***COURSERA: STATS W/ R SPECIALIZATION***

***COURSE 2 - Inference***

**WEEK 4 - Inference for Proportions**

***4.4.1 Inference For Proportions***

**Introduction**

* Inference for categorical variables where **parameter of interest** = a proportion, as opposed to mean
* Gallup poll results where American public was asked about opinion on same-sex marriages
* Poll where data were collected from a variety of countries trying to answer the if most children in the country have the opportunity to learn + grow every day
* Published study on antihypertensive meds + serious falls on nearly 5k Americans > 70 during a 3 year period + found those who taking antihypertensive meds had a 30-40% greater likelihood of experiencing severe fall-related injuries like hip fractures + head trauma
* What is common between these studies = they deal w/ categorical variables like on same-sex couple marriages, whether children in a country have the opportunity to learn + grow every day, + whether patients taking a certain type of med are more likely to have fall-related injuries.
* Ex1: Simple case = categorical variable only has 2 levels we can categorize as a success or failure (**binary**)
* Success != something positive 🡪 could be a patient dying or suffering from a certain type of disease, or somebody graduating from high school
* Doesn't matter the context = important thing is these = **binary categorical variables** = levels can be categorized as either 1 thing or the other
* In this case, parameter of interest = **proportion of success**.
* Ex2: categorical variable w/ > 2 levels.
* Ex: Socioeconomic status tends to be categorized as low, medium, or high
* Ex3: 2 categorical variables that both have only 2 levels.
* IF somebody is male or female, if they decide to pursue a major in the sciences or not
* Ex4: 2 categorical variables, where either 1 or both of have > 2 levels
* Ex: Socioeconomic status (low, med, high) + educational attainment (finished high school, junior college, college, or graduate degree) + look at the relationship between these 2 variables
* In this case, we evaluate whether these variables appear to be dependent or independent.

**Sampling Variability and CLT for Proportions**