Math 263–001 Test 1

Instructor: David Love February 1, 2012

- Show **all** algebraic work to receive full credit.
- Please turn OFF all cell phones, pagers, and other communication devices and put them out of sight.
- All textbooks, notes, etc. must be put away. A 3×5 note card is allowed.

Student's Name (please print): Key

1. Give an example of a categorical variable.

Gender,

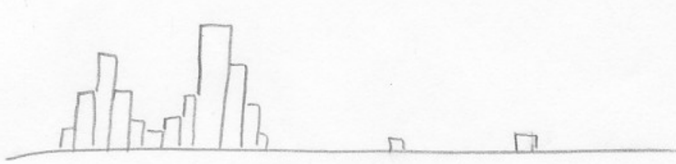
2. Give an example of the other kind of variable. What is the other kind called?

Quantitative – height, weight, jumping distance.

3. What is a "distribution?"

Something that describes what values a variable takes, and how often the values appear.

4. Sketch a histogram for a distribution that is bimodal, with two observations that are outliers.



5. A set of astronomical observations are normally distributed. Their mean is 4,392 and standard deviation is 381. These scores are converted to standard normal z scores. What would be the mean and median of the distribution of the converted scores?

mean & median of z-scores
are 0.

6. Find z in a standard normal distribution so that 18% of observations are above z.

$z = \text{invNorm}(1-0.18) = 0.9154$

7. A newspaper reports a high correlation between city and personal income, with a correlation coefficient of over \$1100. Identify and briefly explain one of several statistical blunders committed.

Correlation is only between quantitative variables, has no units. $-1 \le r \le 1$.

8. In a regression analysis, all the points lie on the line $\hat{y} = 1.5 - 2.1x$. What can you say about the correlation coefficient between x and y?

$r = -1$, because of negative slope

$r^2 = 1$

9. If $r^2 = 0.49$ for a particular Least Squares Regression Line of y on x, then x and y are positively correlated. T/F?

F. $r = 0.7$ or -0.7.

10. When doing linear regression on two variables, it does not matter which is the explanatory variable and which is the response variable. You will get the same correlation coefficient. T/F?

T. $r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$.

No difference between x & y variables

11. When doing a linear regression on two variables, an outlier will always have a large residual. T/F?

F. An outlier could be an influential point.

12. Suppose that the median of a set of income data is \$90,000 and the mean is \$25,000. What does this tell you about the distribution of incomes?

$M > \bar{x}$, so it is skewed left.

$\bar{x} < M$

13. The results of a large high jump competition are normally distributed with mean 4.2 feet and a standard deviation 0.9 feet. In each case you must explain how you got your answer, and determine the associated z-scores.

(i) What fraction of jumps are less than 3 feet?

$$z = \frac{3-4.2}{0.9} = -1.33$$

normalcdf(−E99, z) = $\underline{0.0912}$

(ii) What fraction of jumps are between 3 and 5 feet?

$$z_3 = -1.33$$

$$z_5 = \frac{5-4.2}{0.9} = 0.8889$$

normalcdf(z_3, z_5) = $\underline{0.7218}$

(iii) How high are the highest one-seventh of jumps?

$$z = \text{invNorm}\left(1-\frac{1}{7}\right) = 1.0676$$

$$z = \frac{x-\mu}{\sigma} \longrightarrow x = \mu + z\sigma = 4.2 + 1.0676 \cdot 0.9$$

$$= \underline{5.16 \text{ feet}}$$

14. In Professor X's statistics course, the correlation between the students' total scores before the final examination and their final examination scores was $r = 0.7$. The pre-exam totals for all students in the course has mean 410 and the standard deviation is 45. The final exam scores have mean 158 and standard deviation 12. X has lost a student's final, and the class records indicate that the student had earned 300 points before the exam. X decides to predict the final examination grade from the pre-exam total, and use that prediction to set the student's grade.

$\bar{x} = 410 \quad \bar{y} = 158$

$s_x = 45 \quad s_y = 12$

$r = 0.7$

(i) Find the LSRL from the information provided.

$\hat{y} = b_0 + b_1 x$

$b_1 = r \frac{s_y}{s_x} = \underline{0.187}$

$b_0 = \bar{y} - b_1 \bar{x} = \underline{81.47}$

(ii) Use the LSRL to predict the final examination score for the student whose exam was lost.

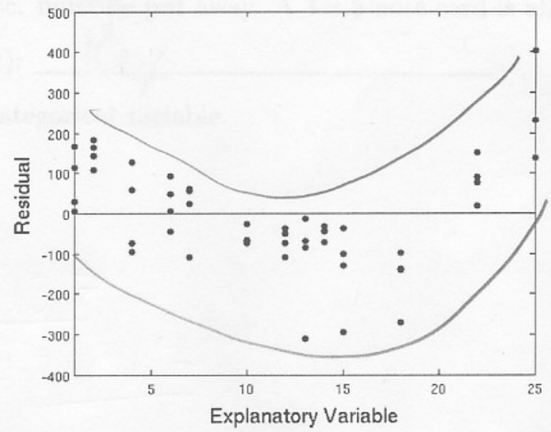
$\hat{y} = 81.47 + 0.187(300) = 137.57$

(iii) Critique this method of assigning the final grade, using the statistical information provided.

$r^2 = 0.7^2 = 0.49$

Less than half of the variation is accoanted for by the line. The real score could be quite different.

15. A Least Squares Regression Line was calculated for a set of observations, and the residual plot is given below. What does the residual plot tell you about using a least squares regression line on this data?



The residuals are curved, suggesting that a LSRL may not be the best tool.