STATISTICS
SECTION II
Part A
Questions 1-5

Spend about 1 hour and 5 minutes on this part of the exam.

Percent of Section II score—75

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. Thirty-four college students were asked how much money they spent on textbooks for the current semester. Their responses are shown in the following stemplot.

1	2 3 3 4 5 5 6 7 8
2	1 2 3 4 5 6 8 8 9 9 9
3	1 2 2 7 8 9
4	1 4 5 7
5	1 3
6	2
7	
8	1

Key: 1 | 2 = $120

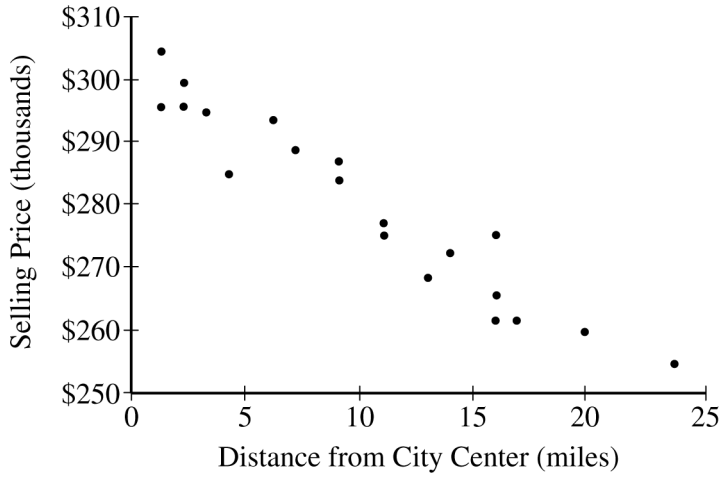
(a) Describe a procedure for identifying potential outliers, and use the procedure to decide whether there are outliers among the responses for the money spent on textbooks.

GO ON TO THE NEXT PAGE.

(b) Based on the stemplot, write a few sentences describing the distribution of money spent on textbooks for the 34 students.

GO ON TO THE NEXT PAGE.

2. A real estate agent working in a large city believes that, for three-bedroom houses, the selling price of the house decreases by approximately $2,000 for every mile increase in the distance of the house from the city center. To investigate the belief, the agent obtained a random sample of 20 three-bedroom houses that sold in the last year. The selling price, in thousands of dollars, and the distance from the city center, in miles, for each of the 20 houses are shown in the scatterplot. The table shows computer output from a regression analysis of the data.



Predictor	Coef	SE Coef	T	P
Constant	301.7	1.80	167.17	0.000
Distance	–2.158	0.149	–14.45	0.000
	S = 4.4336		R-sq = 92.1%	

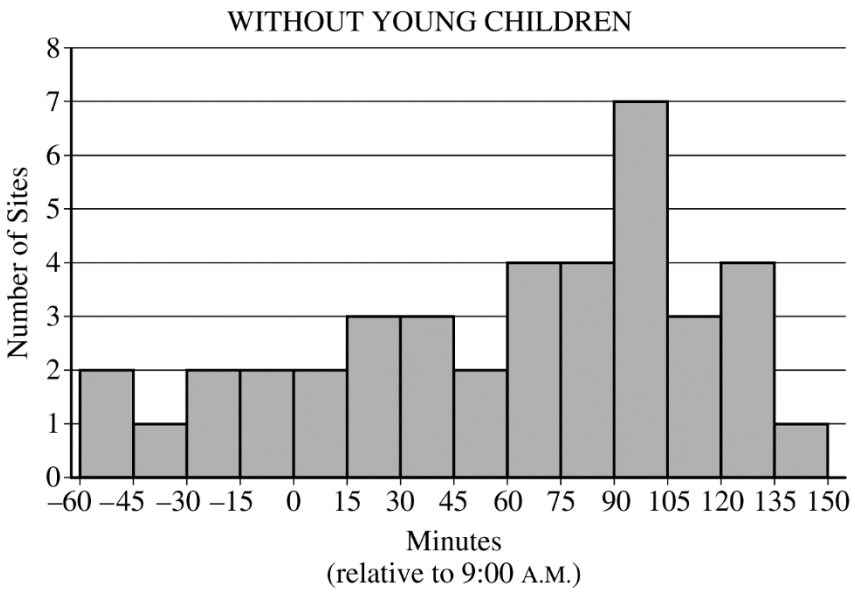
(a) Assume all conditions for inference are met. Construct and interpret a 95 percent confidence interval for the slope of the least-squares regression line.

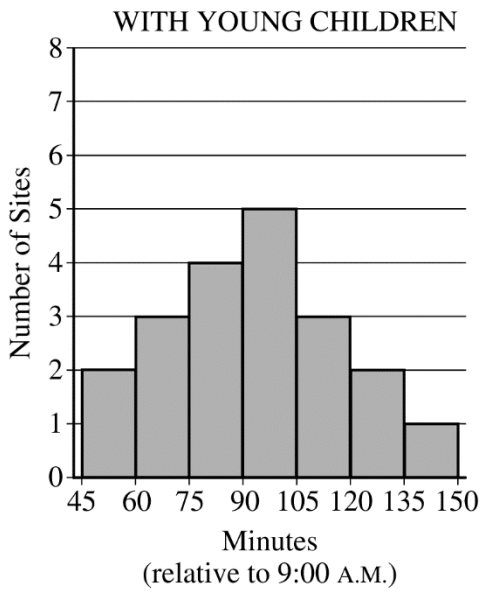
GO ON TO THE NEXT PAGE.

(b) Does the confidence interval contradict the agent's belief about the relationship between selling price and distance from the city center? Justify your answer.

3. River Run campground has sites for people to use for camping. The sites can be reserved for a certain number of days. To help with cleaning and maintenance, the campground requests an exit time (the time at which campers leave the site) of 9 A.M. on the last day of the reservation.

   To estimate the typical exit time, the manager of River Run selected a random sample of 60 sites. Of the selected sites, 40 were reserved by people without young children, and 20 were reserved by people with young children. The following histograms summarize the exit times, recorded as minutes relative to 9 A.M. For example, an exit time of 9:30 A.M. is 30 minutes relative to an exit time of 9 A.M. Each interval contains possible values from the left endpoint up to but not including the right endpoint.





(a) Consider the two histograms.

(i) How many of the 60 sites had an exit time before 8:30 A.M.?

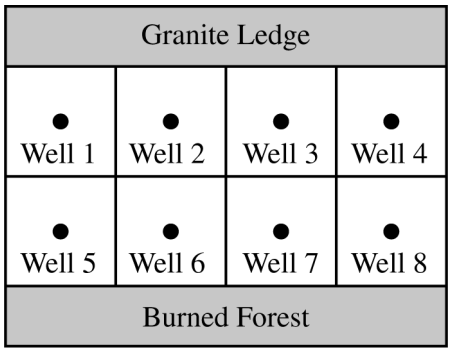
(ii) How many of the 60 sites had an exit time of 11:00 A.M. or later?

(b) Compare the distributions of the exit times for those without young children and those with young children.

(c) Based on the histograms, what is a reasonable estimate of the median exit time for the random sample of 60 sites? Explain your reasoning.

4. Arsenic is a naturally occurring chemical that can enter groundwater through eroding granite or from a burned forest. A health organization recommends drinking water should contain no more than 10 parts per billion (ppb) of arsenic. A company produces filters to clean arsenic from private wells that could be affected by the groundwater.

   The company wants to investigate the effectiveness of a new filter compared to that of an older filter. They will test the filters on a field that is bordered on one side by a granite ledge and on the other side by a burned forest. The field is divided into 8 square plots of equal size, and a well to collect groundwater is drilled in the center of each plot. One filter will be used in each well. The following diagram shows the placement of the wells in the field.



   The company will use four of each type of filter to conduct the investigation. A randomized block design will be used.

   (a) Assuming there is a difference in the effectiveness of the two filters, under what conditions will a randomized block design be better for detecting the difference than a completely randomized design?

(b) Identify the wells, by number, that will be included in each block.

(c) Describe how to assign filters to wells to create a randomized block design.

GO ON TO THE NEXT PAGE.

5. For each day that Sasha travels to work, the probability that she will experience a delay due to traffic is 0.2. Each day can be considered independent of the other days.

   (a) For the next 21 days that Sasha travels to work, what is the probability that Sasha will experience a delay due to traffic on at least 3 of the days?

   (b) What is the probability that Sasha's first delay due to traffic will occur after the fifth day of travel to work?

(c) Consider a random sample of 21 days that Sasha will travel to work. For the proportion of those days that she will experience a delay due to traffic, is the sampling distribution of the sample proportion approximately normal? Justify your answer.

STATISTICS

SECTION II

Part B

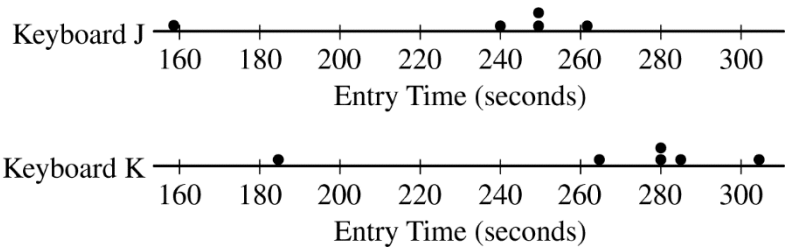
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Researchers are studying two different designs of computer keyboards, J and K, to investigate the effectiveness of the design on the speed of data entry. The researchers believe there is a tendency for people entering data with keyboard J to have faster entry times compared with people entering data with keyboard K. Using the same set of data for entry, the researchers randomly assigned 5 people to keyboard J and 6 people to keyboard K and recorded the number of seconds each person took to enter the data. The following dotplots show the observed entry times for the two keyboards.



(a) Explain why it is <u>not</u> appropriate to conduct a two-sample *t*-test for the difference in population means.

(b) Based on the dotplots, explain why it might be more appropriate to compare population medians instead of population means.

One test used to compare population medians is the Wilcoxon Rank Sum Test. Under the assumption that the shape and variability of the distributions are the same, the test uses the rankings of the combined observed values. To conduct the test, the entry times for keyboards J and K are combined into one group and then ranked from 1 to n_T, the total number of observed values in the combined group. The observed entry times, in seconds, for both keyboard types are shown in the following table.

	Observed Entry Times					
J	158	240	248	251	261	
K	184	267	279	280	284	305

(c) Consider the observed entry times for keyboards J and K.

(i) Complete the following table to assign ranks to the observed entry times for keyboards J and K combined.

Rank	1	2	3	4	5	6	7	8	9	10	11
Keyboard	J	K	J							K	K
Time	158	184	240							284	305

(ii) Use the completed table in (i) to calculate the sum of the ranks assigned to each keyboard.

Sum of ranks for J (SR_J):

Sum of ranks for K (SR_K):

The hypotheses for the Wilcoxon Rank Sum Test are as follows.

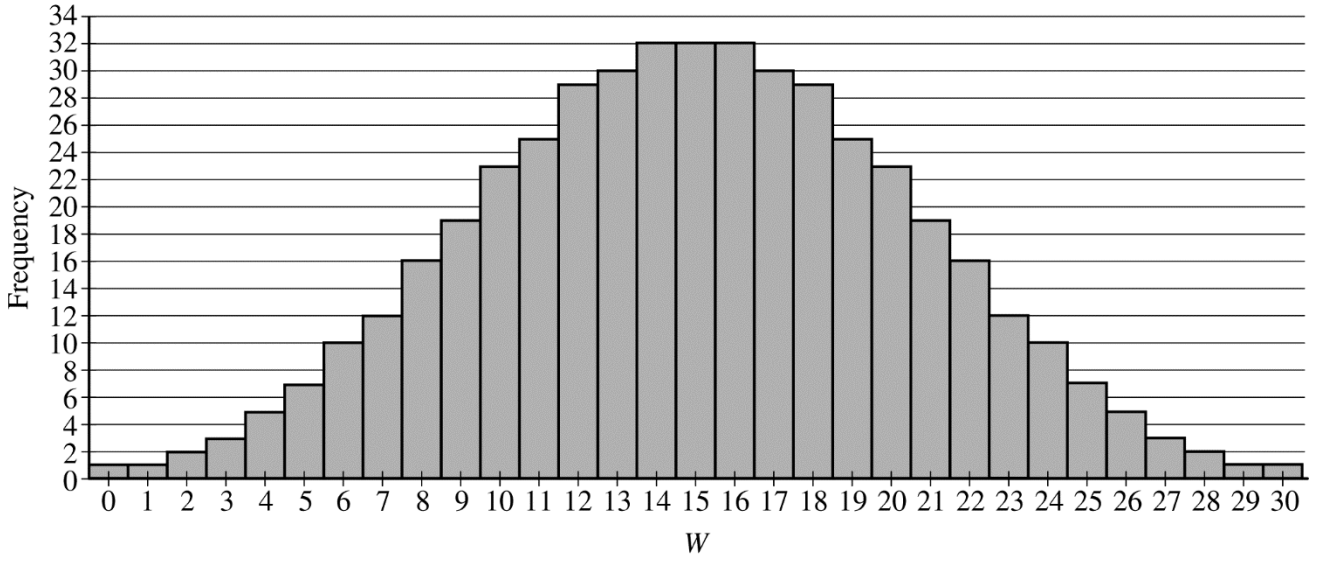
H_0 : The median of the distribution of entry times for all users of keyboard J and the median of the distribution of entry times for all users of keyboard K are the same.

H_a : The median of the distribution of entry times for all users of keyboard J is less than the median of the distribution of entry times for all users of keyboard K.

The test statistic W for the test is $W = SR_J - \frac{n_J(n_J + 1)}{2}$, where n_J is the number of observations for keyboard J.

(d) Use the formula to calculate the test statistic W for the keyboard data.

There are 462 possible assignments of 11 ranks to 5 Js and 6 Ks. If the null hypothesis is true, the 462 assignments are equally likely. The following graph shows the sampling distribution of all possible values of W resulting from the 462 assignments.



(e) The least possible value of W in the sampling distribution is 0.

(i) Find the value of SR_J for $W = 0$.

(ii) Assign 5 ranks to keyboard J and 6 ranks to keyboard K so that $W = 0$. Show your assignments by completing the following table.

Rank	1	2	3	4	5	6	7	8	9	10	11
Keyboard											