## Statistics 31, Section 3, Midterm I Tuesday, September 26, 2000

Name:

Pledge: I have neither given nor received aid on this examination.

Signature: \_\_\_\_\_

Instructions: Do <u>not</u> do any actual numerical calculations (e.g. answers in a form that you would type into an Excel field, with a working answer, are expected).

1. Admissions policies at the Law School and Business School in a major university were compared for gender bias. Here is a breakdown of admissions during one period:

	Law	School	Business	School
	Admitted	Denied	Admitted	Denied
Male:	70	30	3	7
Female:	8	2	40	60

Note that in each school, the percent of females admitted (80% and 40%, respectively) is higher than the percentage of males (70% and 30% respectively). Yet, if the admissions across the schools are aggregated, the percent of females admitted (~44%) is paradoxically MUCH LOWER than the percent of males (~66%).

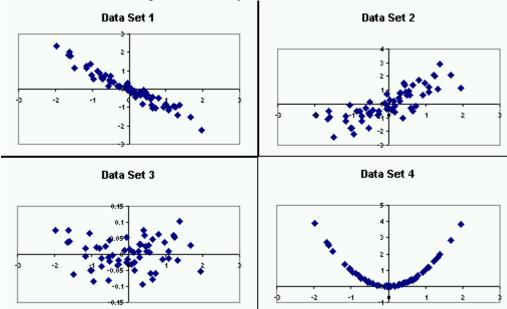
- a. Which of the following is the name of this phenomenon?
  - i. Confounding of Variables
  - ii. Simpson's Paradox
  - iii. The Law of Averages
  - iv. Extrapolation
- b. What is the lurking variable in the aggregated scores?
- c. Explain in 20 words or less why the aggregated percentages show women as being admitted less frequently.

2. Lengths of pregnancies vary approximately according to a Normal distribution with mean 266 days and standard deviation 16 days.

(a) Choose an Excel menu below (only one!), and fill it out to find the pregnancy length at the third quartile of the population.

NORMDIST					
×		🚺 = numbe	r		
Mean		🛐 = numbe	r		
Standard_dev		🛐 = numbe	r		
Cumulative		🛐 = logical			
Returns the normal cumulative distribution for the specified mean and standard deviation.					
Formul	a result =	ОК	Cancel		
NODMINU					
-NORMINV Probability		🚺 = numbe	r		
		🚺 = numbe			
Probability			r		
Probability Mean Standard_dev		s number = number =	r r		
Probability Mean Standard_dev Returns the inverse deviation. Probability		= number = number = e specified meaning	an and standard		

- (b) Write an Excel command to calculate the percent of pregnancies between 250 and 290 days
- (c) Write an Excel command to calculate the percent of pregnancies that are within 1.5 standard deviations of the mean.
- (d) Write an Excel command to calculate how long the longest 10% of pregancies last.
- (e) Use the 68-95-99/7 rule to write an Excel command to calculate values between which the middle 95% of all pregnancies lie.

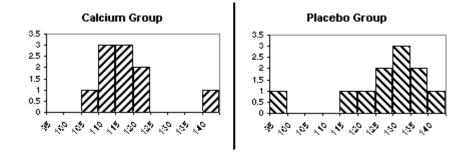


3. Here are scatterplots for 4 Toy Data Sets

Match the data sets to all statements which apply (matches can overlap, can be reused, and may not exist).

a.	Data Set 1 Data Set 2 Data Set 3 Data Set 4	The variables are strongly associated with each other. The variables are moderately associated with each other. The variables are not associated with each other.
b.	Data Set 1 Data Set 2 Data Set 3 Data Set 4	The correlation is approximately $r = -0.95$ The correlation is approximately $r = 0$ The correlation is approximately $r = 0.8$ The correlation is approximately $r = 0.95$
c.	Data Set 1 Data Set 2 Data Set 3 Data Set 4	There are no obvious outliers There is one serious outliers There are two probably outliers There are many outliers
d.	Data Set 1 Data Set 2 Data Set 3 Data Set 4	There is a positive linear relationship. There is a negative linear relationship. There is a curved relationship. There is no apparent relationship.

4. In a medical experiment, one group of men was given calcium, and another group was given a placebo. After some time their blood pressures were recorded and are shown in these two histograms.



Circle one of True, False or N.E.I. ("Not Enough Information") for each of the following:

b.	True	False	N.E.I	The Placebo Group population shows more variability.
C.	True	False	N.E.I	There is a mild positive correlation between Cal. and Pla. Groups

False N.E.I The Calcium Group had higher blood pressures overall.

d. True False N.E.I The Placebo Group has an outlier to the left.

True

a.

- e. True False N.E.I When the outlier is ignored, the Placebo Dist'n is left skewed.
- f. True False N.E.I When the outlier is ignored, the Cal. Dist'n is fairly symmetric.
- g. True False N.E.I The Calcium Distribution has 3 modes.
- h. True False N.E.I The median is larger for the Calcium Group than for the Controls.
- i. True False N.E.I The Inter Quartile Range is larger for the Calcium Group.
- j. True False N.E.I The range is larger for the Calcium Group.

- 5. To understand erosion, water was released on a test bed, at different flow rates, and the amount of eroded soil was measured.
  - a. What is the response variable?

b. What is the explanatory variable?

c. If the x-data values are in Excel cells D4:D24, and the y data values are in the Excel cells E4:E24, write an Excel formula to calculate the y-intercept of the least squares regression line.

d. For data as in (c), write an Excel formula to calculate the slope of the least squares regression line.

e. If the y-intercept and slope from (c) are -3 and 2 (respectively), write an Excel formula to calculate the predicted y value corresponding to a new x value of 27.

f. If the x values range from 5 to 15, is the prediction in (e) likely to be reasonably accurate? Explain why or why not in 20 words or less.