STATISTICS

SECTION II

Part A

Questions 1-5

Spend about 65 minutes on this part of the exam.

Percent of Section II grade—75

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

1. The summary statistics for the number of inches of rainfall in Los Angeles for 117 years, beginning in 1877, are shown below.

N	MEAN	MEDIAN	TRMEAN	STDEV	SE MEAN
117	14.941	13.070	14.416	6.747	0.624

MIN	MAX	Q1	Q3
4.850	38.180	9.680	19.250

   (a) Describe a procedure that uses these summary statistics to determine whether there are outliers.

   (b) Are there outliers in these data?   _____

   Justify your answer based on the procedure that you described in part (a).

   (c) The news media reported that in a particular year, there were <u>only</u> 10 inches of rainfall. Use the information provided to comment on this reported statement.

GO ON TO THE NEXT PAGE.

2. A department supervisor is considering purchasing one of two comparable photocopy machines, *A* or *B*. Machine *A* costs $10,000 and machine *B* costs $10,500. This department replaces photocopy machines every three years. The repair contract for machine *A* costs $50 per month and covers an unlimited number of repairs. The repair contract for machine *B* costs $200 per repair. Based on past performance, the distribution of the number of repairs needed over any one-year period for machine *B* is shown below.

Number of Repairs	0	1	2	3
Probability	0.50	0.25	0.15	0.10

You are asked to give a recommendation based on overall cost as to which machine, *A* or *B*, along with its repair contract, should be purchased. What would your recommendation be? Give a statistical justification to support your recommendation.

3. Every Monday a local radio station gives coupons away to 50 people who correctly answer a question about a news fact from the previous day's newspaper. The coupons given away are numbered from 1 to 50, with the first person receiving coupon 1, the second person receiving coupon 2, and so on, until all 50 coupons are given away. On the following Saturday, the radio station randomly draws numbers from 1 to 50 and awards cash prizes to the holders of the coupons with these numbers. Numbers continue to be drawn without replacement until the total amount awarded first equals or exceeds $300. If selected, coupons 1 through 5 each have a cash value of $200, coupons 6 through 20 each have a cash value of $100, and coupons 21 through 50 each have a cash value of $50.

   (a) Explain how you would conduct a simulation using the random number table provided below to estimate the distribution of the number of prize winners each week.

   (b) Perform your simulation 3 times. (That is, run 3 trials of your simulation.) Start at the leftmost digit in the first row of the table and move across. Make your procedure clear so that someone can follow what you did. You must do this by marking directly on or above the table. Report the number of winners in each of your 3 trials.

   72749 13347 65030 26128 49067 02904 49953 74674 94617 13317

   81638 36566 42709 33717 59943 12027 46547 61303 46699 76423

   38449 46438 91579 01907 72146 05764 22400 94490 49833 09258

Question 1 - Solution

Part (a):

An outlier is any value that is more than 1.5 *IQR below the lower quartile or 1.5 *IQR above the upper quartile. If MIN $< Q_1 - 1.5*$ IQR, there is at least one outlier on the low side and if MAX $> Q_3 + 1.5*$IQR, there is at least one outlier on the high side.

OR

An outlier is any observation that is more than 2 (or 3) standard deviations above or below the mean.

Part (b):

$$IQR = 19.250 - 9.680 = 9.57$$
$$1.5*IQR = 14.355$$

$$Q_1 - 1.5*IQR = 9.680 - 14.355 = -4.675$$
$$Q_3 + 1.5*IQR = 19.250 + 14.355 = 33.605$$

There is at least one outlier on the high side because the maximum value is greater than 1.5*IQR above the upper quartile, Q_3.

OR

mean $- 2*$std.dev $= 1.447$ OR   mean $- 3*$std.dev $= -5.300$
mean $+ 2*$std.dev $= 28.435$ mean $+ 3*$std.dev $= 35.182$
Since $38.180 > 28.435$ (or $38.180 > 35.182$), there is at least one outlier on the high side.

Part (c):

Since $Q_1 = 9.68$ inches, more than 25% of the years had less than 10 inches of rain. Hence, 10 inches of rain is not an unusual value.

OR

Since 10 inches is within one standard deviation of the mean ($Z = -0.732$), 10 inches of rain is not an unusual value.

Scoring

The solution of this problem has four components:

1. **Outlier decision rule:**
   The student must state how to make a decision, using correct boundary values (one of the two standard procedures), that identifies outliers on both sides.

2. **Two boundary values:**
   The student must describe how to get both upper and lower plausible boundary values using his or her decision rule - symbolically, numerically, or graphically.

Note: Incorrect but plausible boundaries:
- Give boundary values where the lower value is less than Q_1 and the upper value is greater than Q_3.

  AND
- Must be of the form:
  Outlier < (location measure) – (multiplier) * (spread measure)
  Outlier > (location measure) + (multiplier) * (spread measure)

3. **Execution of outlier decision rule:**
   The student must give a correct conclusion using his or her outlier rule. Only the upper side must be checked. Boundary values and outlier decision rule <u>must</u> be plausible to receive credit for this part. (Remember, the execution must be consistent with the rule given in part (a).)

4. **Comment on "<u>only</u> 10 inches of rainfall":**
   The student must state that 10 inches of rain is not (or is) an unusual value and must explain why.
   Examples of incorrect comments include stating that 10 inches of rain is not an outlier or basing his or her conclusion on an assumption of normality.

Parts (a) and (b) need to be read together. Credit for components 1 and 2 may be given if found in part (b) instead of part (a). If components 1 or 2 are incorrect in part (a), credit for components 1 and 2 cannot be given based on work in part (b).

   4   **Complete Response**
           All four components are essentially correct.

   3   **Substantial Response**
           Three components are essentially correct.

   2   **Developing Response**
           Two components are essentially correct.

   1   **Minimal Response**
           Only one component is essentially correct.

Remember: Before assigning a score, assess the entire paper holistically.