Lecture July 27: Frequency distributions - Standard Errors and Z-Scores

Today: Frequency Distributions - Standard Errors and Z-Scores

We're working with the normal distribution here

Previously and how it fits

- 68-95-99.7 rule
- Need to tie sample to population **Today**
- Central Limit Theorem - ties sample to normal distribution
- Law of Large Numbers - ties the sample to the population conceptually
- Standard Error - ties sample measures to population
- Z-score - translates Standard Error to probability terms

Tying samples to populations

- Terms
  - statistics - sample
  - parameter - population
- We want to use the *sample* mean/median/etc to *estimate* the *population* version of the same value
- If you hear something like "We used the statistic to estime the parameter" it's the same thing!

Standard Error

- Quantifies the range around population value (parameter) for the sample value (the statistic)
- Usually the **mean** but you can figure a standard error for other statistics like the median
- Distance from the mean is a...

Standard Deviation

- The Standard Error of the Mean:
  - Standard deviation measures dispersion relative to the mean
  - Standard error measures dispersion between the sample mean and the population mean
  - Standard error of the mean is the standard deviation of the sample divided by the square root of the sample size:

$$\frac{s}{\sqrt{n}}$$

Z-Score

- Number of standard deviations (standard errors) from the mean
- Why not just follow the 68-95-99.7 rule and use 2?
  - 68-95-99.7 was just an approximation. Actual value is 95.45%
  - So the precise 5% level for Z-score is 1.96
  - The second answer is - 2 **is** the mental shortcut

```
data1 <- cars
model <- lm(dist ~ speed, data = data1)
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(model, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                 dist
## -----------------------------------------------
## speed                         3.932***
##                                (0.416)
##
## Constant                      -17.579**
##                                (6.758)
##
## -----------------------------------------------
## Observations                     50
## R2                              0.651
## Adjusted R2                     0.644
## Residual Std. Error       15.380 (df = 48)
## F Statistic           89.567*** (df = 1; 48)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```