Angela Meng
Hoai Truong 11AM-12:30PM

February 26, 2014

Hypothesis Testing

Brief Overview: Wednesday's class went over Hypothesis Testing and the Neyman-Pearson framework, two relatively familiar topics from STA 309H, but with much more depth. After looking at the current population data (cps85.R), we sought to find the reason why men were seemingly paid more than women, which introduced the concept of reshuffling **(permtest)**. The analogy of re-shuffling cards then begged the question:

If gender-wage could be reshuffled and reassigned like simple playing cards to mimic a "natural" non-association, in comparison could the difference found within the original data be **caused by chance**?

Data: cps85.R [current population survey- the alternative survey between censuses]
Explanation for difference between male and female wage

1. Men really were paid more than women—sample reflects this pattern
2. Unlucky sample—no real wage premium in the population
    - Boxplot of wage ~ sex, appears to be a wage premium for men
    - Remember the lm function gives you means in baseline offset form!
    - Alternate interpretation: we took an unlucky sample with mostly well-paid men and mostly low-wage paying women. May be attributed to sample variability.

Card Example: each person in the room has 2 cards—represent wage and gender

- Is there an association between the two?
- Take gender cards, shuffled them and dealt them back out and then calculated mean wage
    - Explicit way to break any connection between two variables
    - Very unlikely that each reshuffle will end with men= high salary, women= low salary
    - Breaking variable associations: collect cards, reshuffling, re-dealing them out
- **Back to R:**
    - Original Difference: 7.879 (F) vs 9.995 (M) [= 2.116]
    - Reshuffle1: 8.910 (F) vs. 9.121 (M)
    - Reshuffle2: 9.293 (F) vs. 8.796 (M)
    - Each reshuffle is caused purely by luck, it gives us an example of what a 'natural', completely random difference between F/M wage should be
    - The original difference of 2.116 looks highly unlikely to be caused by chance in comparison to the random reshuffles

Permutation test (shuffling test): creates a 1000 different samples

- R script: Permtest= do(1000)*mean(wage~shuffle(sex), data=cps85
  hist(permtest$M – permtest$F, xlim= c
- In conclusion: by looking at the histogram of the permtest and then comparing the original difference, the permutation test should be random showing a "natural" difference

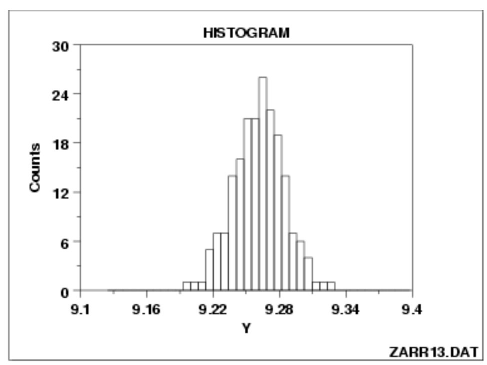
Angela Meng
Hoai Truong 11AM-12:30PM

Bush votes (2004) + Federal Aid Sample

- Looking at states that voted for Bush in 2004 (Red States) and the states that receive more federal aid than they pay in taxes (Green States), there seems to be a coincidental overlap. Two explanations:
    - Null Hypothesis: the states shown a random sample of states that Congress gave money out to
    - Alternative Hypothesis: the green states are a biased sample, showing a relation between federal aid and political alignment
    - 250,000 sample distribution of random map created by randomly drawing 25 states and then matching how many were “red states” in 2004
    - If the sample of interest is outside of the expected interval of chance, it gives strong evidence to believe that the sample’s result was not due to chance.
    - Results: highly unlikely, but still possible that the red/green state overlap was caused by chance

6 steps of Neyman-Pearson hypothesis testing framework (NO P-values)

1. Choose a null hypothesis **[H_o]**
    a. “There is no wage premium for men in the wider population”
    b. “Green states are a random sample of all states caused by chance”
        i. There’s nothing going on, everything is caused by random chance, no effect
2. Choose a summary measure (discrepancy measure, test statistic) **[t]**
   Designed to measure the discrepancy between the null hypothesis and
    a. $t = \bar{x}_{men} - \bar{x}_{women}$
    b. t = # of overlaps between red/green maps
3. Calculate (simulate) the sample distribution of t, assuming that H_o is true
   $P(t|H_o)$ – simulates a histogram
   Similar to bootstrapping, if we could take random samples one after another we would, but since we can’t…we just use reshuffling (permtest)



4. Choose a rejection region **[R]**
   Entirely subjective- it’s all up to your preference!
    a. R = any difference outside +/- $1
5. Calculate the size of R…literally **[α]**
   Size(R) = .05 = α
   Fraction of $P(t|H_o)$ that falls in R
6. Look at <u>your</u> data, and check whether <u>your</u> t fell in <u>your</u> R
   If t falls in R, reject H_o
   If it doesn’t…then don’t!

You can never fully reject the null hypothesis as a possible answer, (you can only hold off breaking up with it.)

Angela Meng
Hoai Truong 11AM-12:30PM

Decision	H_o is TRUE	H_o is FALSE
Reject	**Bad:** False Positive Type I Error α error	**Good: (even better)** **ヽ(*・ω・)ﾉ**
Fail to Reject	**Good:** ☺	**Bad:** False Negative Type II Error Β error

False Positive: Get a positive pregnancy test, but you're not actually pregnant (aka fake scare)
False Negative: You have cancer, but the test doesn't detect it…which is obviously really bad

Balance between False Positive/False Negative: the wider your R (rejection region) is the less likely you'll have a False Positive, BUT the more likely you'll make a False Negative

Smaller rejection region = a more conservative test with less False Positive errors, but more False Negative errors

$$\alpha = P(t \text{ falls in } R \,|\, H_o \text{ true})$$

Study break: once upon a time, Professor Scott's mother called 50 cent, "50 pence"
…now back to statistics:

Juries/trials analogy:

- higher standard – convict less people overall, but let more guilty people free
- Lower standard- convict more people overall, but sentence more innocent people as guilty

How to answer hypothesis testing problems, walk through steps 1-6, don't worry about p-values for now

Wrap Up:
Return to cps85.R, search for an explanation for why women are seemingly paid less than men.
Confounding variables=

- Unions- are men paid more because they are systematically more involved in unions rather than just because they're men?
    a. Use linear modeling: after adjusting for union membership, sexM premium becomes 1.9012- a much smaller difference than before
    b. Keep Union and Wage cards, re-collect gender cards and shuffle/re-distribute them

Permtest3 = do(1000)*lm(wage~union+ shuffle(sex), data=cps85)