9:00 (Homework #1a)

If you would remember back to the beginning of class, we rotated tables to meet some new people and go over the homework (exercises #4). Sean gave a stunning explanation of number 1a discussing dummy variables and analysis of variance.

- **Analysis of Variance** – in this homework, we ran analysis of variance on our model looking at the sum of squares output to determine how much predictive influence each variable has on the model output

The important concepts from this homework (as shown on the board by Dr. Scott) are **analysis of variance, sampling distribution, Monte Carlo Simulation, standard error, bootstrapping, bootstrapped standard error, confidence interval, and frequent coverage property.**

9:45 (Homework #1b)

Next, Chisam and Jonathan explained two different ways to do 1b. Chisam solved the problem without using analysis of variance, sparking a discussion about main effect vs. interaction variables. Jonathan did use analysis of variance, and his explanation of the problem was much like Sean's explanation of 1a.

10:00 (Homework #2)

Dr. Scott explained number 2 homework:

Monte Carlo Simulation: A way of accounting for the uncertainty/determining the accuracy of the data of a small sample from a large population. It answers the question: would the data be the same (or similar at least) if the sample had been different?

- A **Monte Carlo Simulation** includes taking a large number of different random samples and seeing how they compare to the original sample chosen. The final product is a **sampling distribution** that shows the variance of the Monte Carlo samples (or as Dr. Scott put it: "how a statistic changes from sample to sample")
- If the sampling distribution is centered at the truth it is called "unbiased"
- **Standard error** – the standard deviation of the sampling distribution

10:15 (Homework #3)

Because we will rarely actually know the whole population, the above exercise will rarely work because it requires the whole population

To address this problem we use **bootstrapping** – a process that treats the sample that we have as a stand in for the population and run a sampling distribution on this data.

- Bootstrapping helps you assess the accuracy of your sample data to the population data without knowing the population
- To do this, you take samples of 251 from your original sample of 251 with replacement (if you didn't use replacement, you would just get your original sample over again.

- If you do this repeatedly (say 1000 times) you can create a **bootstrapped sampling distribution** from which you can find **bootstrapped standard error.**

10:30 (more bootstrapping explanation)

If you have at least 30 data points per parameter, you can trust your bootstrapping procedure.

- Remember however, that this is not the same as the population distribution. It is merely an estimate.

To find confidence interval of your bootstrapped population estimate, just look at the bootstrapped sampling distribution.

Frequent coverage property – we will discuss this next class

10:45 – Class is out

Summary – the most important thing we did today is learn how to determine and measure uncertainty by running the thought experiment (what would have happened if we had taken a different sample from the population?). To do this, even when we don't have any knowledge of the population outside our sample, we can use bootstrapping.