ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: “Big” Data Analytics
Final Report

Mary Latimer
GTID 902868975

Since I started working for Coca-Cola, the drink preferences of people has been a real interest to me. I grew up around Atlanta, and all of my friends and family growing up preferred Coke. Coming to college, I met a lot of people who preferred Pepsi, who were also from out of state. This made me wonder if Coke vs Pepsi is a regionally induced preference, since the birthplace of Coke is in Atlanta. I think that a greater proportion of people who grew up in Southeastern states prefer Coke as compared to people who grew up in non Southeastern states.

I made a survey collecting data on: drink preferences growing up, drink preferences today, and the state in which they grew up. I collected data from my peers, by sending an online survey out to them through multiple Facebook groups at Georgia Tech. Since the data could possibly be very biased and unbalanced because we are in the Southeast, I think Georgia Tech is a good place to sample because we have so many people out of state and international. To make sure that I have collected unbiased and balanced data(because we are in Atlanta and there are more people from the Southeast than anywhere else, I capped the data collection at 22 samples for each set of data: Southeast and non-Southeast.
My survey questions were: What soda preference did you grow up with? ; What is your soda preference now?

Since I collected qualitative data, to make it easier to analyze, I set Coke=0 and Pepsi=1.

I planned to do Hypothesis testing to analyze the results. The null hypothesis is that there is no difference in the proportion of Southeasterners and non-Southeasterners that prefer Coke. The alternate Hypothesis is that p2 will be greater than p1 (since Coke=0 and Pepsi=1,

and the sampling average will be higher of non-Southeasterners so p2-p1>0). I think that this hypothesis will be true. I will test at alpha=.05, 95% confidence level.
Ho=p2-p1=0
Ha=p2- p1>0
I did 2 separate tests, since I collected 2 sets of information: drink preferences growing up, and drink preferences now.
growing up: p1=0.04347826087; p2=0.1363636364 S=0.3178208631
now: p1= 0.04347826087; p2= 0.1818181818

we will reject if $t_0>t_{\alpha,df-1}$, $t_0>t_{.05,43}$,
Since $t_{.05,43} \approx 2.021$, we can reject the Ho.

t test of “growing up” statistic
$(\bar{x}-0)/(S/\sqrt{n})$; p1: $(0.0929-0)/(0.395/\sqrt{44})=1.56=t_0$
Since $t_{.05,43} \approx 2.021$, we can reject the Ho, since $t_0>t_{.05,43.}$

t test of “now” statistic:
$(\bar{x}-0)/(S/\sqrt{n})$; p1: $(0.1383399209-0)/(0.3178/\sqrt{44})=2.89=t_0$
Since $t_{.05,43} \approx 2.021$, we can reject the Ho, since $t_0>t_{.05,43.}$

To interpret this, from the data collected, it is reasonable to assume that nonSoutherners prefer Pepsi **now** at a higher rate than Southerners do. This makes sense since Coke was invented and headquartered in Atlanta and is heavily marketed here.

t test of “growing up” statistic
$(\bar{x}-0)/(S/\sqrt{n})$; p1: $(0.0929-0)/(0.395/\sqrt{44})=1.56=t_0$
Since $t_{.05,43} \approx 2.021$, we cannot reject the Ho, since $t_0<t_{.05,43.}$
To interpret this, from the data collected it is not reasonable to assume that Southerners preferred Pepsi or Coke **growing up** at a higher rate than non-Southerners did.

These conclusions are somewhat surprising. The results of the “now” test is what I expected— that Southeasterners prefer Coke to Pepsi at a higher rate.
However, when they were growing up, there seems to be no difference in the preference.