Week 7: Central Limit Theorem and Confidence Intervals

Stat 201: Statistics I

March 10, 2019

Central Limit Theorem

R does not have built-in functions for simply calculating standard errors of sampling distributions, but the calculations are fairly simple.

- Example, the "population" of male Metro State students' heights (inches) has a mean of 67.42 and a standard deviation of 5.28. What are the mean and standard error of the distribution of means from 40 subject samples?

```r
s.mean <- 67.42
s.se <- 5.28 / sqrt(40)

s.se
```

```
## [1] 0.8348413
```

Confidence intervals

Similarly, there are not R functions for creating the kinds of confidence intervals we are concerned with here, but we can handle the calculations "by hand".

Recall, the formula for a confidence interval is:

$$CI\,(1-\alpha)\% = x \pm z_{\alpha/2} SE$$

Thus, the three values required are:

- the point estimate x
- the critical z value, defined by the confidence level, which in turned is defined by α
- the standard error of the sampling distribution

Point estimate

The point estimate is simply the statistic used as estimator derived from a sample.

- Find the point estimate x using a 40 subject sample, s:

```r
x <- mean(s)
x
```

```
## [1] 68.9
```

Critical values

Critical values can be found with the `qnorm` function as we have learned previously. It is also acceptable to use the known critical values for common values of α (i.e., for $\alpha = 0.05$, $z_{\alpha/2} = 1.96$).

- Find critical value for $\alpha = 0.01$ (99% confidence level):

```r
alpha = 0.01
crit.z <- qnorm(1 - alpha/2)
crit.z
```

```
## [1] 2.575829
```

Standard error

Often when we are finding confidence intervals, we do not yet have information about the population parameters (if we did, we would not have to do a confidence interval). Therefore, we use the sample standard deviation as an estimate for the population standard deviation. Then, the standard error can be calculated as in the previous slide.

- Find the standard error of the sampling distribution using a 40 subject sample, s:

```r
s.sd <- sd(s)
s.sd
```

```
## [1] 5.550699
```

```r
s.se <- s.sd / sqrt(40)
s.se
```

```
## [1] 0.8776425
```

Calculate confidence interval

Once the three values are given or calculated, a confidence interval can be constructed.

- Create a 99% confidence interval for the population mean from the sample *s*:

```
ci.lower <- x - crit.z * s.se
ci.lower
```

```
## [1] 66.63934
```

```
ci.upper <- x + crit.z * s.se
ci.upper
```

```
## [1] 71.16066
```

The 99% confidence interval is (66.64, 71.16).

Calculate confidence interval, alternate method

With R's vector arithmetic, we can create a confidence interval more simply, or at least with less code.

- Create a 99% confidence interval for the population mean from the sample *s*:

```
ci <- x + c(-1,1) * crit.z * s.se
ci
```

```
## [1] 66.63934 71.16066
```

This method results in a vector, where the lower bound is the first value and the upper bound is the second:

```
ci[1]
```

```
## [1] 66.63934
```

Confidence intervals for proportions

When creating confidence intervals for proportions, the point estimate (proportion from sample $\hat{p}$) and critical value are found the same way as for means. However, binomial standard deviations are calculated from the proportion value, so the standard error needs to be calculated from the sample proportion.

- Create a 95% confidence interval for the population proportion from a sample with a proportion of $\hat{p} = 0.38$ and a sample size of 36:

```r
crit.z <- 1.96
p.hat <- 0.38
p.se <- sqrt(p.hat * (1-p.hat) / 36)

ci <- p.hat + c(-1,1) * crit.z * p.se
ci
```

```
## [1] 0.2214404 0.5385596
```