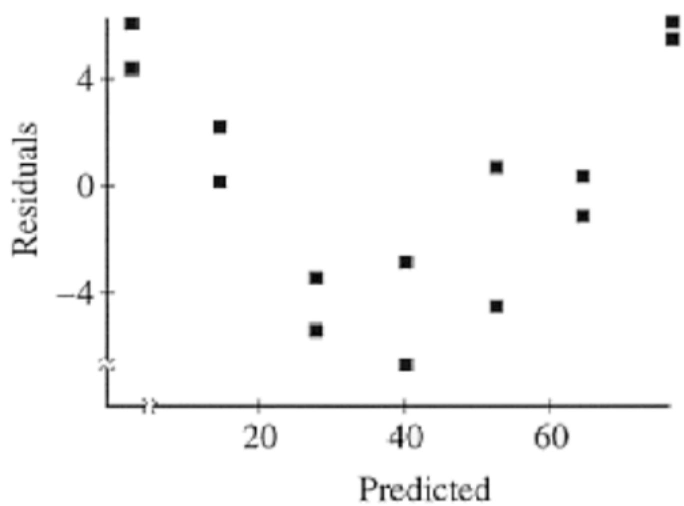
1998 #4

In a study of the application of a certain type of weed killer, 14 fields containing large numbers of weeds were treated. The weed killer was prepared at seven different strengths by adding 1, 1.5, 2, 2.5, 3, 3.5, or 4 teaspoons to a gallon of water. Two randomly selected fields were treated with each strength of weed killer. After a few days, the percentage of weeds killed on each filed was measured. The computer output obtained from fitting a least squares regression line to the data is shown below. A plot of the residuals is provided as well.

Dependent variable is: percent killed
R squared = 97.2%   R squared (adjusted) = 96.9%
$s = 4.505$ with $14 - 2 = 12$ degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	8330.16	1	8330.16	410
Residual	243.589	12	20.2990	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**Prob**
Constant	-20.5893	3.242	-6.35	≤ 0.0001
No. Teaspoons	24.3929	1.204	20.3	≤ 0.0001



(a) What is the equation of the least squares regression line given by this analysis? Define any variables used in this equation.

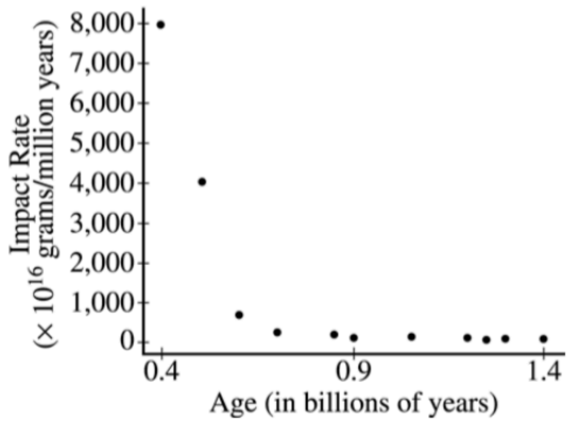
(b) If someone uses this equation to predict the percentage of weeds killed when 2.6 teaspoons of weed killer are used, which of the following would you expect?

- ❍ The prediction will be too large.
- ❍ The prediction will be too small.
- ❍ A prediction cannot be made based on the information given on the computer output.

Explain your reasoning.

2004 Form B #1

The Earth's Moon has many impact craters that were created when the inner solar system was subjected to heavy bombardment of small celestial bodies. Scientists studied 11 impact craters on the Moon to determine whether there was any relationship between the age of the craters (based on radioactive dating of lunar rocks) and the impact rate (as deduced from the density of the craters). The data are displayed in the scatterplot below.



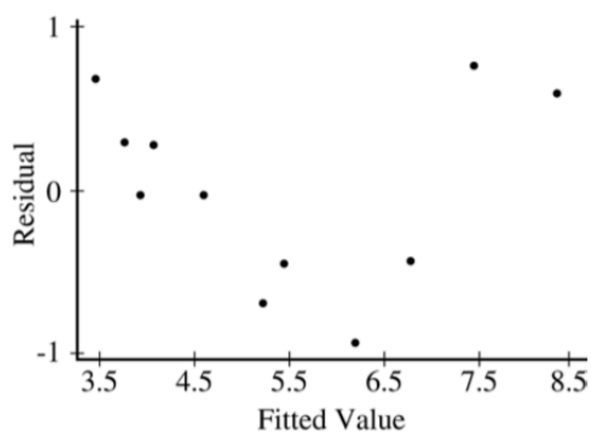
(a) Describe the nature of the relationship between impact rate and age.

Prior to fitting a linear regression model, the researchers transformed both impact rate and age by using logarithms. The following computer output and residual plot were produced.

Regression Equation: ln(rate) = 4.82 – 3.92 ln(age)

Predictor	Coef	SE Coef	T	P
Constant	4.8247	0.1931	24.98	0.000
ln(age)	–3.9232	0.4514	–8.69	0.000

S = 0.5977 R-Sq = 89.4% R-Sq (adj) = 88.2%

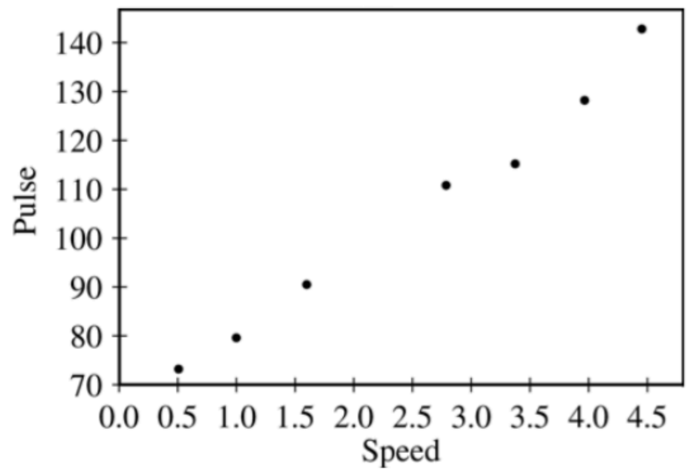


(b) Interpret the value of r^2.

(c) Comment on the appropriateness of this linear regression for modeling the relationship between the transformed variables.

2005 Form B #5

John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087   R-Sq = 98.7%   R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(a) Using the regression output, write the equation of the fitted regression line.

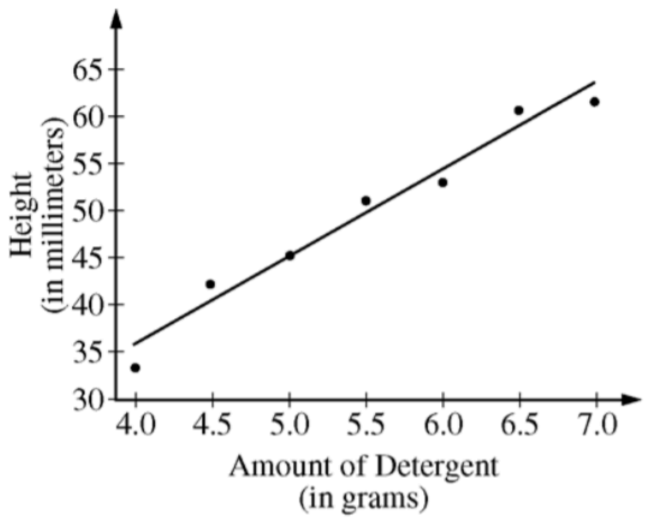
(b) Do your estimates of the slope and intercept parameters have meaningful interpretations in the context of this question? If so, provide interpretations in this context. If not, explain why not.

(c) John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.

2006 #2

A manufacturer of dish detergent believes the height of soapsuds in the dishpan depends on the amount of detergent used. A study of the suds' heights for a new dish detergent was conducted. Seven pans of water were prepared. All pans were of the same size and type and contained the same amount of water. The temperature of the water was the same for each pan. An amount of dish detergent was assigned at random to each pan, and that amount of detergent was added to the pan. Then the water in the dishpan was agitated for a set amount of time, and the height of the resulting suds was measured.

A plot of the data and the computer output from fitting a least squares regression line to the data are shown below.



Predictor	Coef	SE Coef	T	P
Constant	−2.679	4.222	−0.63	0.554
Amount	9.5000	0.7553	12.58	0.000

S = 1.99821   R–Sq = 96.9%   R–Sq(adj) = 96.3%

(a) Write the equation of the fitted regression line. Define any variables used in this equation.

(b) Note that $s = 1.99821$ in the computer output. Interpret this value in the context of this study.

(c) Identify and interpret the standard error of the slope.

2011 #5

Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

AP STATISTICS: SELECT #6 QUESTIONS

These questions should take approximately 25-30 minutes. On the AP Exam, they are scored out of 4, but are worth 12.5 out of the 100 total points. Ideally, you want to score at least a 1 on each of these.

2006

6. A manufacturer of thermostats is concerned that the readings of its thermostats have become less reliable (more variable). In the past, the variance has been 1.52 degrees Fahrenheit (F) squared. A random sample of 10 recently manufactured thermostats was selected and placed in a room that was maintained at 68°F. The readings for those 10 thermostats are given in the table below.

Thermostat	1	2	3	4	5	6	7	8	9	10
Temperature (°F)	66.8	67.8	70.6	69.3	65.9	66.2	68.1	68.6	67.9	67.2

(a) State the null and alternative hypotheses that the manufacturer is interested in testing.

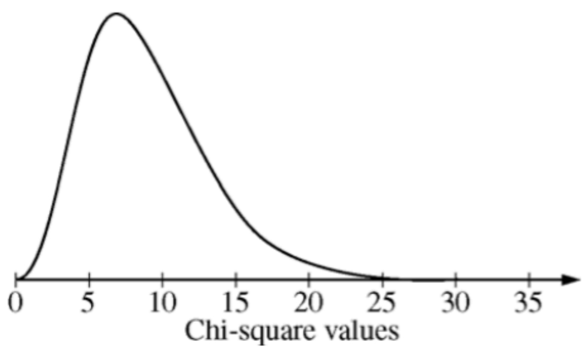
It can be shown that if the population of thermostat temperatures is normally distributed, the sampling distribution of $\frac{(n-1)s^2}{\sigma^2}$ follows a chi-square distribution with $n-1$ degrees of freedom.

(b) Calculate the value of $\frac{(n-1)s^2}{1.52}$ for these data.

(c) Assume that the population of thermostat temperatures follows a normal distribution. Use the test statistic $\frac{(n-1)s^2}{1.52}$ from part (b) and the chi-square distribution to test the hypotheses in part (a).

(d) For the test conducted in part (c), what is the smallest value of the test statistic that would have led to the rejection of the null hypothesis at the 5 percent significance level?

Mark this value of the test statistic on the graph of the chi-square distribution below. Indicate the region that contains all of the values that would have led to the rejection of the null hypothesis.

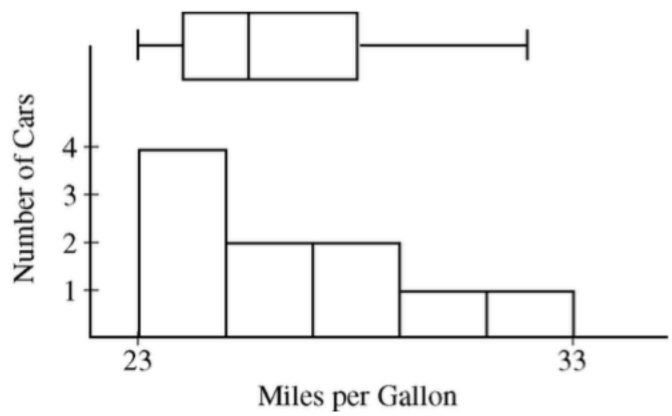


2009

6. A consumer organization was concerned that an automobile manufacturer was misleading customers by overstating the average fuel efficiency (measured in miles per gallon, or mpg) of a particular car model. The model was advertised to get 27 mpg. To investigate, researchers selected a random sample of 10 cars of that model. Each car was then randomly assigned a different driver. Each car was driven for 5,000 miles, and the total fuel consumption was used to compute mpg for that car.

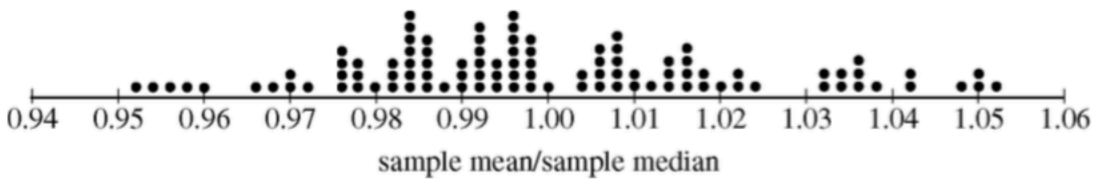
(a) Define the parameter of interest and state the null and alternative hypotheses the consumer organization is interested in testing.

One condition for conducting a one-sample *t*-test in this situation is that the mpg measurements for the population of cars of this model should be normally distributed. However, the boxplot and histogram shown below indicate that the distribution of the 10 sample values is skewed to the right.



(b) One possible statistic that measures skewness is the ratio $\frac{\text{sample mean}}{\text{sample median}}$. What values of that statistic (small, large, close to one) might indicate that the population distribution of mpg values is skewed to the right? Explain.

(c) Even though the mpg values in the sample were skewed to the right, it is still possible that the population distribution of mpg values is normally distributed and that the skewness was due to sampling variability. To investigate, 100 samples, each of size 10, were taken from a normal distribution with the same mean and standard deviation as the original sample. For each of those 100 samples, the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was calculated. A dotplot of the 100 simulated statistics is shown below.



In the original sample, the value of the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was 1.03. Based on the value of 1.03 and the dotplot above, is it plausible that the original sample of 10 cars came from a normal population, or do the simulated results suggest the original population is really skewed to the right? Explain.

(d) The table below shows summary statistics for mpg measurements for the original sample of 10 cars.

Minimum	Q1	Median	Q3	Maximum
23	24	25.5	28	32

Choosing only from the summary statistics in the table, define a formula for a different statistic that measures skewness.

What values of that statistic might indicate that the distribution is skewed to the right? Explain.

2010

6. Hurricane damage amounts, in millions of dollars per acre, were estimated from insurance records for major hurricanes for the past three decades. A stratified random sample of five locations (based on categories of distance from the coast) was selected from each of three coastal regions in the southeastern United States. The three regions were Gulf Coast (Alabama, Louisiana, Mississippi), Florida, and Lower Atlantic (Georgia, South Carolina, North Carolina). Damage amounts in millions of dollars per acre, adjusted for inflation, are shown in the table below.

HURRICANE DAMAGE AMOUNTS IN MILLIONS OF DOLLARS PER ACRE

	Distance from Coast				
	< 1 mile	1 to 2 miles	2 to 5 miles	5 to 10 miles	10 to 20 miles
Gulf Coast	24.7	21.0	12.0	7.3	1.7
Florida	35.1	31.7	20.7	6.4	3.0
Lower Atlantic	21.8	15.7	12.6	1.2	0.3

(a) Sketch a graphical display that compares the hurricane damage amounts per acre for the three different coastal regions (Gulf Coast, Florida, and Lower Atlantic) and that also shows how the damage amounts vary with distance from the coast.

(b) Describe differences and similarities in the hurricane damage amounts among the three regions.

Because the distributions of hurricane damage amounts are often skewed, statisticians frequently use rank values to analyze such data.

(c) In the table below, the hurricane damage amounts have been replaced by the ranks 1, 2, or 3. For each of the distance categories, the highest damage amount is assigned a rank of 1 and the lowest damage amount is assigned a rank of 3. Determine the missing ranks for the 10-to-20-miles distance category and calculate the average rank for each of the three regions. Place the values in the table below.

ASSIGNED RANKS WITHIN DISTANCE CATEGORIES

	Distance from Coast					
	< 1 mile	1 to 2 miles	2 to 5 miles	5 to 10 miles	10 to 20 miles	Average Rank
Gulf Coast	2	2	3	1		
Florida	1	1	1	2		
Lower Atlantic	3	3	2	3		

(d) Consider testing the following hypotheses.

H_0: There is no difference in the distributions of hurricane damage amounts among the three regions.

H_a: There is a difference in the distributions of hurricane damage amounts among the three regions.

If there is no difference in the distribution of hurricane damage amounts among the three regions (Gulf Coast, Florida, and Lower Atlantic), the expected value of the average rank for each of the three regions is 2. Therefore, the following test statistic can be used to evaluate the hypotheses above:

$$Q = 5\left[\left(\bar{R}_G - 2\right)^2 + \left(\bar{R}_F - 2\right)^2 + \left(\bar{R}_A - 2\right)^2\right]$$

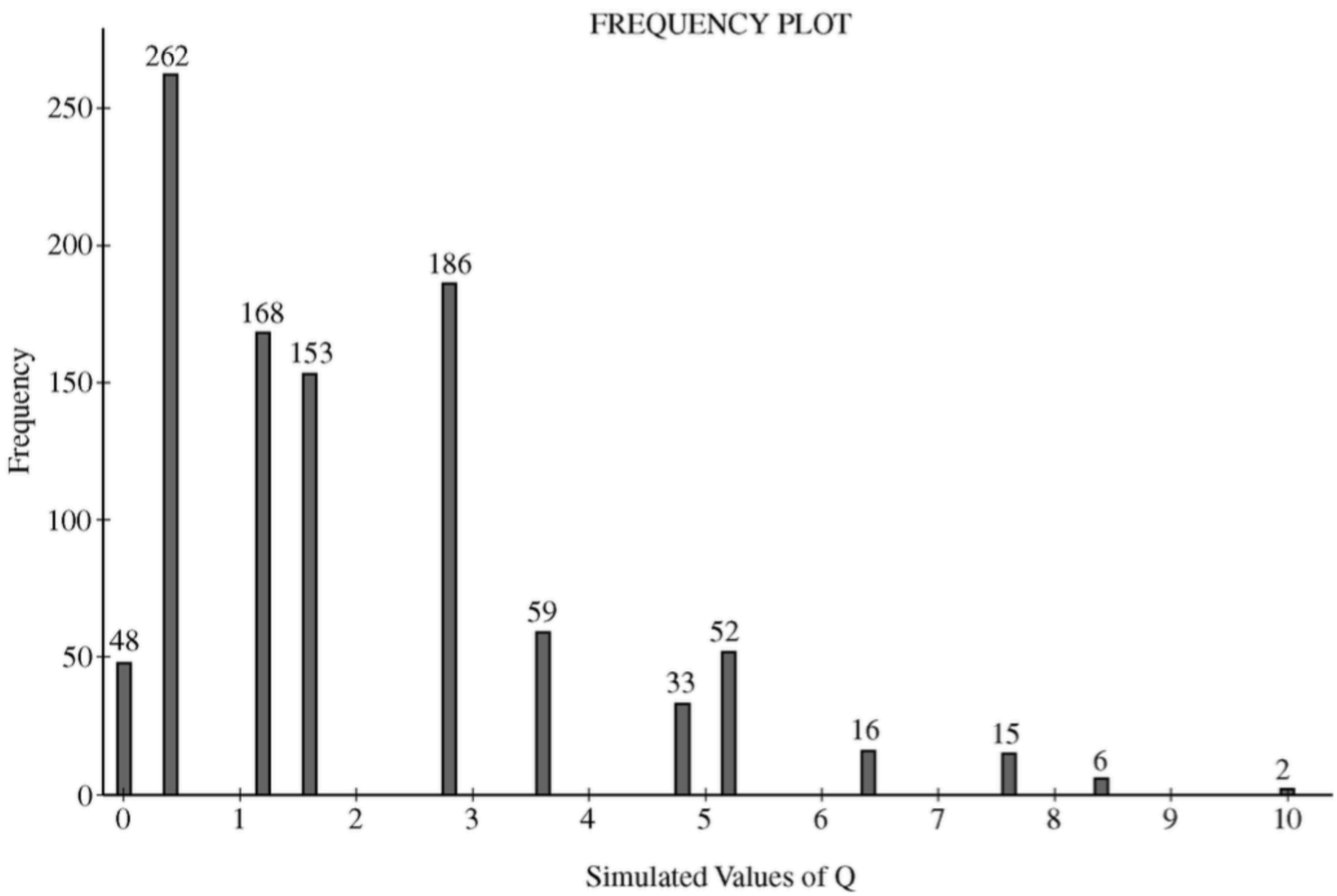
where $\bar{R}_G$ is the average rank over the five distance categories for the Gulf Coast (and $\bar{R}_F$ and $\bar{R}_A$ are similarly defined for the Florida and Lower Atlantic coastal regions).

Calculate the value of the test statistic Q using the average ranks you obtained in part (c).

(e) One thousand simulated values of this test statistic, Q, were calculated, assuming no difference in the distributions of hurricane damage amounts among the three coastal regions. The results are shown in the table below. These data are also shown in the frequency plot where the heights of the lines represent the frequency of occurrence of simulated values of Q.

Frequency Table for Simulated Values of Q

Q	Frequency	Cumulative Frequency	Percent	Cumulative Percent
0.0	48	48	4.80	4.80
0.4	262	310	26.20	31.00
1.2	168	478	16.80	47.80
1.6	153	631	15.30	63.10
2.8	186	817	18.60	81.70
3.6	59	876	5.90	87.60
4.8	33	909	3.30	90.90
5.2	52	961	5.20	96.10
6.4	16	977	1.60	97.70
7.6	15	992	1.50	99.20
8.4	6	998	0.60	99.80
10.0	2	1000	0.20	100.00



Use these simulated values and the test statistic you calculated in part (d) to determine if the observed data provide evidence of a significant difference in the distributions of hurricane damage amounts among the three coastal regions. Explain.

2011

6. Every year, each student in a nationally representative sample is given tests in various subjects. Recently, a random sample of 9,600 twelfth-grade students from the United States were administered a multiple-choice United States history exam. One of the multiple-choice questions is below. (The correct answer is C.)

> In 1935 and 1936 the Supreme Court declared that important parts of the New Deal were unconstitutional. President Roosevelt responded by threatening to
>
> (A) impeach several Supreme Court justices
> (B) eliminate the Supreme Court
> (C) appoint additional Supreme Court justices who shared his views
> (D) override the Supreme Court's decisions by gaining three-fourths majorities in both houses of Congress

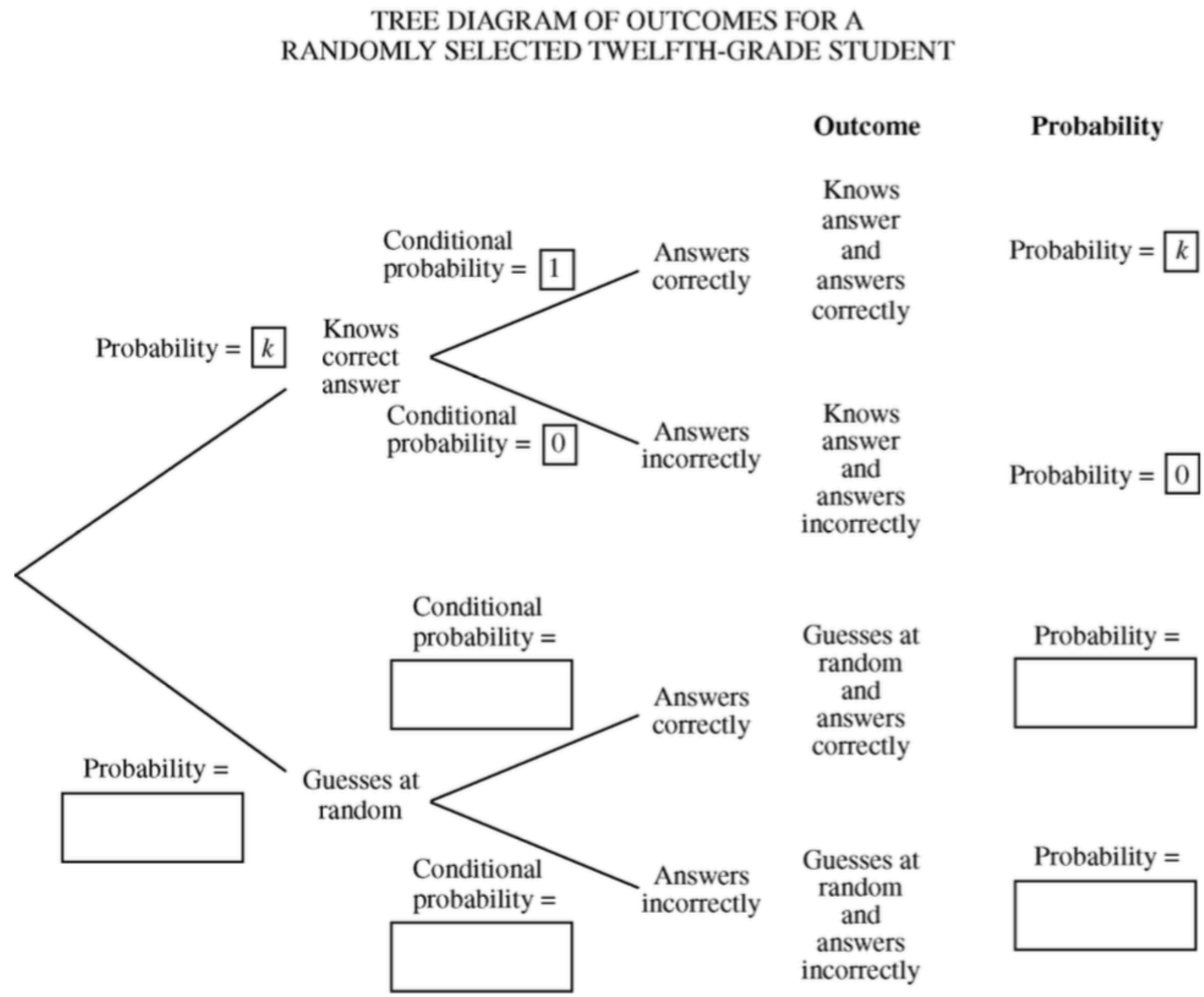
Of the 9,600 students, 28 percent answered the multiple-choice question correctly.

(a) Let p be the proportion of all United States twelfth-grade students who would answer the question correctly. Construct and interpret a 99 percent confidence interval for p.

Assume that students who actually know the correct answer have a 100 percent chance of answering the question correctly, and students who do not know the correct answer to the question guess completely at random from among the four options.

Let k represent the proportion of all United States twelfth-grade students who actually know the correct answer to the question.

(b) A tree diagram of the possible outcomes for a randomly selected twelfth-grade student is provided below. Write the correct probability in each of the five empty boxes. Some of the probabilities may be expressions in terms of k.



(c) Based on the completed tree diagram, express the probability, in terms of k, that a randomly selected twelfth-grade student would correctly answer the history question.

(d) Using your interval from part (a) and your answer to part (c), calculate and interpret a 99 percent confidence interval for k, the proportion of all United States twelfth-grade students who actually know the answer to the history question. You may assume that the conditions for inference for the confidence interval have been checked and verified.

2012

6. Two students at a large high school, Peter and Rania, wanted to estimate μ, the mean number of soft drinks that a student at their school consumes in a week. A complete roster of the names and genders for the 2,000 students at their school was available. Peter selected a simple random sample of 100 students. Rania, knowing that 60 percent of the students at the school are female, selected a simple random sample of 60 females and an independent simple random sample of 40 males. Both asked all of the students in their samples how many soft drinks they typically consume in a week.

   (a) Describe a method Peter could have used to select a simple random sample of 100 students from the school.

   Peter and Rania conducted their studies as described. Peter used the sample mean $\overline{X}$ as a point estimator for μ. Rania used $\overline{X}_{\text{overall}} = (0.6)\overline{X}_{\text{female}} + (0.4)\overline{X}_{\text{male}}$ as a point estimator for μ, where $\overline{X}_{\text{female}}$ is the mean of the sample of 60 females and $\overline{X}_{\text{male}}$ is the mean of the sample of 40 males.

   Summary statistics for Peter's data are shown in the table below.

Variable	N	Mean	Standard Deviation
Number of soft drinks	100	5.32	4.13

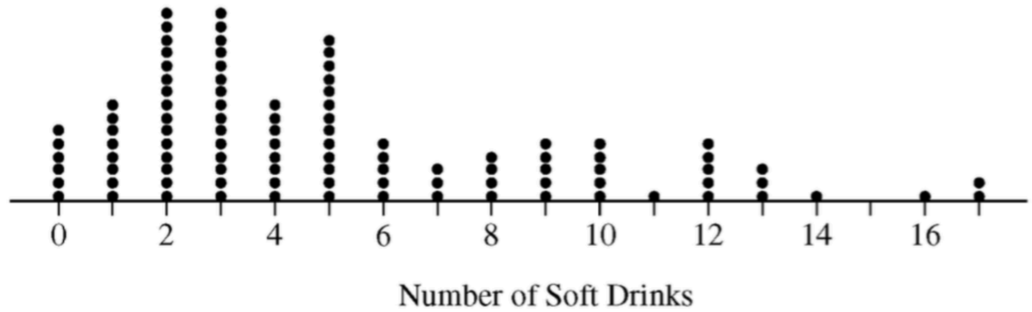
   (b) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution (sometimes called the standard error) of Peter's point estimator $\overline{X}$.

Summary statistics for Rania's data are shown in the table below.

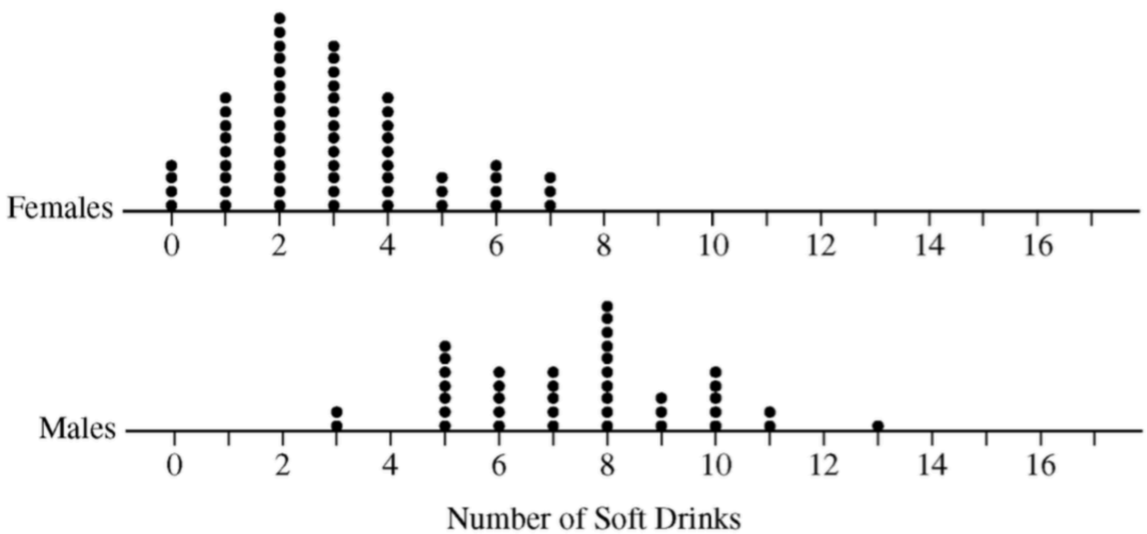
Variable	Gender	N	Mean	Standard Deviation
Number of soft drinks	Female	60	2.90	1.80
	Male	40	7.45	2.22

(c) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution of Rania's point estimator $\overline{X}_{\text{overall}} = (0.6)\overline{X}_{\text{female}} + (0.4)\overline{X}_{\text{male}}$.

A dotplot of Peter's sample data is given below.



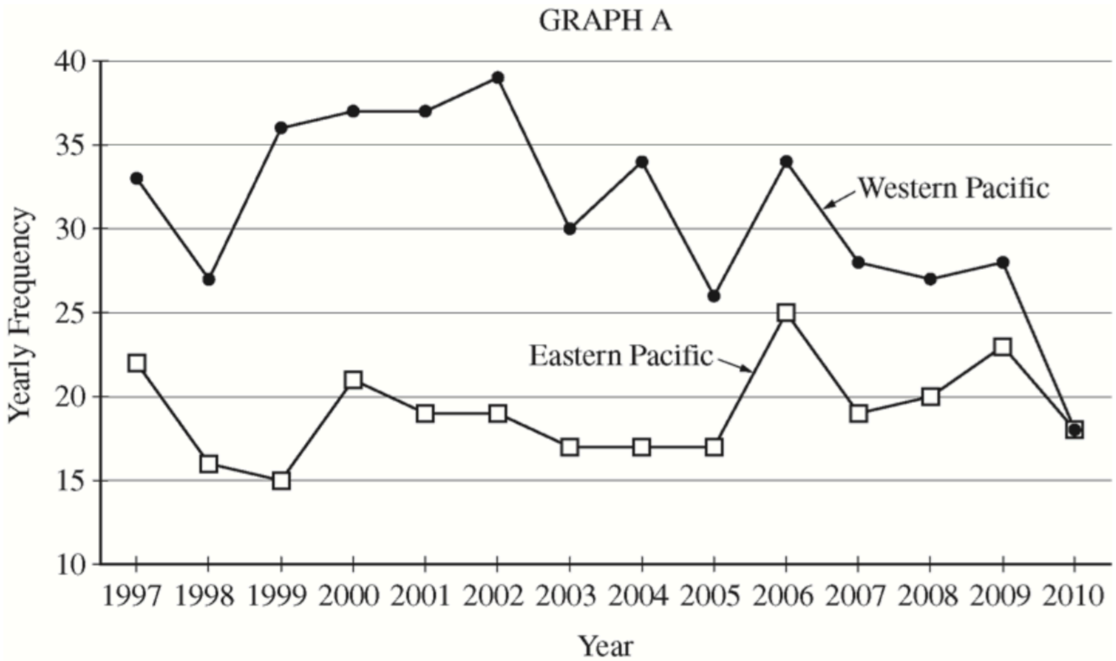
Comparative dotplots of Rania's sample data are given below.



(d) Using the dotplots above, explain why Rania's point estimator has a smaller estimated standard deviation than the estimated standard deviation of Peter's point estimator.

2013

6. Tropical storms in the Pacific Ocean with sustained winds that exceed 74 miles per hour are called typhoons. Graph A below displays the number of recorded typhoons in two regions of the Pacific Ocean—the Eastern Pacific and the Western Pacific—for the years from 1997 to 2010.



(a) Compare the distributions of yearly frequencies of typhoons for the two regions of the Pacific Ocean for the years from 1997 to 2010.

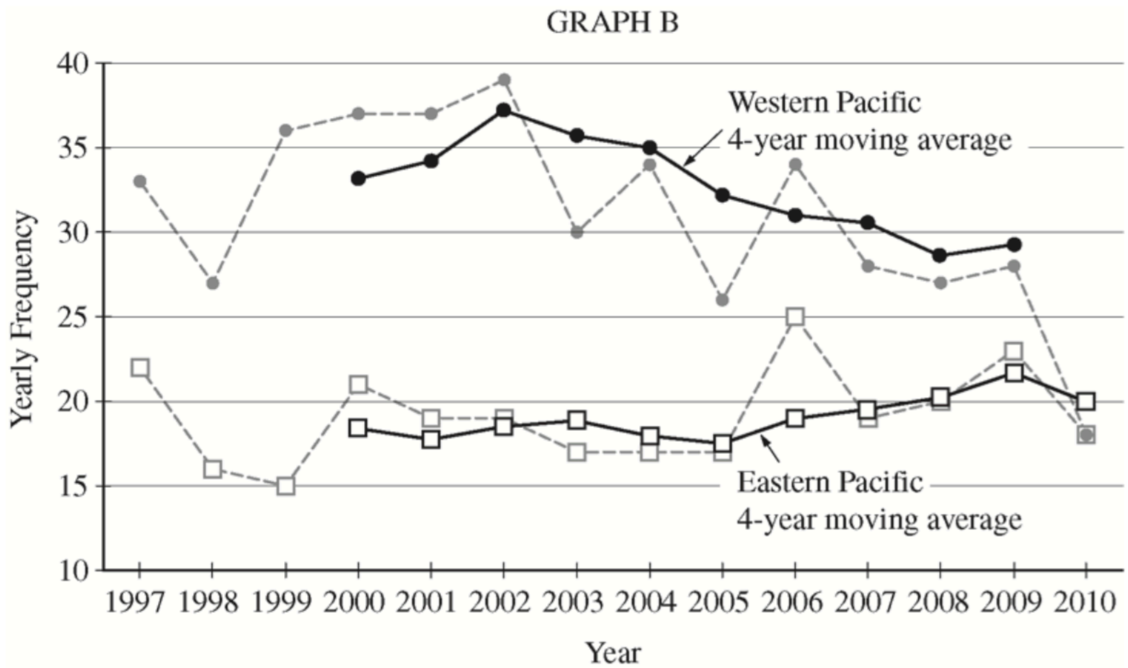
(b) For each region, describe how the yearly frequencies changed over the time period from 1997 to 2010.

A moving average for data collected at regular time increments is the average of data values for two or more consecutive increments. The 4-year moving averages for the typhoon data are provided in the table below. For example, the Eastern Pacific 4-year moving average for 2000 is the average of 22, 16, 15, and 21, which is equal to 18.50.

Year	Number of Typhoons in the Eastern Pacific	Eastern Pacific 4-year moving average	Number of Typhoons in the Western Pacific	Western Pacific 4-year moving average
1997	22		33	
1998	16		27	
1999	15		36	
2000	21	18.50	37	33.25
2001	19	17.75	37	34.25
2002	19	18.50	39	37.25
2003	17	19.00	30	35.75
2004	17	18.00	34	35.00
2005	17	17.50	26	32.25
2006	25	19.00	34	31.00
2007	19	19.50	28	30.50
2008	20	20.25	27	28.75
2009	23	21.75	28	29.25
2010	18	20.00	18	

(c) Show how to calculate the 4-year moving average for the year 2010 in the Western Pacific. Write your value in the appropriate place in the table.

(d) Graph B below shows both yearly frequencies (connected by dashed lines) and the respective 4-year moving averages (connected by solid lines). Use your answer in part (c) to complete the graph.



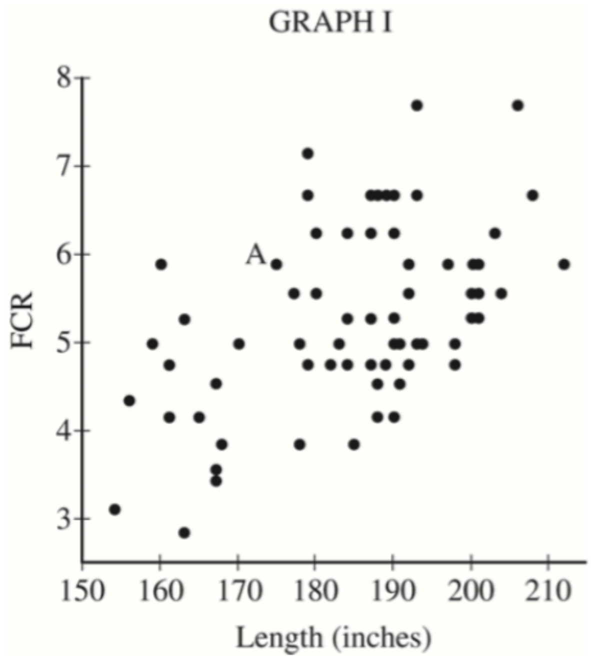
(e) Consider graph B.

i) What information is more apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

ii) What information is less apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

2014

6. Jamal is researching the characteristics of a car that might be useful in predicting the fuel consumption rate (FCR); that is, the number of gallons of gasoline that the car requires to travel 100 miles under conditions of typical city driving. The length of a car is one explanatory variable that can be used to predict FCR. Graph I is a scatterplot showing the lengths of 66 cars plotted with the corresponding FCR. One point on the graph is labeled A.



Jamal examined the scatterplot and determined that a linear model would be a reasonable way to express the relationship between FCR and length. A computer output from a linear regression is shown below.

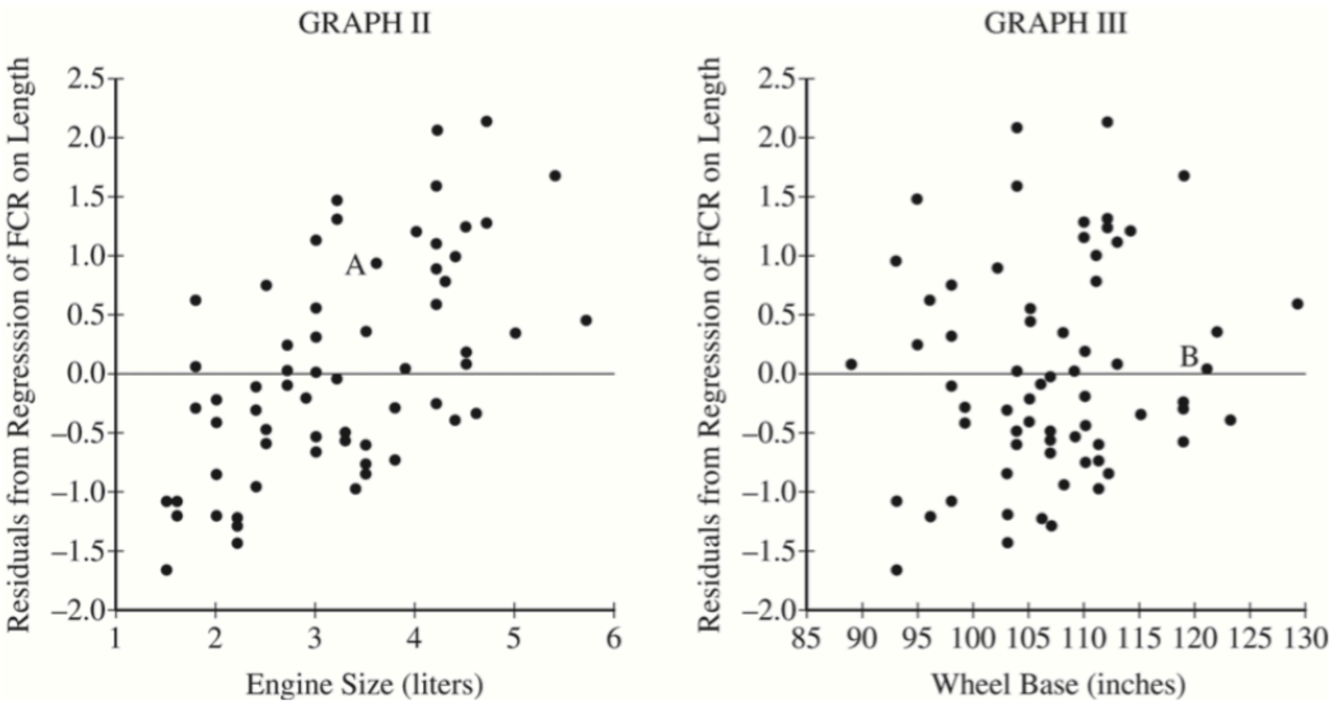
Linear Fit
FCR = −1.595789 + 0.0372614 * Length

Summary of Fit

RSquare	0.250401
Root Mean Square Error	0.902382
Observations	66

(a) The point on the graph labeled A represents one car of length 175 inches and an FCR of 5.88. Calculate and interpret the residual for the car relative to the least squares regression line.

Jamal knows that it is possible to predict a response variable using more than one explanatory variable. He wants to see if he can improve the original model of predicting FCR from length by including a second explanatory variable in addition to length. He is considering including engine size, in liters, or wheel base (the length between axles), in inches. Graph II is a scatterplot showing the engine size of the 66 cars plotted with the corresponding residuals from the regression of FCR on length. Graph III is a scatterplot showing the wheel base of the 66 cars plotted with the corresponding residuals from the regression of FCR on length.



(b) In graph II, the point labeled A corresponds to the same car whose point was labeled A in graph I. The measurements for the car represented by point A are given below.

FCR	Length (inches)	Engine Size (liters)	Wheel Base (inches)
5.88	175	3.6	93

(i) Circle the point on graph III that corresponds to the car represented by point A on graphs I and II.

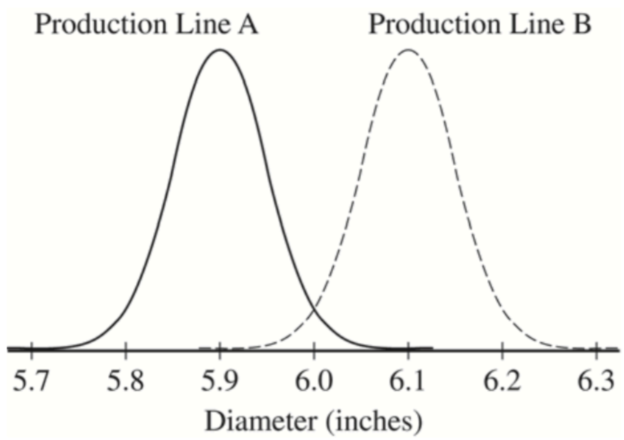
(ii) There is a point on graph III labeled B. It is very close to the horizontal line at 0. What does that indicate about the FCR of the car represented by point B?

(c) Write a few sentences to compare the association between the variables in graph II with the association between the variables in graph III.

(d) Jamal wants to predict FCR using length and one of the other variables, engine size or wheel base. Based on your response to part (c), which variable, engine size or wheel base, should Jamal use in addition to length if he wants to improve the prediction? Explain why you chose that variable.

2015

6. Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.



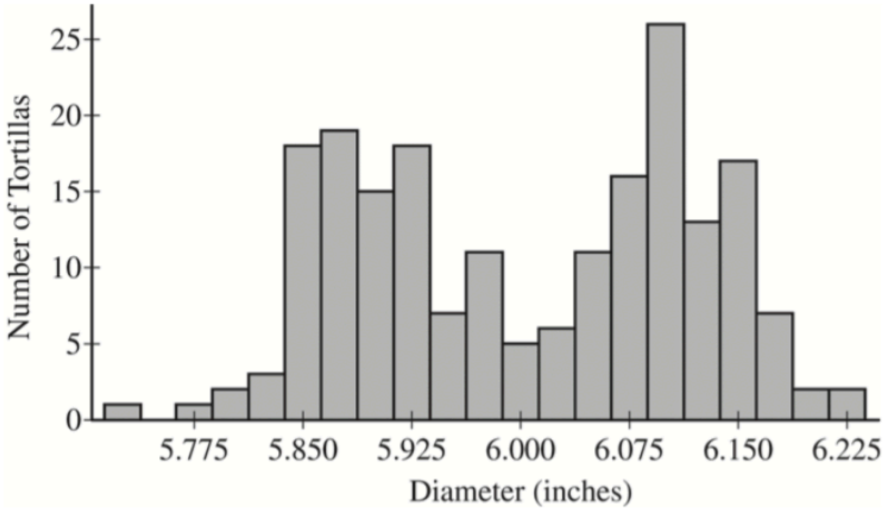
The tortillas produced at the factory are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected and the diameters are measured. From the sample of 200 tortillas, the manager of the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

<u>Method 1</u>: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

<u>Method 2</u>: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

(a) Will a sample obtained using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Explain why or why not.

(b) The figure below is a histogram of 200 diameters obtained by using one of the two sampling methods described. Considering the shape of the histogram, explain which method, Method 1 or Method 2, was most likely used to obtain a such a sample.



(c) Which of the two sampling methods, Method 1 or Method 2, will result in less variability in the diameters of the 200 tortillas in the sample on a given day? Explain.

Each day, the distribution of the 200,000 tortillas made that day has mean diameter 6 inches with standard deviation 0.11 inch.

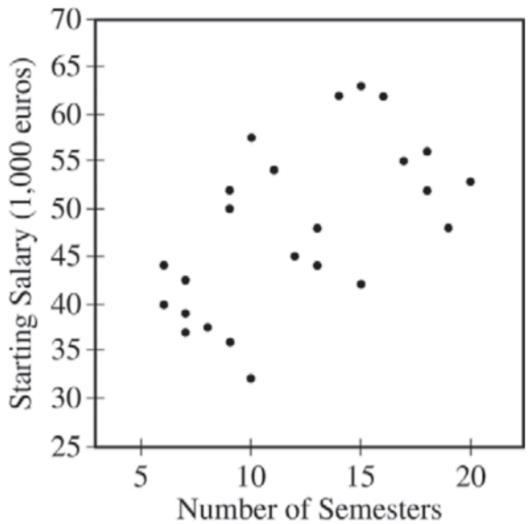
(d) For samples of size 200 taken from one day's production, describe the sampling distribution of the sample mean diameter for samples that are obtained using Method 1.

(e) Suppose that one of the two sampling methods will be selected and used every day for one year (365 days). The sample mean of the 200 diameters will be recorded each day. Which of the two methods will result in less variability in the distribution of the 365 sample means? Explain.

(f) A government inspector will visit the facility on June 22 to observe the sampling and to determine if the factory is in compliance with the advertised mean diameter of 6 inches. The manager knows that, with both sampling methods, the sample mean is an unbiased estimator of the population mean. However, the manager is unsure which method is more likely to produce a sample mean that is close to 6 inches on the day of sampling. Based on your previous answers, which of the two sampling methods, Method 1 or Method 2, is more likely to produce a sample mean close to 6 inches? Explain.

2016

6. A newspaper in Germany reported that the more semesters needed to complete an academic program at the university, the greater the starting salary in the first year of a job. The report was based on a study that used a random sample of 24 people who had recently completed an academic program. Information was collected on the number of semesters each person in the sample needed to complete the program and the starting salary, in thousands of euros, for the first year of a job. The data are shown in the scatterplot below.



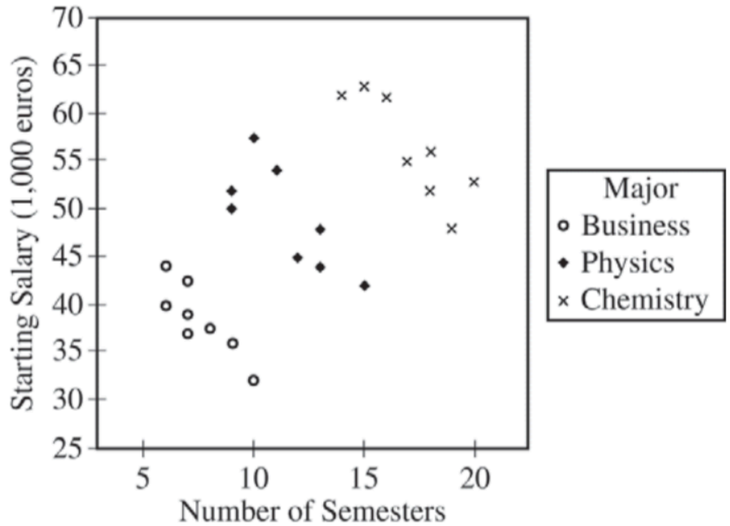
(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

The table below shows computer output from a linear regression analysis on the data.

Predictor	Coef	SE Coef	T	P
Constant	34.018	4.455	7.64	0.000
Semesters	1.1594	0.3482	3.33	0.003
S = 7.37702	R-Sq = 33.5%		R-Sq(adj) = 30.5%	

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below.



(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the business majors.

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

2017

6. Consider an experiment in which two men and two women will be randomly assigned to either a treatment group or a control group in such a way that each group has two people. The people are identified as Man 1, Man 2, Woman 1, and Woman 2. The six possible arrangements are shown below.

Arrangement A	
Treatment	Control
Man 1 Man 2	Woman 1 Woman 2

Arrangement B	
Treatment	Control
Man 1 Woman 1	Man 2 Woman 2

Arrangement C	
Treatment	Control
Man 1 Woman 2	Man 2 Woman 1

Arrangement D	
Treatment	Control
Woman 1 Woman 2	Man 1 Man 2

Arrangement E	
Treatment	Control
Man 2 Woman 2	Man 1 Woman 1

Arrangement F	
Treatment	Control
Man 2 Woman 1	Man 1 Woman 2

Two possible methods of assignment are being considered: the sequential coin flip method, as described in part (a), and the chip method, as described in part (b). For each method, the order of the assignment will be Man 1, Man 2, Woman 1, Woman 2.

(a) For the sequential coin flip method, a fair coin is flipped until one group has two people. An outcome of tails assigns the person to the treatment group, and an outcome of heads assigns the person to the control group. As soon as one group has two people, the remaining people are automatically assigned to the other group.

(i) Complete the table below by calculating the probability of each arrangement occurring if the sequential coin flip method is used.

Arrangement	A	B	C	D	E	F
Probability						

(ii) For the sequential coin flip method, what is the probability that Man 1 and Man 2 are assigned to the same group?

The six arrangements are repeated below.

Arrangement A	
Treatment	Control
Man 1	Woman 1
Man 2	Woman 2

Arrangement B	
Treatment	Control
Man 1	Man 2
Woman 1	Woman 2

Arrangement C	
Treatment	Control
Man 1	Man 2
Woman 2	Woman 1

Arrangement D	
Treatment	Control
Woman 1	Man 1
Woman 2	Man 2

Arrangement E	
Treatment	Control
Man 2	Man 1
Woman 2	Woman 1

Arrangement F	
Treatment	Control
Man 2	Man 1
Woman 1	Woman 2

(b) For the chip method, two chips are marked "treatment" and two chips are marked "control." Each person selects one chip at random without replacement.

(i) Complete the table below by calculating the probability of each arrangement occurring if the chip method is used.

Arrangement	A	B	C	D	E	F
Probability						

(ii) For the chip method, what is the probability that Man 1 and Man 2 are assigned to the same group?

(c) Sixteen participants consisting of 10 students and 6 teachers at an elementary school will be used for an experiment to determine lunch preference for the school population of students and teachers. As the participants enter the school cafeteria for lunch, they will be randomly assigned to receive one of two lunches so that 8 will receive a salad, and 8 will receive a grilled cheese sandwich. The students will enter the cafeteria first, and the teachers will enter next. Which method, the sequential coin flip method or the chip method, should be used to assign the treatments? Justify your choice.

2018

6. Systolic blood pressure is the amount of pressure that blood exerts on blood vessels while the heart is beating. The mean systolic blood pressure for people in the United States is reported to be 122 millimeters of mercury (mmHg) with a standard deviation of 15 mmHg.

   The wellness department of a large corporation is investigating whether the mean systolic blood pressure of its employees is greater than the reported national mean. A random sample of 100 employees will be selected, the systolic blood pressure of each employee in the sample will be measured, and the sample mean will be calculated.

   Let μ represent the mean systolic blood pressure of all employees at the corporation. Consider the following hypotheses.

$$H_0 : \mu = 122$$
$$H_a : \mu > 122$$

   (a) Describe a Type II error in the context of the hypothesis test.

   (b) Assume that σ, the standard deviation of the systolic blood pressure of all employees at the corporation, is 15 mmHg. If $\mu = 122$, the sampling distribution of $\overline{x}$ for samples of size 100 is approximately normal with a mean of 122 mmHg and a standard deviation of 1.5 mmHg. What values of the sample mean $\overline{x}$ would represent sufficient evidence to reject the null hypothesis at the significance level of $\alpha = 0.05$?

   The actual mean systolic blood pressure of all employees at the corporation is 125 mmHg, not the hypothesized value of 122 mmHg, and the standard deviation is 15 mmHg.

   (c) Using the actual mean of 125 mmHg and the results from part (b), determine the probability that the null hypothesis will be rejected.

   (d) What statistical term is used for the probability found in part (c) ?

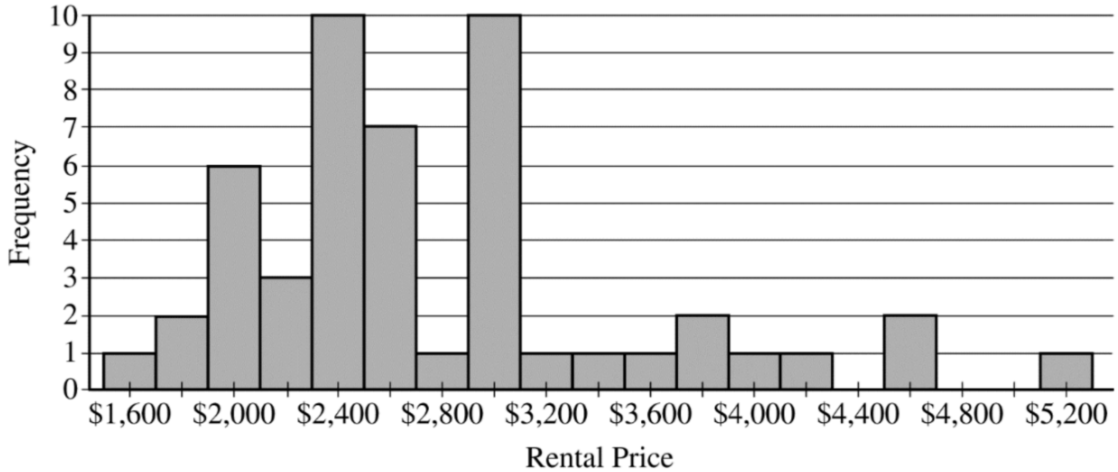
   (e) Suppose the size of the sample of employees to be selected is greater than 100. Would the probability of rejecting the null hypothesis be greater than, less than, or equal to the probability calculated in part (c) ? Explain your reasoning.

2019

6. Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

   (a) Describe the population for which it is appropriate for Emma to generalize the results from her sample.

   The distribution of the 50 rental prices of the available apartments is shown in the following histogram.



   (b) Emma wants to estimate the typical rental price of a one-bedroom apartment in the city. Based on the distribution shown, what is a disadvantage of using the mean rather than the median as an estimate of the typical rental price?

   (c) Instead of using the sample median as the point estimate for the population median, Emma wants to use an interval estimate. However, computing an interval estimate requires knowing the sampling distribution of the sample median for samples of size 50. Emma has one point, her sample median, in that sampling distribution. Using information about rental prices that are available on the Web site, describe how someone could develop a theoretical sampling distribution of the sample median for samples of size 50.

   Because Emma does not have the resources to develop the theoretical sampling distribution, she estimates the sampling distribution of the sample median using a process called bootstrapping. In the bootstrapping process, a computer program performs the following steps.

   - Take a random sample, with replacement, of size 50 from the original sample.
   - Calculate and record the median of the sample.
   - Repeat the process to obtain a total of 15,000 medians.

Emma ran the bootstrap process, and the following frequency table is the bootstrap distribution showing her results of generating 15,000 medians.

Bootstrap Distribution of Medians					
Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65
2,497.5	136	2,735	35	2,985	12
2,500	1,899	2,747.5	619	2,987.5	1
2,522.5	2	2,750	2	2,995	6
2,525	945	2,795	278	3,000	2
2,550	1,673	2,812.5	16	3,062.5	3

The bootstrap distribution provides an approximation of the sampling distribution of the sample median. A confidence interval for the median can be constructed using a percentage of the values in the middle of the bootstrap distribution.

(d) Use the frequency table to find the following.

(i) Value of the 5th percentile:

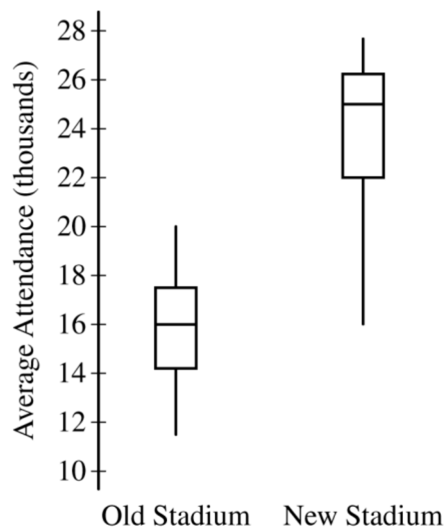
(ii) Value of the 95th percentile:

(e) Find the percentage of bootstrap medians in the table that are equal to or between the values found in part (d).

(f) Use your values from parts (d) and (e) to construct and interpret a confidence interval for the median rental price.

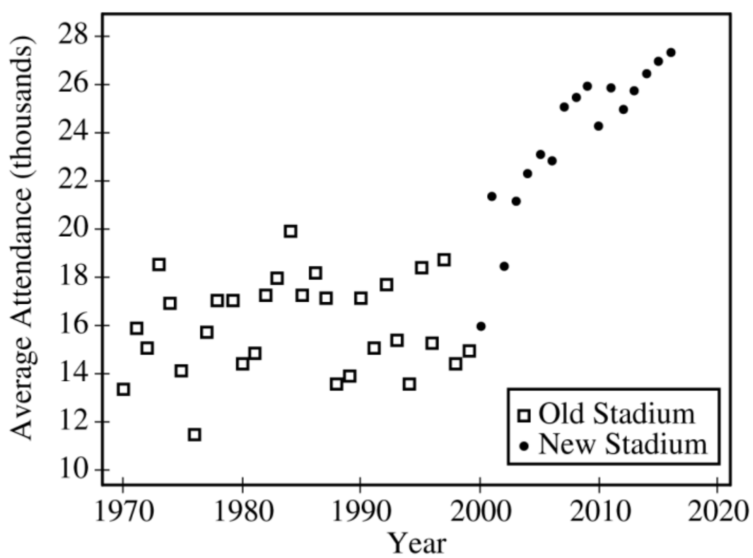
2021

Attendance at games for a certain baseball team is being investigated by the team owner. The following boxplots summarize the attendance, measured as average number of attendees per game, for 47 years of the team's existence. The boxplots include the 30 years of games played in the old stadium and the 17 years played in the new stadium.



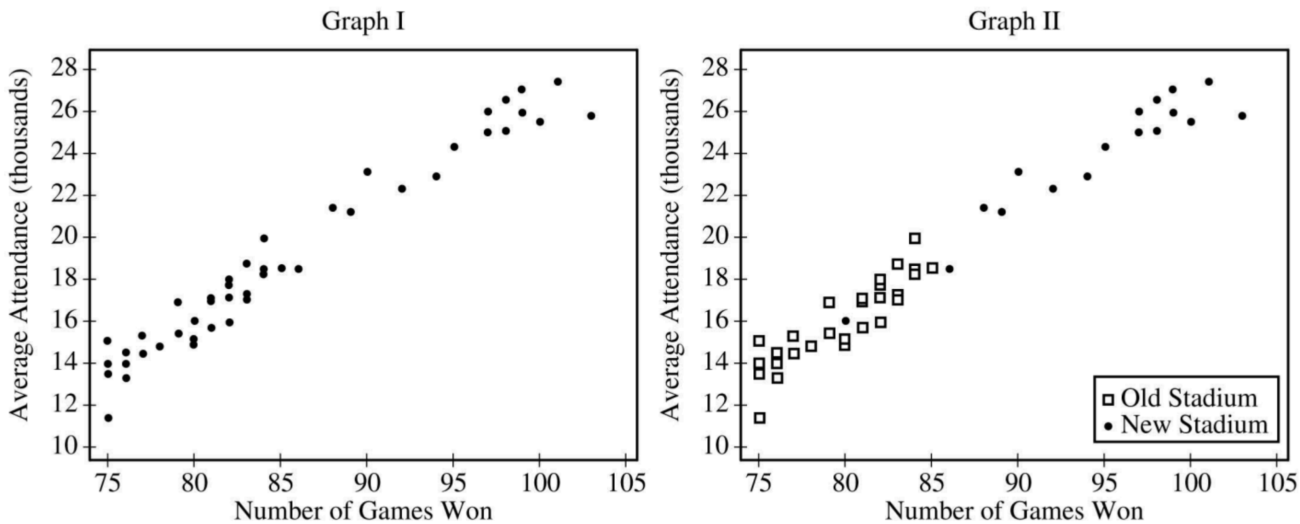
(a) Compare the distributions of average attendance between the old and new stadiums.

The following scatterplot shows average attendance versus year.



(b) Compare the trends in average attendance over time between the old and new stadium.

(c) Consider the following scatterplots.



(i) Graph I shows the average attendance versus number of games won for each year. Describe the relationship between the variables.

(ii) Graph II shows the same information as Graph I, but also indicates the old and new stadiums. Does Graph II suggest that the rate at which attendance changes as number of games won increases is different in the new stadium compared to the old stadium? Explain your reasoning.

(d) Consider the three variables: number of games won, year, and stadium. Based on the graphs, explain how one of those variables could be a confounding variable in the relationship between average attendance and the other variables.

AP STATISTICS – LAST MINUTE CRAM SHEET

If you're going to cram, here are some good ideas.

SCRIPTS/INTERPRETATIONS

1. Standard deviation – *If we randomly select many, many* ____________ *and calculate the* __________ *(average, proportion, etc.), we will typically (or "on average") be off by/within* _____ *of the true* ________________ *(average, proportion, etc.*).

Example: A factory randomly samples 100 tortillas and calculates the average diameter. Suppose the sampling distribution of the sample mean has a standard deviation of 0.13 inch.
Interpretation of 0.13: *If we randomly sampled 100 tortillas many times and calculated the average diameter, we will typically be off by (within) about 0.13 inch of the true average tortilla diameter for all tortillas in the factory.*

2. Kinds of samples
Simple random – no pre-grouping
Stratified – strategically break into <u>different</u> groups, get <u>some</u> individuals from <u>each group</u> (number individuals in *every* strata)
Cluster – break into <u>similar</u> groups, get <u>all</u> individuals from <u>some</u> groups (number *the clusters*, use all individuals from selected clusters)

3. Kinds of experiments
Completely randomized – no pre-grouping
Blocking – strategically break into <u>different</u> groups, then randomly assign <u>within</u> each group
- pick variable that you expect will be associated with the response variable

Matched pairs – blocking experiment where blocks are of size 1 or 2

4. Scope of inference
What aspect of a study allows you to infer results to a <u>larger population</u>? Random SAMPLING.
Why? Sample should be <u>representative</u> of the population.
What aspect of a study allows you to conclude <u>causation</u>? Random ASSIGNMENT.
Why? Should create groups that are similar in every way that you can't directly control, so you can attribute any differences to the <u>one</u> variable being manipulated.

5. "In a few sentences, compare/contrast/describe the distributions."
Use <u>comparison</u> words (more, less, larger, smaller).

SHAPE: Left skewed, right skewed, approximately normal, roughly symmetric, unimodal, bimodal
CENTER:
- *MEDIAN* if given: stemplot, dotplot, boxplot, histogram (give approximate or find the class it's in)
- *MEAN* if given: density curve, sampling distribution

VARIABILITY:
- BOXPLOT: IQR, range, outliers (remember that outliers DO count for range, if you're going to use that)
- DOTPLOT/STEMPLOT: Where is data clustered? Where are the most extreme values? Are there any gaps? Range *maybe.*
- HISTOGRAM: Where is data clustered? Where are the most extreme values? Are there any gaps? *You can't calculate range or IQR.*

6. r^2 – ___*% of the variability in the y-variable (in context) can be accounted for by the least-squares regression line relating the y-variable to the x-variable*.

Example: Suppose a least-squares regression line is formed to model the relationship between engine size and fuel consumption rate of 50 cars and trucks. The hypothesis is that engine size can help predict fuel consumption rate. The r value is 0.8 and $r^2 = 0.64$. Interpret $r^2 = 0.64$: *64% of the variability in fuel consumption rate can be accounted for by the least-squares regression line relating fuel consumption rate to engine size.*

7. $\boldsymbol{r}$ – Tells the strength and direction *of a linear relationship*, but not *whether* the form is linear.

For the example in #2 about engine size and fuel consumption rate, interpret $r = 0.8$ (this is $\sqrt{r^2}$).
There is a strong, positive relationship between fuel consumption rate and engine size.

8. Confidence <u>interval</u> – *We are ___% confident that the interval from _____ to ______ (units) captures the true ________________ for the population of ______________.*

9. Confidence <u>level</u> – *In repeated samplings of this same size, ____% of the constructed ____% confidence intervals would capture the true _____________ (parameter in context).*

Example: A sample of 9,600 Americans looks to estimate the proportion of adults who can name at least 3 Supreme Court justices. A 95% confidence interval about the sample proportion is 0.24 ± 0.03.

Interpret the <u>interval</u>: *We are 95% confident that the interval from 0.21 to 0.27 captures the true proportion of <u>all</u> U.S. adults who can name at least 3 Supreme Court Justices.*

Interpret the confidence <u>level</u>: *If we took many samples of 9,600 Americans and created a 95% interval about each sample proportion who can name at least 3 Supreme Court Justices, 95% of these intervals would capture the true proportion for <u>all</u> U.S. adults.*

10. P-value – *If the null hypothesis is true, the probability that we would see a statistic this extreme or more (in the direction of H_a) due to chance is ______.*

Example: Researchers believe that, when driving, adults are more likely to miss an exit if they are talking on the phone than they are if they are simply talking to a passenger. They randomly assign 50 adults to talk on the phone while using a driving simulator and another 50 adults to talk to a passenger while using a driving simulator. 18 of the adults talking on the phone miss the exit, while 12 of the adults talking to a passenger miss the exit. The hypothesis are $H_0: p_{phone} - p_{pass} = 0$ versus $H_a: p_{phone} - p_{pass} > 0$. The P-value for a two-proportion z-test is P = 0.095.

Interpret the P-value of 0.095: (Note: The leg work here is $\hat{p}_{phone} - \hat{p}_{pass} = \frac{18}{50} - \frac{12}{50} = 0.12$.)
If there is no difference in the proportions of adults who would miss an exit while talking on the phone and who would miss an exit while talking to a passenger, the probability we would get a statistic as high or higher than 0.12 by chance is 0.095.

11. Concluding a hypothesis test
Because P = _____ >/< _____ = α, we fail to reject/reject the <u>null</u> hypothesis (in context). We do not/do have convincing evidence of the <u>alternative</u> hypothesis (in context).

To help you keep them straight:
$P < \alpha$ → *Reject* H_0 → *Convinced* of H_a → *Statistically significant* evidence of H_a
$P > \alpha$ → *Fail to reject* H_0 → *Not convinced* of H_a → *Not statistically significant* evidence of H_a

12. Power
If this <u>specific value of the alternative is true</u>, there is a ____% chance that we will <u>correctly reject the null hypothesis</u>.

Example: A quality control manager at an apple sauce company must turn away a truck if more than 8% of the apples on the truck are defective. He randomly samples some apples and calculates what percent have defects. The power of the test against $p = 0.12$ is 0.84.
Interpret the power of 0.84: *If actually 12% of the apples on the truck have defects, there is an 84% chance the manager will correctly reject the null hypothesis and turn away the truck.*