# IS606 Homework 6

*Daniel Dittenhafer*

*October 27, 2015*

### 6.6 2010 Healthcare Law (p313)

**(a) We are 95% confident that between 43% and 49% of Americans in this sample support the decision of the U.S. Supreme Court on the 2010 healthcare law.** More accurately, I would say we are 100% confident that 46% of the sample support the decision.

**(b) We are 95% confident that between 43% and 49% of Americans support the decision of the U.S. Supreme Court on the 2010 healthcare law.** Given the 3% margin of error at 95% confidence and point estimate of 46%, the interval of 43% - 49% seems correct to me as an estimate of the population parameter.

**(c) If we considered many random samples of 1,012 American, and we calculated the sample proportions of those who support the decision of the U.S. Supreme Court, 95% of those sample proportions will be between 43% and 49%.** Not so much. Rather, 95% of those proportions would would contain the population proportion.

**(d) The margin of error at a 90% confidence level would be higher than 3%.** The reverse would be true. At a lower confidence level (90% vs 95%), the margin of error would narrow (be smaller) due to the use of a smaller Z score in margin of error computation.

### 6.12 Legalization of marijuana (Part I)

*The 2010 General Social Survey asked 1,259 US residents: "Do you think the use of marijuana shoudl be made legal or not?" 48% of the respondents said it should be made legal.*

**(a) Is 48% a sample statistic or a population parameter? Explain.** The 48% is a sample statistic because it is the portion of respondents to the survey… the sample.

**(b) construct a 95% confidence interval for the proportion of US residents who think marijuana should be made legal, and interpret it in the context of the data.** The following `R` code constructs the 95% confidence interval.

```
n <- 1259
p <- 0.48
z <- qnorm(0.975)
se <- sqrt( (p * (1-p) )/ n)
# Assuming the sample observations are independent, and
# checking success-failure conditino of inference.
succeses <- p * n
succeses > 10
```

```
## [1] TRUE
```

```r
failures <- (1-p) * n
failures > 10
```

```
## [1] TRUE
```

```r
# Compute the margin of error
me <- z * (se)
me
```

```
## [1] 0.02759672
```

```r
# Construct the 95% Confidence Interval
ci95 <- data.frame(lb=p - me, ub=p + me)
ci95
```

```
##          lb        ub
## 1 0.4524033 0.5075967
```

**(c) A critic points out that this 95% confidence interval is only accurate if the statistic follows a normal distribution, or if the normal model is a good approximation. Is this true for these data? Explain.** From the text, "the distribution of $\hat{p}$" is nearly normal when the distribution of 0's and 1' is not too strongly skewed for the sample size." In our case, the sample proportion is close to 0.5, and so the distribution of 0's and 1's would not be skewed very much at all. Additionally, we have already checked the success-failure condition.

**(d) A news piece on this survey's findings states, "Majority of American's think marijuana should be legalized." Based on your confidence interval, is this news piece's statement justified?** I don't believe the news piece's statement is justified. The proportion of American could be as low as 45.2%.

### 6.20 Legalize Marijuana, Part II

*As discussed in Exercise 6.12, the 2010 General Social Survey reported a sample where about 48% of US residents thought marijuana should be makde legal. If we wanted to limit the margin of error of a 95% confidence interval to 2%, about how many Americans would we need to survey?*

```r
p <- 0.48
me <- 0.02
z <- qnorm(0.975)

se <- me / z

# Compute the required n
n <- (p * (1-p)) / se^2
n
```

```
## [1] 2397.07
```

We would need to survey 2398 Americans to achieve a 2% margin of error for a 95% confidence interval.

### 6.28 Sleep deprivation, CA vs OR Part I (p319)

*Calculate a 95% confidence interval for the difference between the proportions of Californias and Oregonians who are sleep deprived and interpret it in context of the data.*

```r
pCA <- 0.08
pOR <- 0.088
pDiff <- pOR - pCA
nCA <- 11545
nOR <- 4691
# Compute standard error and margin of error for the proportion difference.
SE <- sqrt( ((pCA * (1 - pCA)) / nCA) +  ((pOR * (1 - pOR)) / nOR))
me <- qnorm(0.975) * SE
# Construct the 95% confidence interval.
ci95 <- data.frame(lb=pDiff - me, ub=pDiff + me )
ci95
```

```
##             lb         ub
## 1 -0.001497954 0.01749795
```

The 95% confidence interval on the difference in proportions between CA and OR is -0.001498 - 0.017498. This interval overlaps 0, therefore we can conclude with a 95% confidence level that the proportions are not statistically different. In other words, CA and OR population proportion might be equal given the results from this sample.

### 6.44

### 6.48