**Computing and Graphics in Applied Statistics February 26, 2020**

**Sample questions for Test 1** (I indicated 5 sample questions were to be given but these 4 nicely capture the essence of the likely questions to appear on the test).

**1.** Given the follow data: y1=1, y2=3, y3=4, y4=6, y5=7, y6=18

Compute the 1/6 or (16.7%) trimmed mean

**Sol1:** Give mean of the middle numbers because there are 6 numbers and you are asked for a 1/6th trimmed mean = (3+4+6+7)/4

**2.** Suppose a coin with Probability(Head) = 0.2 is tossed until a Head appears. Outline the steps of a Monte Carlo simulation that you would use to obtain the expected number of tosses of this coin until the first Head appears.

**Sol2:** Basically, just use the law of large numbers and say that we need to run the experiment large number of times

* Write a procedure that simulates tossing of a coin with prob(Head) = 0.2
* Record the number of trials until a head appears and you record that
* Repeat it a large number of times, say 5000 times
* Estimate = average => sigma(5000)/5000 = sum of the observations / number of observations

**3.** What is meant by an adaptive estimator?

**Sol3:** Basically, uses the information in the sample as part of the procedure. You take the sample and based on that (you do some trimming and other methods) you learn something. You let the CI and trimming to be dictated by the sample. Computing 95% CI by using a formula is not adaptive.

**4.** A researcher is interested in estimating the probability of cure represented by Y=1 after receiving a given dose(mg) X of a drug. Note that Y=0 represents the event of no cure.

Prob(Y=1)= (βx) / (1 + (βx) )

The following 3 data points: (X,Y)= (1,0), (2,1), (4,1) were observed. If there are only 2 possible choices for β, 0.5 or 1.0, find the Maximum Likelihood Estimator for β.

**Sol4:** compute the likelihood at 0.5 and 1.0

L0.5 = looking at (1,0) => P(Y = 0) = 1 – P(Y = 1) = ((1 – 0.5(1)) / (1 + 0.5(1))) = ans1

L1.0 = looking at (1,0) => repeat the above process and substitute x and beta = ans2

The max is the final ans

**Notes:**

1. random number between 0 and 1 that has equal density. Generate a random number between 0 and 1 keep going till u hit the graph and the x value at that is the random value u get with cdf, cumulative distribution function
2. hw4 is imp
3. cdf distributions always go from 0 to 1. E in ecdf means u define it from sample
4. trimming deletes the outliers and winsorizing replaces outliers. Winsorizing takes the value next to it and replaces the extreme max and min with it
5. residuals = data – fit
6. sigma = eta (0, greekSigma^2) 🡪 see what this is
7. error terms aren’t than do robust regression and then do bootstrapping to get CI
8. error terms are normally distributed – normal theory
9. p value should always be 0.05 which is 95% CI. If the var is less than p value, then and only then it could be useful. Otherwise not
10. R^2 goes up as u keep adding variables in your regression modal and residual(error) numbers go down. How much variations in y from the x is what explained by R^2 value?
11. As R^2 goes up, R^2(adj) does not always go up