Psychology 105, Advanced Research Methods, Homework #3

Available 3/10/2021, due 3/18/2021

Once again, you will continue to work with the Statlab50 data set that you created for Homework #1.

**Part One: Linear regression**

1. In class, we looked at the regression of Peabody on Raven for Jack's Statlab50 sample. This began with the creation of a perfect scatterplot (good axes labels, good title, good aspect ratio, 5% extra white space provided on both axes). Produce a similar scatterplot for your own Statlab50 sample.
2. Estimate the linear regression of Peabody on Raven in your sample.
3. Add the regression line to your plot using the abline() function.
4. Interpret the slope: What does it say about the conditional mean of Raven as a function of Peabody?
5. Test the null hypothesis that the slope equals zero, using an alpha level of .05. What do you conclude about the relationship between Peabody and Raven scores?
6. List the assumptions necessary for that inference to be valid. (Powerpoint from March 9 will be informative for this question.)
7. Assess each assumption, or state why you lack sufficient information to assess it. Are you concerned about the validity of your inference?

**Part Two: Using simulation to learn about probability distributions**

We have encountered various probability distributions in class (e.g., Bernoulli, normal, uniform, chi square). The exponential distribution is one that we have not discussed. Exponential random variables are sometimes used to model time until an event occurs or time between events. The exponential distribution is governed by a single *rate* parameter, which must be greater than zero. So we might speak of an "exponential(2.5)" distribution, meaning an exponential distribution with rate 2.5.

In R, it is easy to sample from the exponential distribution. The syntax is: rexp(size of sample, rate parameter). For example, to create a variable called "x" that contains half a million draws from the exponential distribution with rate parameter 1.0, we would enter

```
x <- rexp(500000, 1.0)
```

We could then do things like plot a histogram of x to see the shape of the distribution, find the mean and variance of x, and so on.

**Your first task** in Part Two is to use simulations in R to determine the relation between the rate parameter and the mean and variance of the distribution. Do this by simulating several large samples (at least 500000) of exponential random variables with different rate parameters, calculating the mean and variance of each sample, and looking for the pattern. (Note: "var(x)" will give the sample variance of x.) It may be easiest to see the pattern if you try rate parameters 1, 2, 5, and 10.

Try to describe the pattern as a mathematical function. For example, if you noticed that a large exponential sample with rate 1 had a variance of 0.98, a sample with rate 2 had a variance of 4.03, and a sample with rate 5 had a variance of 24.87, you might hazard a guess that the variance of an exponential random variable is the squared value of its rate parameter. (Note: that's not actually true!)

**Your second task** is to describe the skew of the exponential distribution. Specifically, does the symmetry of the distribution appear to change when you change the rate parameter? If so, how? Use histograms and Pearson's index of skew to discuss symmetry. Recall that in R, you can calculate Pearson's index for a variable called "x" by typing

```
3*(mean(x)-median(x))/sd(x)
```

or by using the "pskew()" function that we wrote.