**7/20/15**

2 Main Themes:

--Regression analysis

--Statistical software

--Diagnostic plots

\*This class will usually assume a normal distribution

\*Not proving things in this class

**Review Questions**

1. Give examples of continuous variables and categorical variables.

--Continuous: time, height, weight (anything that is a real number in a range)

--Categorical: food, gender, brands

--Ordinal: year in school (1 2 3 or 4)

2. Difference b/w population and sample

--Use a sample to infer characteristics about a population

--Sample is data you collect

3. What is the name of each of these Greek letters? What do they represent in statistics?

α – alpha = type 1 error aka “level” (how often your going to reject the hypothesis if it’s actually true)

β – beta = type 2 error (probability of accepting the null hypothesis when it’s actually false). Also represents regression parameters.

ε – epsilon = random error in regression equation (random measurement error in a measurement model)

θ – theta = arbitrary population parameter in a probability distribution (don’t have to worry b/c as

κ – kurtosis = thick tails (determine whether our regression is good or not)

μ –

ν –

π – pi = (important for logistic regression)

ρ – rho = population correlation

σ – sigma = standard deviation (how spread out your normal distribution is; fairly specific; find inflection points 🡪 critical point is where first derivative = 0; where curve flattens out, horizontal slope. Inflection point when second derivative = 0; where slope reaches maximum value and stops increasing. Sigma is the difference b/w critical point and inflection point.

φ – psi = represent normal density function

χ – chi = logistic regression

ω – omega = represents a point in the sample space

4. What is a random variable?

--Random variable is a process for obtaining random outcomes. (Flip a coin tell me how many Heads you get; the answer is a random variable)

9. Define these terms:

--Discrete Random Variable = Flip a coin and tell me how many heads you get? (Process for getting a random outcome and there is only a FINITE amount of outcomes)

--Continuous Probability Density = Cant calculate something exactly. Imagine a normal distribution for height, can’t ask the probability of a specific height, but CAN ask the probability of a range of heights.

--Uniform Distribution = equal probability of getting something within an equally sized range

--Parameter vs Statistic = Parameter 🡪 Population; Statistic 🡪 Sample

**Practice Problem**

To get a 95% confidence interval:

--Compute Z score for sample mean

-1.96 <= Z <= 1.96

Z = (x-bar – mu)/SEx

SEx = Sx / sqr(n)

**Working with R**

Dnorm() gives densitiy

Pnorm() gives probability