New Draper Office Hours: MWF 11:45-12:45
BE 367C

Standard Deviation:
Empirical Rule: helps to estimate SD

ex.  
SD 0? too small
SD 50? too large (all graph)

\[
\begin{array}{c}
\text{mean} \\
30 \\
50 \\
70 \\
\text{mean} + 2 \text{SDs}
\end{array}
\]

Should be \( \approx 95\% \): no, still too small

SD = 20, looks about right

\[
\begin{array}{c}
\text{mean} \\
10 \\
30 \\
50 \\
50 + (2 \cdot 20)
\end{array}
\]

Empirical Rule for estimating SD
always starting at the mean: should capture:
go 1 SD either way 2/3 of data, 68%
go 2 SD either way most/95% of data
go 3 SD either way tail 99.7% of data
Normal Curve / Gaussian Distribution

\[ f(y) \]

Looking at our butterfly data set -
What % \( \leq 3.56 \) cm?  
\( \sim 8\% \)
Count 2124 = 8.3%  
This is the exact answer.

If we want the approximate answer:

The normal curve is dependent on the value of the mean and the SD.

Standard Normal Curve

Using Empirical Rule

\( \sigma = 1 \)  
mean: 0

Using the Empirical Rule:  
68% \( \sigma \leq 1 \)
68% \( \sigma \leq 2/3 \) yes, part 1

\( \{ \text{See pages L-34 and L-36 for Standard Normal Table. Table gives decimals} \} \)

FACTS:
1. All normal curves are symmetric
2. Total area under each normal curve is 100% = 1
3. All normal curves satisfy Empirical Rule exactly
Bc curve is symmetric, area to left of $-1 = area right of $+1$; both $= 16\%$

Total area must add up to $100\%$.

To calculate the actual value for the area under curve $[1, 1]$

$100\% - (2 \cdot 16\%) = 68\%$

$84\% - 16\% = 68\%$

Should be equal!

Example:

Wing Length

$X = \frac{3.66 \text{cm} - 4.0 \text{cm}}{0.3 \text{cm}} = \frac{\# - \text{mean}}{\text{SD}}$

$= -1.47 = X$

So now we want to know area to left of $-1.47$. 

\(\text{(Standard units } z)\)

\[\text{Converting to std units } \frac{Y - \bar{Y}}{s}\]

\(\text{Note: units should cancel, no units}\)
Using negative chart,
\(-1.4 \quad 0.07 \rightarrow 0.0708 \approx 7\%

Section 2: Experimental Design

Controlled Experiment = control group (C) and a treatment group (T), and the experimenters control who goes into which group(s).

"Is this difference practically significant?"
Comparing Data 1 absolute 2 percentage

We want the subjects in T to be similar to C (as similar as possible) in all relevant ways except for the T/C distinction.

To encourage similarity: randomize

Randomized Controlled Trial (RCT) Experiment
Flowchart for Classifying Experimental Designs

Was a comparison made between 2+ treatment groups? no → "experience not experiment" (neither)
o → observational study: vital to control for PCFs

Did investigators have control over who got into T and C groups no →

Controlled Experiment: Did investigators assign subjects to T and C at random? no →

judgmental allocation: suspect bias

yes → Randomized Controlled Trial: strongest design

Bias: systematic tendency to over/underestimate the truth
Back to the cortex experiment

\[ Y \quad \text{outcome} \quad : \quad \text{cortex weight} \]
\[ X \quad \text{treatment} \quad : \quad T \quad \text{vs} \quad C \quad 1/0 \]
\[ Z \quad \text{potential confounding factor (PCF)} \]
\[ \text{Genetic background} \]

Positively associated: \( u \uparrow, v \uparrow \) on avg (vice versa)
Negatively associated: \( u \uparrow, v \downarrow \)

PCF: in an experimental setting w/ treatment variable \( X \) and outcome variable \( Y \), any third variable \( Z \) may plausibly be associated both w/ \( X \) and \( Y \)

ex. \( Y \) (outcome): cortex weight \( X \) (treatment) \( C: \text{depriv} \)
\( Z \) (genetics) \( \uparrow \quad \uparrow \quad \checkmark \)
\( Z \uparrow \quad \uparrow \quad \uparrow \quad \checkmark \)

PCFs are the enemy in experimental design bc they cause bias in conclusions.

How to defeat PCFs?
Hold them **CONSTANT** in TIC comparison.