Introduction to Statistics in Psychology: PSY 201

Greg Francis, PhD
Department of Psychological Sciences

EXAM 1

Name _____

Total points on the exam is 100. The exam will count as 10% of your class grade. Write your answers on the exam. Including all of your intermediate work will give you the best chance of getting partial credit (if necessary). Use the back of the page if necessary. The problem with your lowest score will be dropped from your grade.

(1) In a study of how infants respond to visual stimuli, researchers monitored the eyes of 16 6-month-old infants while each infant viewed pictures of spiders and flowers. Pupil dilation was measured because increased pupil dilation indicates arousal. The resulting mean pupil dilation for flowers was 0.03 millimeters, and the mean dilation for spiders was 0.14 millimeters. This study has been interpreted as evidence that 6-month-old infants respond with more arousal to spiders than to flowers.

a) Name a descriptive statistic in the study.

b) What part of the study uses inferential statistics?

c) How are the data, inferential statistics, and population related?

d) Suppose you decide to sum each infant's pupil dilations for both flowers and spiders. If you knew the variance of pupil dilations for both flowers and for spiders, what else would need to know to compute the variance of the sum of dilations? Write the equation you would use.

(2) In the StatLab Assignment on the Horizontal Vertical Illusion, you adjusted the size of a horizontal line until it appeared to match the size of a fixed target line. The target line was sometimes oriented horizontally and sometimes oriented vertically. For the class, the mean match size for horizontal targets was 99.3, and the mean match size for vertical targets was 105.6.

a) Identify the independent variable. What is its scale of measurement? Is it quantitative or qualitative?

b) Identify the dependent variable. What is its scale of measurement? Is it quantitative or qualitative?

c) Given the types of variables involved, would a bar chart or line graph provide a better way of representing the data? Justify your answer.

(3) Sketch an example of each of the following

a) A distribution with negative skew.

b) A distribution with high kurtosis.

c) A bimodal distribution.

d) An ogive, plotting cumulative frequency against upper limit of each class interval.

(4) The web site Pop vs Soda (popvssoda.com) collects data via an online survey. Participants are asked to report what U.S. county they are from and what generic word they use to describe carbonated soft drinks. The current results for Indiana are displayed in the following table.

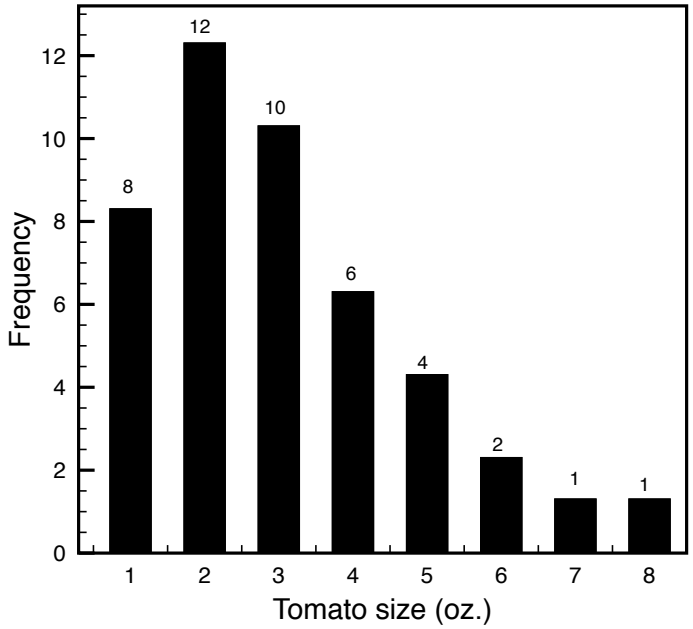
	Pop	Soda	Coke	Other
Number of responses	5591	1485	1989	363

a) Given this study, identify the population. Justify your answer using the definition of "population."

b) Given this study, identify the sample. Justify your answer using the definition of "sample."

c) Discuss some advantages and disadvantages of using an online survey to gather this kind of data.

(5) Scientists create a hybrid tomato by crossing a cherry tomato plant that produces 1 oz tomatoes with a beefsteak plant that produces 9 oz tomatoes. You make a frequency distribution of the resulting tomato sizes by plotting number of tomatoes against fruit size in ounces.



a) What are the mode, median, and mean of the data?

b) Name one advantage and one disadvantage of presenting the data in a graph compared to presenting frequencies in a table.

c) If this data was presented in a cumulative frequency histogram, what would be the cumulative frequency for the final category? Justify your answer.

(6) There are 9 scores in a dataset. If the data set is treated as a *sample*, then the variance is 25.

a) Treat the data set as a *population* and compute the variance.

b) Treat the data set as a *population* and compute the standard deviation.

c) Compute the sample standard deviation.

d) Explain why the equation for calculating the population variance differs from the equation for calculating sample variance.

(7) For the data in the following sample:
2, 3, 5, 6, 8, 9

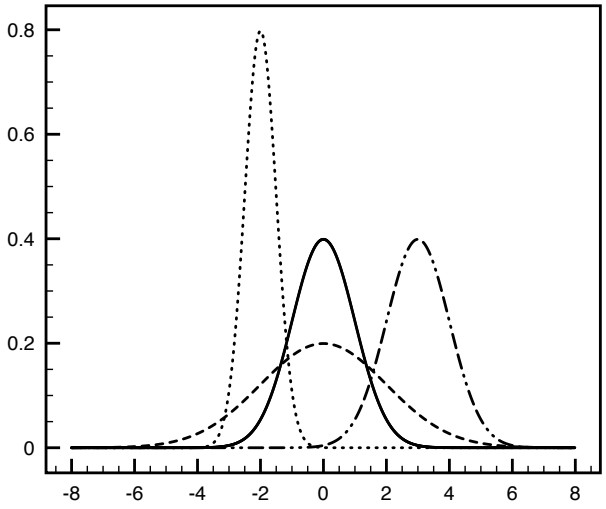
a) Compute the mean.

b) Compute the median.

c) Compute the standard deviation.

d) Compute the P_{35}.

(8) In the graph below, label the following distributions:



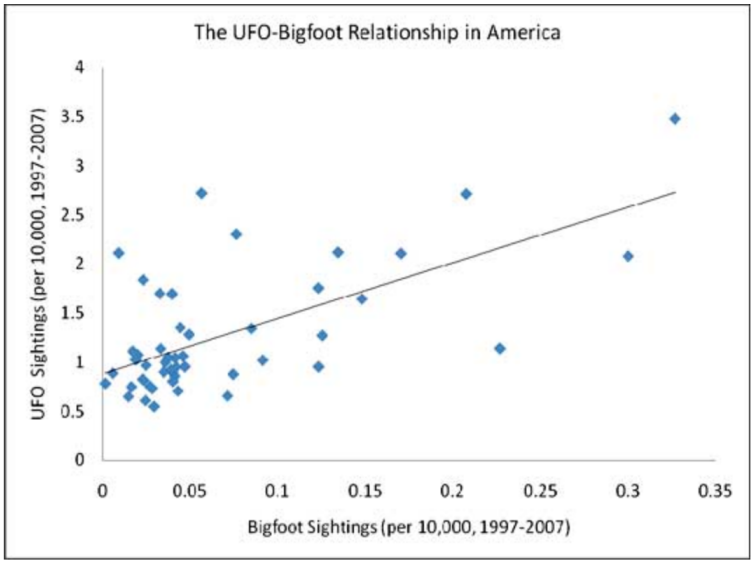
a) Label the standard normal distribution with the letter "A."

b) Label the distribution with a mean of 3 with the letter "B."

c) Label the distribution with a standard deviation of 2 with the letter "C."

d) Label the distribution having the largest area above the value 4 with the letter "D."

(9) At freakonomics.com, the "intriguing pattern" below was reported. The figure plots total UFO sightings in the U.S. for each state per 10,000 residents from 1997-2007 against total bigfoot sightings in each state per 10,000 residents for the same period.



a) Is the correlation between bigfoot sightings and UFO sightings positive or negative?

b) Can we conclude from this relationship that UFO sightings cause bigfoot sightings? Justify your answer.

c) Sketch a scatterplot of these variables if there were *no correlation* between them.

d) What is the Pearson correlation coefficient of the variables you sketched in part (c)?

(10) All of the following questions refer to a standard normal distribution.

a) What is the total area under the distribution?

b) Define the *Xth* percentile. How do you find it?

c) What percentile is the mean? Justify your answer.

d) The 80th percentile is 0.841. What is the 20th percentile?

(11) Five professional clowns were surveyed to see how many balls they could juggle and how many years they have been gainfully employed.

Clown	Balls (X)	Years employed (Y)	XY	X^2	Y^2
Bobo	6	25	150	36	625
Antsy	4	2	8	16	4
Toodles	5	6	30	25	36
Luna	10	9	90	100	81
Topsy	5	3	15	25	9
Sum	30	45	293	202	755

a) Compute the Pearson r.

b) Suppose there is another data set (call it "Data Set 2") with an r value that is equal to half the value of the r you computed in part (a). What is the r value of Data Set 2?

c) Is the relationship between the variables in Data Set 2 half as strong as the relationship between the variables in the above table? Justify your answer.

You may find the following equations useful.

$$P_X = X_{I_R} + (X_{I_R+1} - X_{IR})\, F_R$$

$$PR_X = \frac{cf + (0.5)f}{n} \times 100$$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$s^2 = \frac{\sum_i (X_i - \overline{X})^2}{n-1}$$

$$s^2 = \frac{\sum_i X_i^2 - [(\sum_i X_i)^2/n]}{n-1}$$

$$\sigma^2 = \frac{\sum_i X_i^2 - [(\sum_i X_i)^2/n]}{n}$$

$$\text{skew} = \frac{\sum_{i=1}^{N} (X_i - \mu)^3}{N\sigma^3}$$

$$\text{kurtosis} = \frac{\sum_{i=1}^{N} (X_i - \mu)^4}{N\sigma^4} - 3$$

$$Y = bX + A$$

$$x_i = X_i - \overline{X}$$

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

$$r = \frac{\sum_i X_i Y_i - \frac{1}{n} \sum_i X_i \sum_i Y_i}{\sqrt{\left[\sum_i X_i^2 - \frac{\left(\sum_i X_i\right)^2}{n}\right] \left[\sum_i Y_i^2 - \frac{\left(\sum_i Y_i\right)^2}{n}\right]}}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$z = \frac{X - \mu}{\sigma}$$

$$X = z\sigma + \mu$$

$$z = \frac{X - \overline{X}}{s}$$

$$X = z(s) + \overline{X}$$