

Statistics 571: Statistical Methods
Fall 2000 – Midterm Exam
Ramón V. León

Name: _____

This exam is closed book and closed notes. However, you can use a few pages of personal notes to help you with this exam. You can also use the statistical tables in the back of the book.

1. An Internal Review Service (IRS) examiner has a file of 30 returns to examine. Unknown to him, 5 of the returns have errors. Suppose he randomly selects returns to examine from the file
 - a. What is the probability that the first return does not have an error, but the second return does?

 - b. What is the probability that the second return has an error?

 - c. What is the probability that each of the first three returns has an error?

2. A survey is made of sources of news for people in a metropolitan area. It is of interest whether people obtain their news from television (Event T), a newspaper (Event N), or radio (Event R). Suppose that 76% of the people obtain news from television, 62% from newspapers, 46% from radio, 44% from television and newspapers, 28% from television and radio, 20% from newspapers and radio, and 5% from television, newspapers, and radio. (Hint: use a Venn diagram)
 - a. Find the proportion who obtain news from television, but not radio

 - b. Find the proportion who do not obtain news from either radio or newspapers

 - c. Find the proportions who do not obtain news from any of these three sources.

3. In a survey, 1000 adults were asked whether they favored an increase in the state income tax if the additional revenues went to education. In addition, it was noted whether the person lived in a city, suburb, or rural part of the state. These are the counts:

	Increase Income Tax		Total
	Yes	No	
City	100	300	400
Suburb	250	150	400
Country	50	150	200
	400	600	1000

- What is the probability that a person living in the city favors a tax increase?
 - What is the probability that a person living the suburb favors a tax increase?
 - What is the probability that a person living in the country favors a tax increase?
 - Does the result indicate that the place of residence and the opinion about tax increase are independent? Why?
4. An experiment measures the number of particle emissions for a radioactive substance. The number of emissions has a Poisson distribution with rate $\lambda = 0.20$ per week.
- What is the probability of zero emissions in a week?
 - What is the probability of at least one emission in a week?
 - What is the probability of zero emissions in a year?

6. Confounding is present in each of the following situations. Explain the nature of the confounding and the why the conclusions drawn may not be valid.
 - a. A cross-country coach thinks that a particular technique will improve the times of his runners. As an experiment he offers an extra daily practice to work on the technique for runners who want to participate. At the end of the season the coach concludes that the technique is effective, because runners who participated in the extra practices have faster average times than those who did not.

 - b. In a geriatric study samples of people over 65 from a general community and from a retirement community are followed. At the end of two years it is found that a larger proportion of the people continue to reside in the retirement community than in the general community. It is concluded that the retirement community living allows the elderly to maintain their independence longer than does the general community living.

7. A study monitored women from 1980 to 1990 and found that those who were modest coffee drinkers in 1980 (up to six cups of coffee daily) were not at an increase risk of heart disease over non-coffee drinkers after adjustment for smoking and other known risk factors. In contrast, some observational studies which studied the effects of coffee consumption on heart disease have not accounted for the fact that many heavy coffee drinkers are also smokers, a known risk factor for heart disease. Identify the response, explanatory, and confounding variables in these studies.

12. A survey collected the following information on new college graduates: sex, age at graduation, time to complete degree, future plans (graduate school, employment, indefinite), citizenship, class rank, and grade point average.

a. Classify each variable as categorical (nominal or ordinal) or numerical (discrete or continuous).

b. Classify the numerical variables as ratio scale or interval scale.

13. Daily rainfall in millimeters was recorded over 47 year period in Turrumurra, Sydney, Australia. For each year, the day with the greatest rainfall was identified. The most rainfall in a day in each year is shown below.

1468, 909, 841, 475, 846, 452, 3830, 1397, 556, 978, 1715, 747, 909, 2002, 1331, 1227, 2543, 2649, 1781, 1717, 2718, 584, 1859, 1138, 2675, 1872, 1359, 1544, 1372, 1334, 955, 1849, 719, 1737, 1389, 681, 1565, 701, 994, 1188, 962, 1564, 1800, 580, 1106, 880, 850

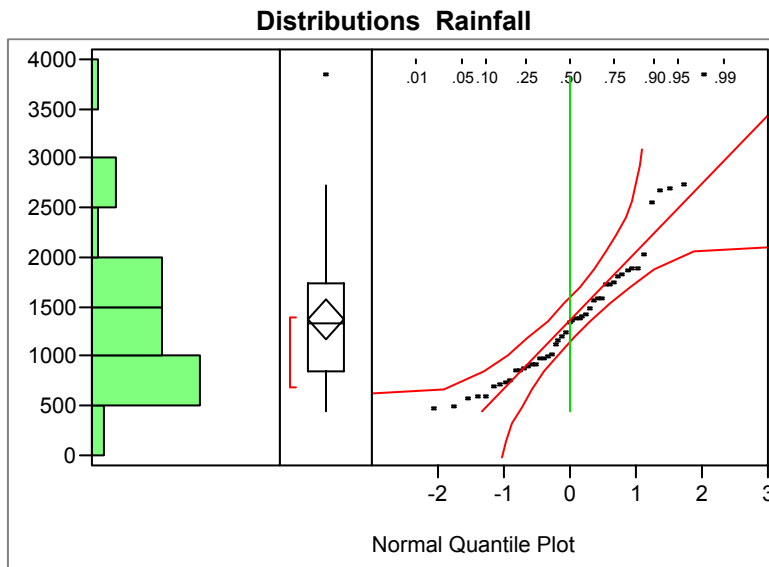
With the help of the JMP output next page:

a. Provide the five-number summary of these data. Does this summary suggest a symmetric or a skewed distribution?

b. Calculate a rough estimate of SD from IQR and compare it with the sample SD.

c. The five smallest observations are 452, 475, 556, 580, 584. The five largest observation are 2543, 2649, 2675, 2718, 3830. Use this information and the value of the sample mean in the JMP output to calculate the 10% trimmed mean. Compare it with the sample mean? Does this comparison suggest any outliers

- d. According to the fence rule are there any outliers in the data? Are there any extreme outliers in the data? Justify your answer.
- e. Comment on the shape of the normal plot. What kind of departure from normality do you see? Notice that this normal plot is sideways when compared to the ones given in your book.



Quantiles			Moments		
100.0%	maximum	3830.0	Mean		1369.106
99.5%		3830.0	Std Dev		693.670
97.5%		3607.6	Std Err Mean		101.182
90.0%		2564.2	upper 95% Mean		1572.775
75.0%	quartile	1737.0	lower 95% Mean		1165.438
50.0%	median	1331.0	N		47.000
25.0%	quartile	850.0			
10.0%		583.2			
2.5%		456.6			
0.5%		452.0			
0.0%	minimum	452.0			

14. A park district surveyed teenagers in a large metropolitan area regarding participation in a preteen soccer program and subsequent teenage drug use. The results are classified by family income as shown in the following table.

Family Income	Played Soccer	Drug Use	
		Yes	No
Low	Yes	10	40
	No	30	120
Medium	Yes	10	90
	No	10	90
High	Yes	6	144
	No	2	48

- a. Combine the income levels to construct a 2×2 table of drug use (columns) by soccer participation (rows). Compare the proportions of drug users among the teenagers who did and did not participate in soccer. Why would it be misleading to conclude from this table that involvement in soccer lowers the rate of teenage drug use?

Soccer Participation	Drug Use	
	Yes	No
Yes		
No		

- b. For each income level compare the proportion of drug users among teenagers who did and did not play soccer. What do these proportions indicate? Calculate the adjusted (for income) proportion of drug user for soccer players and nonplayers.

15. A computer network manager wants to model how access time in milliseconds (y) for data files varies with the number of simultaneous user (x) accessing the files. Based on 50 paired observations, the following summary statistics are obtained:

$$\bar{x} = 8.7, s_x = 2.5, \bar{y} = 15.3, s_y = 3.8, r = 0.8$$

Give the equation of the least-squares straight line to estimate the access time as a function of the number of simultaneous users. What is the estimated access time if there are 10 simultaneous users?