Week 9 Live Session

w203 Instructional Team

Announcements

Feedback about Quiz 1.

We made a lot of changes to week 9 this semester. We're trying new activities out, so let us know what you think!

Common hypothesis testing errors

For each of the following scenarios, explain the key error in the statistical procedure.

a) Bill hypothesizes that the average student drinks between 100 and 200 grams of caffeine during a take-home lab. He measures mean caffeine intake for a random sample of 50 lab-takers, then computes the p-value associated with his hypothesis.

b) Mike likes peanuts. Mike likes peanuts so much that he conducts a study to show how peanut allergies are an NIH sponsored hoax. He recruits 20 toddlers and randomly assigns each into two groups: peanut butter and brown sugar paste. To Mike's delight, he fails to find evidence for a difference between the groups (p = .34). Mike concludes by accepting the null hypothesis (that peanut allergies do not exist).

c) Anne replicates Mike's study and estimates a p-value of .03, she concludes that the alternative hypothesis has a 97% chance of being true.

d) Tim asks 50 passengers on the 8am Staten Island Ferry to complete his survey about attitudes toward atheists. He finds a statistically significant difference between attitudes toward atheists and attitudes toward scientologists (p = .04). Huzzah! Tim concludes that the US public is more fearful of atheists than scientologists.

Comparing Means

The file united_states_senate_2014_v2.csv contains data on the 100 members of the US senate that served in 2014. We will consider this group to be a sample (for example, from some generative process that creates senators).

```
S = read.csv("united_states_senate_2014_v2.csv")
summary(S)
```

```
##           Senator.Names    Gender          State            Party
##  Alan Franken    : 1    Female:20   Alabama   : 2   Democrat   :53
##  Amy Klobuchar   : 1    Male  :80   Alaska    : 2   Independent: 2
##  Angus King      : 1                Arizona   : 2   Republican :45
##  Barbara Boxer   : 1                Arkansas  : 2
##  Barbara Mikulski: 1                California: 2
##  Benjamin Cardin : 1                Colorado  : 2
##  (Other)         :94                (Other)   :88
##             Religion  Campaign.Money.Raised..millions.of...
##  Protestant     :49   Min.   : 0.100
##  Catholic       :27   1st Qu.: 4.575
```

```
##  Jewish         :10   Median : 7.550
##  Other Christian: 7   Mean   : 9.645
##  Mormon         : 2   3rd Qu.:13.800
##  Unaffiliated   : 2   Max.   :44.200
##  (Other)        : 3
##  Campaign.Money.Spent..millions.of...   NRA.Rating
##  Min.   : 0.200                        A      :34
##  1st Qu.: 2.975                        F      :34
##  Median : 6.000                        A+     : 9
##  Mean   : 8.227                               : 5
##  3rd Qu.:12.225                        AQ     : 5
##  Max.   :43.400                        C      : 3
##                                        (Other):10
```

You have three questions that you would like to answer with a statistical test. For each question, answer the following using the dataset and your background knowledge:

1. Are the assumptions for a t-test met? (you may want to review unit 9.5)
2. Is a paired test or an unpaired test more appropriate?
3. (Unless you argue that a t-test is clearly invalid), conduct a t-test in R and interpret the results.

Question 1: Is there a difference between the amount of money a senator raises and the amount spent?

Question 2: Do female Democratic senators raise more or less money than female Republican senators?

Question 3: Does the NRA prefer male senators or female senators?

Reproducibility

1. You have a dataset of the number of Facebook status updates by day of the week. You run 7 different t-tests, one for posts on Monday (versus all other days), or for Tuesday (versus all other days), etc. Only the test for Sunday is significant, with a p-value of .045, so you throw out the other tests. Should you conclude that Sunday has a significant effect on number of posts? (How can you address this situation responsibly when you publish your results?)
2. As before, you have a dataset of the number of Facebook status updates by day of the week. You do a little EDA and notice that Sunday seems to have more "status updates" than all other days, so you recode your "day of the week" variable into a binary one: Sunday = 1, All other days = 0. You run a t-test and get a p-value of .045. Should you conclude that Sunday has a significant effect on number of posts?
3. Suppose researcher A tests if Monday has an effect (versus all other days), Researcher B tests Tuesday (versus all other days), and so forth. Only Researcher G, who tests Sunday finds a significant effect with a p-value of .045. Should you conclude that Sunday has a significant effect on number of posts?
4. What if researcher G above is a sociologist that chooses to measure the effect of Sunday based on years of observing the way people behave on weekends? Researcher G is not interested in the other tests, because Sunday is the interesting day from her perspective, and she wouldn't expect any of the other tests to be significant.
5. Many observers have noted that as studies yielding statistically significant results are repeated, estimated effect sizes go down and often become insignificant. Why is this the case?