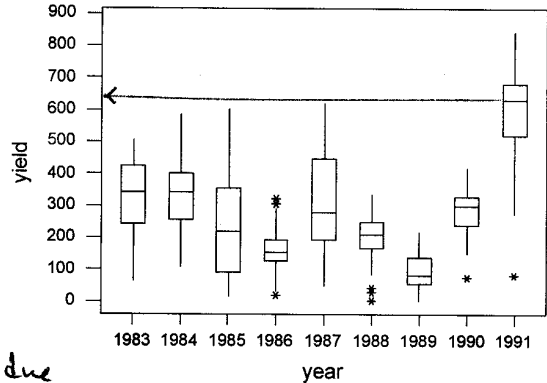


STA261 Exam I (Fall 2001)

Name Key Row: _____

Instructions: You must show all work to receive full credit, except in cases where calculator stat functions are used. Be sure your answers are in the context of the problems. Completeness and conciseness will benefit you. Point values in <>.

1. A small vineyard on South Bass Island near Lake Erie has kept records of their grape production over the years since the owners restarted the winery. The vineyard has 52 individual rows of vines of equal length, and the yield of grapes (in pounds) has been recorded for each row for each year ($n = 52$ per year). At right is a boxplot display of the yields for each of the years 1983 through 1991.



- a) Visually estimate the median yield per row for 1991. Would you say that the mean is a better measure of center than the median for the 1991 data? State why or why not. <6>

Median \approx 640 lbs.

The median is probably better due to the one vastly under-producing row in the vineyard. This outlier would lower \bar{y} .

- b) Which year would you say produced the most consistent yield across rows of the vineyard? The least consistent? List at least two real factors that might explain this consistency (or inconsistency) within the rows of the vineyard within a given year, and briefly discuss them. <2+2+4+4>

Most consistent: 1989

Least consistent: 1985, 1987 or 1991

Reason 1: Wind damage / Pest damage: probably happens most often at the edge of the vineyard, resulting in inconsistent yields across all rows.

Reason 2: Terrain: if some rows are in low spots, moisture will be higher there, producing higher yields

- c) Why do you tend to see outliers on the low side, but not on the high side? Offer possible reasonable explanations. <6>

It is much more reasonable to suspect that there are many factors which might inhibit yield in a specific row (pests, wind, etc); not so reasonable to suspect that one (or few) rows would produce an abundance

2. Rose, Jack and Cal are three friends who were enrolled in three different sections of STA261 at Miami. At the end of the term both Rose and Jack got B's (3.0 on a 4.0-point scale) and Cal got a B- (2.7 on a 4.0-point scale). A summary of class performance is given below. Based on this info, which of the three friends had the best performance in STA261 relative to their section? Offer justification. <8>

	class GPA	class SD
Rose's class	2.689	0.787
Jack's class	2.590	0.922
Cal's class	2.323	0.674

$$\text{Rose: } \frac{3.0 - 2.689}{.787} = 0.395$$

$$\text{Jack: } \frac{3.0 - 2.590}{.922} = 0.444$$

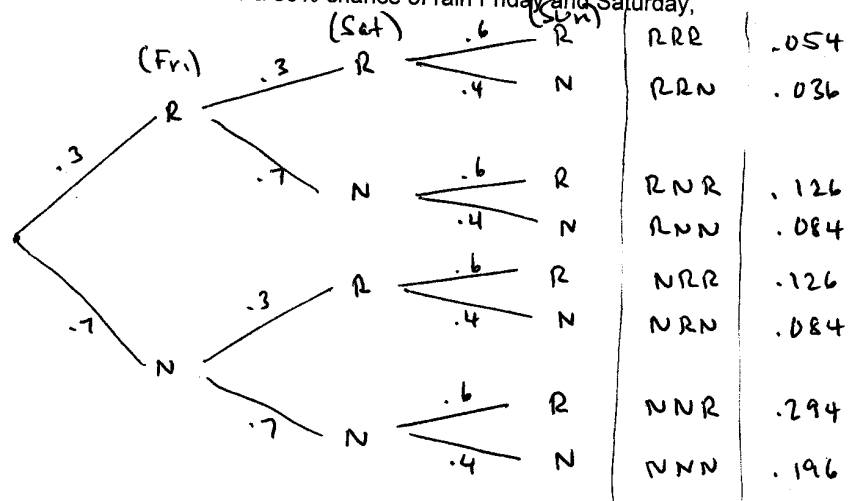
$$\text{Cal: } \frac{2.7 - 2.323}{.674} = 0.559$$

Cal did best: his standardized score is highest.

3. Suppose you have a camping trip scheduled this coming Friday, Saturday and Sunday at Hueston Woods State Park. You check the weather forecast and find that there is a 30% chance of rain Friday and Saturday, but a 60% chance on Sunday.

- a) At right, build a probability tree for the weather conditions for this upcoming weekend. <8>

R = rain
N = no rain



- a) What is the probability of a total washout this weekend? <5>

$$(.3)(.3)(.6) = \underline{.054}$$

- b) What is the probability that you have at least one day without rain? <6>

$$1 - .054 = \underline{.946} \quad (\text{note that this is the 'opposite' event of 'a washout'})$$

4. A national sample survey in July 1994 examined public opinion in the period before the 1994 elections. The survey interviewed 4,800 people age 18 and older nationwide by telephone. One question asked was whether they agreed with this statement: "Some people say we should have a third major political party in this country in addition to the Democrats and Republicans. Do you agree?"

- a) Define the apparent target population of interest. <4>

All adult Americans.

- b) For what population parameter might the survey want to make an inference? <4>

The true proportion of adult Americans who believe we should have a third major political party.

- c) Suppose the sample proportion responding 'yes' to the question was $2510/4800 = 0.523$. What is the generic name statisticians give to any quantity computed from a sample? <4>

a statistic

- d) Does the value 0.523 given in part c) necessarily equal the value of the population parameter in part b)? Briefly state why or why not? <4>

no. Sampling error via sample selection; biases in sample collection process.

- e) Think of two realistic ways in which **sampling bias** might creep in to a survey such as the one conducted above, and briefly discuss them. <3+3>

1. People without telephones would be excluded. This might bias against poorer Americans.
2. People might not be home when called. This might bias against people with long working hours or more active lifestyles.

Instructions: Clearly write the letter corresponding to the most appropriate response in the space provided. Each problem is worth 3 points.

The stock market has done well during most of the past decade. Below are the percent total returns (change in price plus dividends paid) for the Standard & Poor's (S&P) stock index. [Questions 5 through 10 refer to these data.]

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Return	31.7	-3.1	30.5	7.6	10.1	1.3	37.6	23.0	33.4	28.6

- D 5. The median return during the period 1989-1998 was:
- a] 5.7 b] 20.07 c] 23.0 d] 25.8 e] 28.6
- D 6. The range of these returns is:
- a] 3.1 b] -3.1 c] 34.5 d] 40.7 e] 33.4
- A 7. The mean return is:
- a] 20.07 b] 20.69 c] 22.3 d] 25.8 e] 10.1
- C 8. The standard deviation of the returns is:
- a] 13.75 b] 13.98 c] 14.74 d] 20.07 e] 25.8
- D 9. To give the five-number summary of these returns, you would report:
- a] the mean, median, third quartile, standard deviation and coefficient of variation.
 b] the mean, standard deviation, coefficient of variation, and the minimum and maximum.
 c] the median, mean, standard deviation, and the largest and smallest individual returns.
 d] the median, the two quartiles, and the largest and smallest individual returns.
- A 10. Stock returns are measured in percent. What are the units of measurement of the mean, the median, the quartiles and the standard deviation?
- a] All are also measured in percent.
 b] All are measured in percent except the standard deviation, which is measured in squared percent.
 c] All are measured in percent except the standard deviation, which is a number that has no units.
 d] All are measured in percent except the quartiles, which are one-fourth the size they should be.
- B 11. The frequency interpretation of probability basically states that $P\{E\}$ may be viewed as the proportion of times that event E would occur:
- a] in a random sample of size n .
 b] in an indefinitely large number of repetitions of a chance operation.
 c] in any number of repetitions of a chance operation.
 d] in a probability tree.
- A 12. The main advantage of boxplots over histograms is that:
- a] boxplots make it easy to compare several distributions.
 b] boxplots show more detail about the shape of the distribution.
 c] boxplots use the five-number summary, whereas histograms use the mean and standard deviation.
 d] boxplots show skewed distributions, whereas histograms only show symmetric distributions.
- B 13. The Empirical Rule is a guideline that talks about the proportion of observations that would fall within:
- a] a certain number of SDs from the mean for any distribution.
 b] a certain number of SDs from the mean for any roughly symmetric unimodal distribution.
 c] the "box" in a regular boxplot.
 d] the "box" in a modified boxplot.