Homework 02
Psychology 310

1. (25 points). Answer the following questions about the normal distribution:

   (a) (5) A student received a grade of 76 in a course where the class average was 70, and the standard deviation 10. If the class distribution was approximately normal in shape, what was the student's approximate percentile rank?

   (b) (5) If SAT scores have a mean of 500 and a standard deviation of 100, approximately what percentage of students obtain SAT scores between 550 and 650?

   (c) (5) What SAT score is approximately at the 99th percentile?

   (d) (10) *Statistical Detective.* Suppose that the class distribution in a large course is almost exactly normal in shape. Joe got an 88 and had a percentile rank of 79.8, while Felicia got a 75 and had a percentile rank of 40.1. From this information, estimate the mean and standard deviation of the class distribution. (*Hint.* The percentile ranks tell you the Z-scores, and they are a function of the mean and standard deviation.)

2. (15 points). For each of the following expressions, for variables X and Y and constants a,b,c,d tell (i) whether or not the expression is a linear combination of X and Y, and (ii) if the expression is a linear combination, what are the linear weights for X and for Y.

   (a) $(9X - 6Y)/3$

   (b) $(X + Y)^2$

   (c) $4(X + Y)$

   (d) $a^2 X + b^2 Y$

   (e) $2XY - 5X$

3. (15 points) The sample mean $\overline{X}_\bullet$ is a linear combination of the N observations, which are generally assumed to be independent and hence uncorrelated, and which are assumed to each have the same mean μ

and variance σ^2. Suppose a population has a normal distribution, a mean μ of 100 and a standard deviation σ of 15. If you take samples of size $N = 25$ from this population, and compute the sample means $\overline{X}_\bullet$, they will, over repeated samples, have a distribution known as the sampling distribution of the sample mean. What will be the mean and variance of this sampling distribution? *Hint.* Suppose N is 2 instead of 25, and apply the linear combination theory we covered in class. Then generalize to $N = 25$.

4. (10 points) Given the same information as the preceding problem, what is the probability that a sample mean $\overline{X}_\bullet$ based on a random sample of size 25 will be greater than 108?

5. (10 points) Given the following data, find the linear regression equation for predicting Y from X, and then compute the predicted value of Y when $X = 4$. You may use R and any of its functions to perform the calculations.

X	Y
5	7
4	11
8	14
7	11
2	5
3	6
9	16
11	22
10	18
6	13

6. (10 points) In the data set *TwoGroup.csv*, available online, test the following two hypotheses, using the approach shown in the *Multiple Regression* lecture notes:

   (a) The two sexes have equal regression line slopes
   (b) The two sexes have equal regression line intercepts

   Note that females are coded 1 and males 0 in this file.

7. (15 points) Using the data and analysis results from the previous problem, produce a graph similar to the one at the end of the *Multiple Regression* lecture notes, showing the data points color-coded by sex,

and with separate regression lines for males and females. The data are coded Male = 1, Female = 0. How could I have set up the data file so that this would be self-evident (i.e., "self documenting")? (*Hint* (for one of several possible answers): In R, 1 means `TRUE` and 0 means `FALSE`).