

Statistics Practice Final

Fall 2006

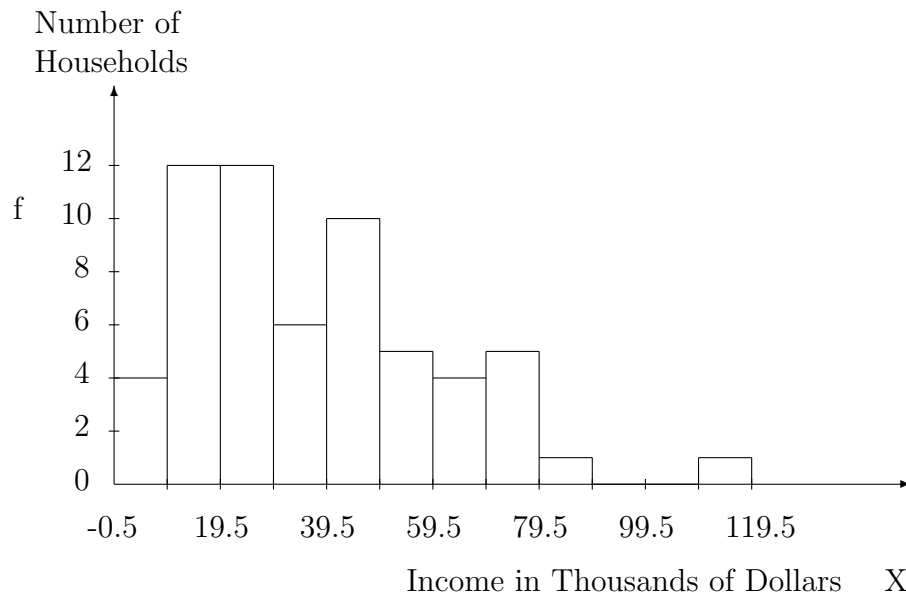
Name: _____

1. Pictured below is a histogram¹ of incomes for sixty Saskatchewan households in 1994.

a) Describe the shape of the data. Are there any outliers?

b) Is the mean bigger than, less than, or equal to the median?

c) What is a better measure of center for this data: mean or median?



¹ www.urgegina.ca/~gingrich

2. Here is a list of US Presidents, ordered by their age at inauguration. **Give** the 5-number summary of this data and draw a **boxplot** to represent this data.

42	51	55	60
43	51	55	61
46	51	55	61
46	52	56	61
47	52	56	61
48	54	56	64
49	54	57	64
49	54	57	65
50	54	57	68
51	54	57	69
51	55	58	

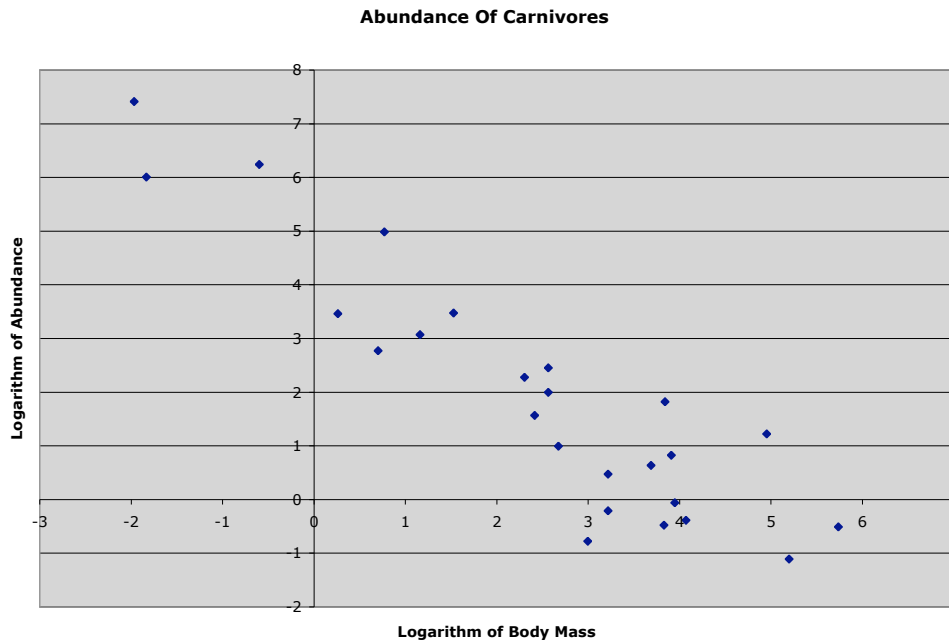
3. The distribution of heights of women aged 20 to 29 is approximately Normal with mean 64 inches and standard deviation 2.7 inches.

a) Between what heights do 95% of young women fall?

b) What percent of young women are taller than 72 inches?

c) What percent of young women have heights between 63 and 65 inches?

4. Here is a scatterplot relating the body mass of carnivores to their abundance per 10,000 kg of



a) Describe the **form, direction, and strength** of the relationship.

b) The correlation is $r = -.9136$. What does this tell you about the data?

c) The correlation squared is $r^2 = .83$. What does this tell you about the data?

d) The mean for the logarithm of the body mass is 2.45. The mean for the logarithm of Abundance is 1.93. The standard deviation for the logarithm of body mass is 2.03 and for the logarithm of abundance it is 2.33. What is the equation for the least squared regression line?

5. Listed below are 3 possible pairings of variables and 3 possible correlations. Match the correlation to the variables it probably describes. **Give a reason for each answer.**

- a. Calcium Intake and Strength of Bones
- b. Time spent partying and GPA
- c. Temperature outside and number of babies born

- i. $r = .75$
- ii. $r = -0.1$
- iii. $r = -.66$

6. Psychologists compared the age at which a child says his or her first word to their performance on the Gesell Adaptive Test, taken much later. The explanatory variable is the age at which the child says his or her first word. The response variable is the test score. The researchers found that the least squares regression line:

$$\hat{y} = 109.8738 - 1.1270x$$

describes the data. If a child says their first word at age 25 months, what do you predict their test score will be? Show all of your work.

7. What is Simpson's Paradox?

8. Below are several scenarios asserting causation between two variables. In each case list possible lurking variables whose effects may be confounded with those of the explanatory variable.

a) A study finds that high school students who take the SAT, enroll in an SAT coaching course, and then take the SAT a second time raise their SAT mathematics scores from a mean of 521 to a mean of 561. Therefore, taking an SAT coaching course will improve your score.

b) High schools that have well-funded music and theatre programs have higher graduation rates. Therefore, to improve the graduation rate at a high school we should provide better funding for the music and theatre programs.

c) Students at almost all colleges and universities get higher grades than was the case 20 years ago. Therefore, college professors now have lower standards than they did 20 years ago.

9. Each of the following studies or inferences has at least one serious error. What is the error in each situation?

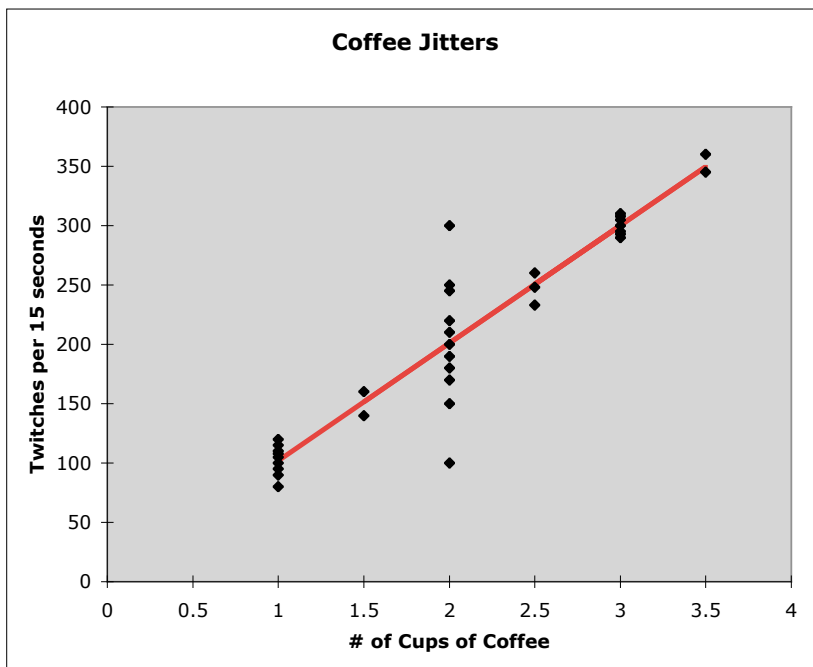
a) Pedestrians on Sunset Boulevard are stopped and asked their opinions about the current TV shows. The answers are used to make decisions about next seasons programming.

b) A senator receives an enormous number of petitions asking him to vote “No” on a bill requiring all street signs to be in both English and Spanish. Do to the overwhelming response the senator votes “No”.

c) A polling company calls random telephone numbers at 2 PM on a Saturday and asks people what kind of car they drive.

d) A scientist is measuring how plants respond to different types of soil. Each week she holds each plant against a ruler and measures the height.

e) Scientists give cups of coffee to randomly selected participants. They obtain the following scatterplot with least squares regression line shown. They use the regression line to predict how many twitches per 15 seconds someone who drinks 2.75 cups of coffee will have. They use the methods discussed in class to write down a 95% confidence interval for their prediction.



10. You want to design a study to find out if students who eat a full breakfast every morning get higher grades than those who eat no breakfast or a minimal breakfast.

a) Describe an observational study that will attempt to answer this question. Keep in mind that the college will not give you access to students' grades.

b) Describe an experiment that will attempt to answer this question. You should thoroughly describe what the control and treatment groups are and how you are assigning subjects to each group. You should also specify your explanatory and response variables and how you will measure them.

11. In a jug there are 5 blue marbles, 3 white marbles, and 1 red marble. You will randomly choose marbles from the jug. Be sure you show all of your work.

a) What is the probability of drawing a blue marble or a red marble?

b) Write down the sample space for the result of drawing two marbles from the jug.

12. Four randomly chosen rats from the NYC subway system are weighed. If the weight of all rats in the NYC subway system follow a normal distribution with mean .5 lb and standard deviation .02 lb what is the probability that the mean weight of the four randomly chosen rats is bigger than .8 lb ?

13. Below are several scenarios where a researcher will be estimating a parameter using a statistic. The options for types of inference are:

- a) Confidence Interval for a single mean
- b) Inference Test for a single mean
- c) Confidence Interval for the difference of two means
- d) Inference Test for the difference of two means
- e) Confidence Interval for a single proportion
- f) Inference Test for a single proportion
- g) Confidence Interval for the difference of two proportions
- h) Inference Test for the difference of two proportions
- i) χ^2 test for more than 2 proportions or for categorical variables
- j) Confidence Interval for the slope of a regression line
- k) Inference Test for the slope of a regression line
- l) ANOVA

For each of the following scenarios, write next to it the letter of the inference method which is **best** applicable. **If you choose a method which is an Inference Test you must say whether you want a 2 sided or 1 sided alternative hypothesis or if that decision is not applicable.** You may not need all of the methods listed above and some methods may be used more than once.

- i) 100 men and 100 women were surveyed. 65% of the surveyed men thought that *Lord of the Flies* described their life while only 63% of women thought that their lives were described by *Lord of the Flies*. Do the same proportion of men as of women think that their lives are described by *Lord of the Flies*?
- ii) 25 UCSB students were fed popcorn every morning for breakfast for a week and then made to run a mile at the end of the week. One month later the same students were fed hotdogs and baked beans every morning for breakfast for a week and then at the end of the week forced to run a mile. Which breakfast enables students to run the mile faster?
- iii) Randomly selected men at a certain company are categorized according to marital status (married, widowed, divorced, never married) and their incomes are classified as high, medium, or low. Do the same proportion of men in each marital category have high incomes?

iv) 35 plants in a laboratory have their heights measured 1 day before and 1 day after being injected with coffee. What are the smallest and largest changes that we would expect to see in a plant injected with coffee?

v) 250 randomly selected men have their head hairs counted and their intelligence measured. We want to know if having more hair is associated with less intelligence (in men).

vi) The number of curse words uttered by randomly selected children in urban, suburban, and rural settings is counted. Is there a difference between the average number of curse words uttered by children in these settings?

vii) 75% of surveyed Americans said they would rather rot their brains in front of a television rather than hear a beautiful French Horn recital. What are the possible percentages of Americans who would rather rot their brains?

14. The nitrogen level of air trapped inside amber from the late Cretaceous era was measured. A sample of 9 measurements was taken. The mean percent nitrogen for the sample was 60%.

What is a 98% confidence interval for the level of nitrogen in the air in the late Cretaceous era?

15. A random digit dialing telephone survey of 880 drivers asked “Recalling the last ten traffic lights you drove through, how many of them were red when you entered the intersections?” Of the 880 respondents, 171 admitted that at least one light had been red. Give a 95% confidence interval for the proportion of all drivers who ran one or more of the last ten lights they met.

16. A study by the national Athletic Trainers Association surveyed 1679 high school freshmen and 1366 high school seniors in Illinois. Results showed that 34 of the freshmen and 24 of the seniors had used anabolic steroids. Is there a significant difference between the proportions of freshman and seniors who have used steroids? Set up and perform a significance test for the difference of two proportions to answer this question.

17. A study of the career plans of young women and men sent questionnaires to al 722 members of the senior class in the College of Business Admin. at the University of Illinois. One question asked which major within the business program the student had chosen. Here are the data from the students who responded.

	Female	Male	Total
Accounting	68	56	124
Administration	91	40	131
Economics	5	6	11
Finance	61	59	120
Total	225	161	386

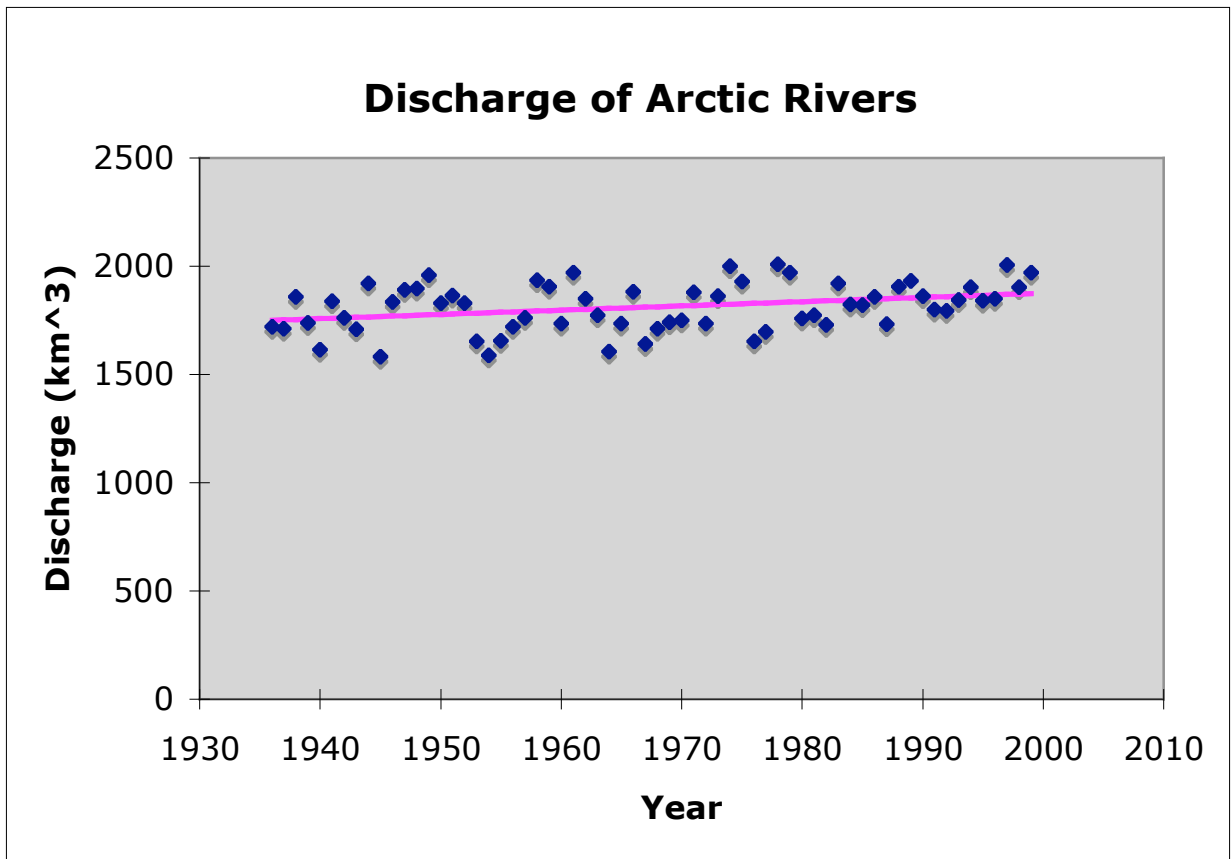
a) Here is a table of expected counts. Complete it

	Female	Male	Total
Accounting			124
Administration			131
Economics	6.4	4.6	11
Finance	70	50	120
Total	225	161	386

b) **State** the degrees of freedom for the χ^2 statistic.

c) The value of χ^2 for this data is 12.25. **Set up** and **Perform** an inference test to decide if there is a significant difference between the majors of the females and males.

18. If global warming is occurring, arctic rivers should be discharging more water now than they did 20 years ago. Below is a scatterplot of river discharge versus time with a least-squares regression line drawn in. Also below is the Excel output about the regression line.



SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.33439264							
R Square	0.11181843							
Adjusted R Sq	0.09749293							
Standard Error	104.002592							
Observations	64							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	84429.0055	84429.0055	7.80554702	0.00692068			
Residual	62	670625.432	10816.5392					
Total	63	755054.438						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-2056.7695	1384.68737	-1.4853674	0.14251371	-4824.72	711.181123	-4824.72	711.181123
X Variable 1	1.966163	0.70374909	2.79384091	0.00692068	0.55938863	3.37293738	0.55938863	3.37293738

a) Give the equation of the regression line.

b) Give a 95% confidence interval for the slope of the true regression line.

c) Set up and perform an inference test with $\alpha = .01$ to determine if the true regression line has slope equal to zero.

19. Researchers recruited 45 women who said they were dog lovers. Fifteen of the subjects were randomly assigned to each of three groups to do a stressful task alone (C group), with a good friend present (F group), or with their dog present (P group). The subject's mean heart rate during the task is one measure of the effect of stress. Here is the data and the Excel ANOVA output.

Is there a significant (at $\alpha = .05$) difference between the heart rates of the three groups? Explain your answer.

D	E	F
C group	F group	P group
84.738	99.692	69.169
87.231	91.354	68.862
84.877	83.4	70.169
80.369	100.877	64.169
91.754	102.154	58.692
87.446	89.815	79.662
87.785	80.277	69.231
73.277	98.2	75.985
84.523	101.062	86.446
77.8	76.908	97.538
70.877	97.046	85
90.015	88.015	69.538
99.046	81.6	70.077
75.477	86.985	72.262
62.646	92.492	65.446
	99.692	

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
C group	15	1237.861	82.5240667	85.4067026		
F group	16	1469.569	91.8480625	69.3114947		
P group	15	1102.246	73.4830667	99.3973152		
ANOVA						
Source of Variat	SS	df	MS	F	P-value	F crit
Between Grou	2613.08307	2	1306.54153	15.4900443	8.5823E-06	3.21448033
Within Groups	3626.92867	43	84.3471784			
Total	6240.01174	45				