STA371H Notes for Wednesday, February 18, 2015

SAP Case Study

- Enterprise software company that competes directly with Oracle
- *Nucleus Research* Study: "SAP customers are 20% less profitable than their industry peers."
    - Download SAP.zip. Read the *Nucleus Research* Study and try to replicate the claim put forth in the quote above. Then, evaluate that claim—do you agree, or is the claim misleading? Data is from sap.csv.
    - Replicating the Claim:
        - They took the average of the firm return on equity (12.6%) and that of the industry return on equity (15.69%). The "20% less profitable" claim was arrived by calculating a ratio between these two numbers.
        - R-Script Used:

```
library(mosaic)

sap <- read.csv("~/College/Spring
2015/STA371H/sap/sap.csv")

#Can you replicate the claim?
mean(sap$roefirm)
mean(sap$roeindustry)

#They basically just took groupwise means of the
ROE's.
```

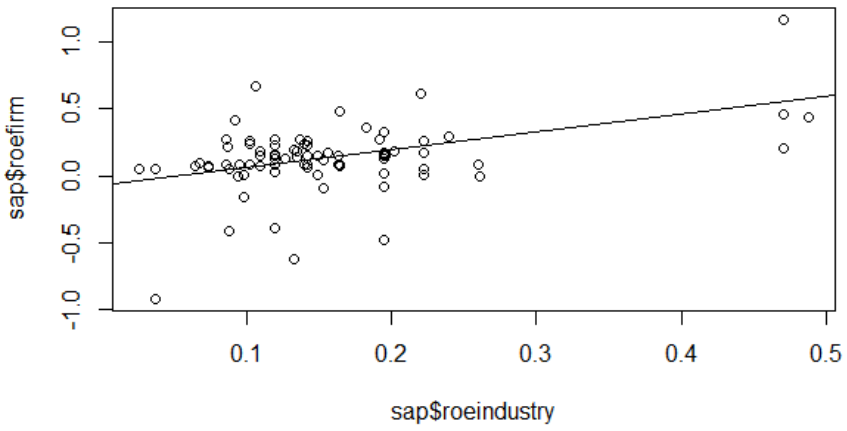
    - Evaluating the Claim:
        - If there were no systematic difference between SAP firms and industry firms, then a plot of the ROE's of the firms against the ROE's of the industries should theoretically yield a slope of one and an intercept of zero. However, a slope greater than one (1.219) and an intercept less than zero (-0.065) suggests that in comparing companies with small ROE's,

SAP firms seem to perform slightly worse than their industry peers, while in companies with large ROE's, SAP firms seem to perform better than their industry peers.

- R-Script Used:

```
lm1 = lm(sap$roefirm~sap$roeindustry)
summary(lm1)
plot(sap$roefirm~sap$roeindustry)
abline(lm1)
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.06504    0.05287  -1.230    0.222
sap$roeindustry  1.21925    0.29332   4.157 8.12e-05 ***
```



- To account for sampling variability, we need to calculate confidence intervals. These were obtained by multiplying the standard errors displayed above by two and adding to/subtracting from the calculated slope and intercept values.
    - Intercept: [-0.19, 0.06]
    - Slope: [0.68, 1.96]
    - Because the confidence intervals of the intercept and the slope pass through zero and one respectively, it cannot be concluded that

there is a systematic difference between the ROE's of SAP firms and those of their industry peers.

*The most difficult part of statistics is knowing how and when we should use the statistical tools we know in order to answer a question.

Frequentist Coverage Property (Fig 5.7 of Course Packet)

- This is what makes a confidence interval a confidence interval.
- Analogous to an assembly line: 99% of random samples coming off an assembly of phones work correctly, but the 99% refers to a property of the entire procedure rather than a property of an individual subject.
- If under repeated application of the same procedure, 80% of the intervals produced by this procedure contain the true value, then we can say that each individual interval has a confidence level of 80%.
- Bootstrapped confidence intervals do satisfy this property.

Normal Linear Regression Model

- $y_i = \beta_0 + \beta_1 x_1 + \epsilon_i$
- We can assume that the residuals follow some sort of normal distribution centered around zero:
    - $\epsilon_i \sim N(0, \sigma^2)$
- Similar to bootstrapping, this method also gets us error bars for our estimates.