

NAME: \_\_\_\_\_

Math 1300

Exam #1

2/27/08

You may use calculators, R, web applets from the course webpage, and a table of z-values for this exam. 10 questions, 100 points.

- (10 pts) 1. The table below shows approximate sales of Vanilla Ice albums. Compute the mean and median of the sales data. Why is there such a large difference in the two values?

Year	Title	Sales (000's)
1989	Hooked	48
1990	To The Extreme	17000
1991	Extremely Live	500
1991	Cool As Ice	500
1994	Mind Blowin'	45
1998	Hard to Swallow	500
2001	Bi-Polar	20
2005	Platinum Underground	15
2011	W.T.F.	25

- (10 pts) 2. Suppose you spent a week at a gas station and recorded the model year of every car that stopped for gas. Explain why the distribution you find would be skewed to the left. Make a sketch of what this distribution might look like.

- (10 pts) 3. Suppose you are designing a study of SLU on-campus dining options by interviewing students.

- What are the individuals in the study?
- Give examples of quantitative and categorical variables you would want to measure in the study (give two of each).

- (10 pts) 4. The length of dog pregnancies from conception to birth varies according to a distribution that is approximately normal with mean 63 days and standard deviation 2 days.

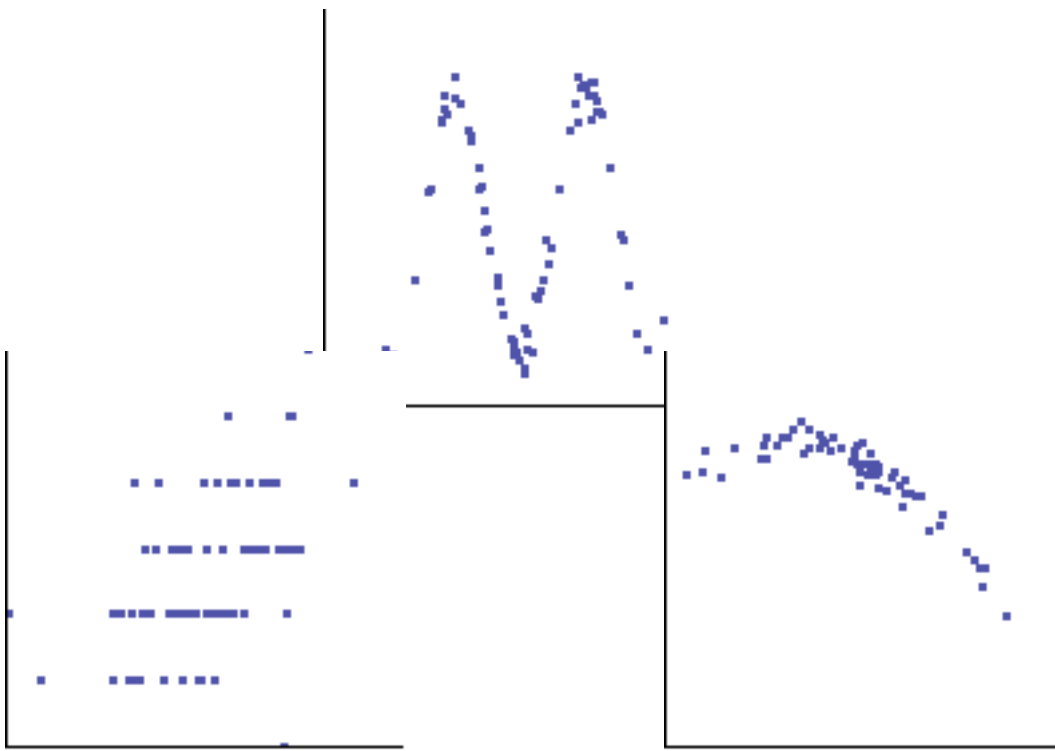
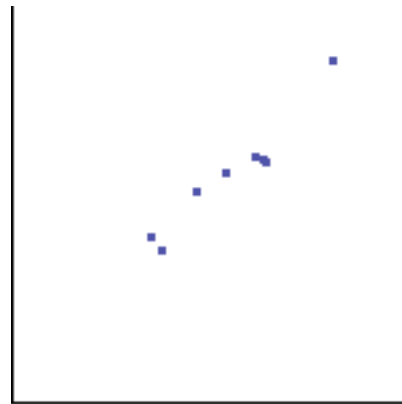
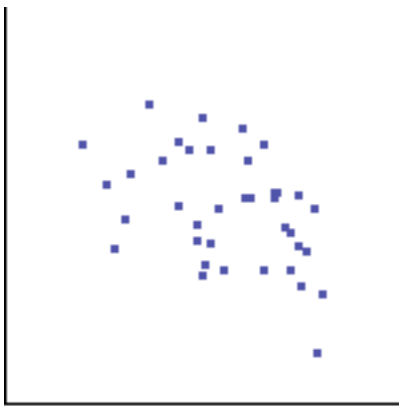
- What percentage of pregnancies last 60 days or less?
- What percentage of pregnancies last 66 days or more?

(10 pts) 5. Continuing the previous problem.

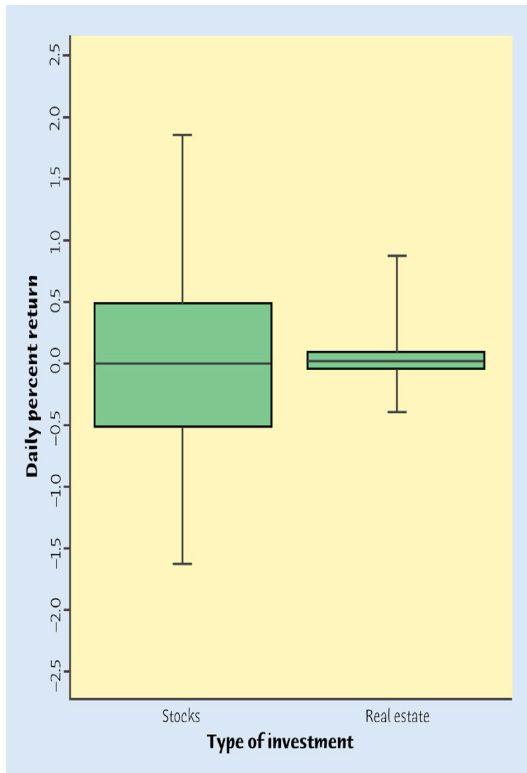
- a. What range covers the shortest 90% of dog pregnancy lengths?
- b. Accurately sketch the distribution of dog pregnancy lengths and shade the area corresponding to the 90% range of pregnancy lengths.

(10 pts) 6. Matching. Match the scatterplot with its description.

- a. Strong positive association.
- b. Strong non-linear association with  $r=0$
- c. Strong non-linear association with  $r < 0$
- d. Weak negative association.
- e. Weak positive association.



- (10 pts) 7. The figure below shows the distributions of daily percent returns for 14 months in 2004 and 2005 on two different types of mutual fund investment: Stocks (on the left) and Real Estate (on the right).



- About what were the highest and lowest daily returns on the stock fund?
- The median return was about the same on both investments. About what was the median return?
- What is the most important difference between the two distributions?

Questions 8,9, and 10 use the data frame `faithful` which is built in to R. It contains data on eruptions of the Old Faithful geyser taken in August, 1985. The variables are:  
`eruptions`: Duration (in minutes) of the eruption.  
`waiting`: Delay (in minutes) until the next eruption.

- (10 pts) 8. Make a histogram of `eruptions`.
- What are the mean and SD of this distribution?
  - Describe the overall shape of the distribution.

The Yellowstone Park Service uses the duration of each eruption to predict the delay until the next eruption, so that park visitors can plan to view the geyser.

(10 pts) 9. Make a scatterplot of eruptions ( $x$ ) against waiting ( $y$ ) and add the regression line to your plot.

a. Does there appear to be a positive or negative correlation ?

b. What is the correlation coefficient  $r$ ?

c. Do you think a line is a good choice to fit the data?

(10 pts) 10a. What is the equation of the regression line to estimate waiting from eruptions?

b. Use your equation to estimate the delay after an eruption of duration 2 minutes.