AMS 7
Lecture 4
6/27/16

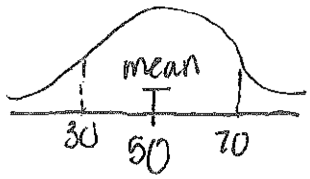
New Draper Office Hours: MWF 11:45-12:45
BE 357C

Standard Deviation:
Empirical Rule: helps to estimate SD

ex.

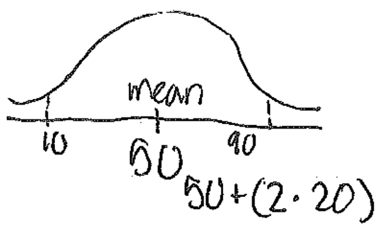
SD5? too small

SD50? too large (all graph)



mean ± 2 SDS

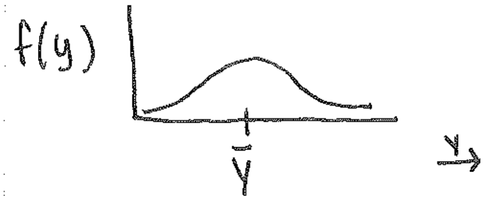
should be ~95%: no, still too small



SD = 20, looks about right

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	all 99.7% of data

Normal Curve / Gaussian Distribution

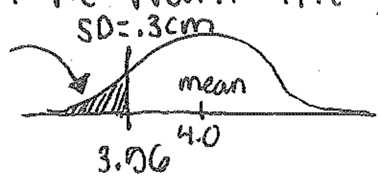


ex. Looking at our butterfly data set - What % ≤ 3.56 cm? ~8%

count 2/24 = 8.3%

this is the <u>exact</u> answer

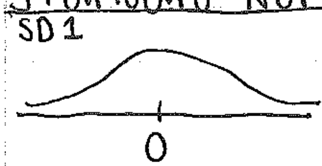
if we want the <u>approximate</u> answer:



butterfly wing length

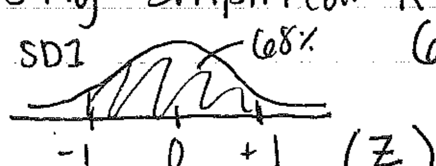
The normal curve is dependent on the value of the mean and the SD.

Standard Normal Curve



mean: 0
SD: 1

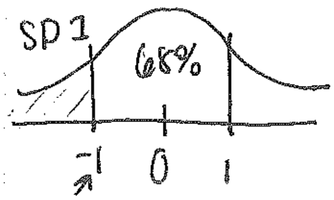
Using Empirical Rule



68% → 2/3 yes, part 1

{See pages L-34 and L-36 for Standard Normal Table. Table gives <u>decimals</u>}

FACTS:
1. all normal curves are symmetric
2. total area under each normal curve is 100% = 1
3. ALL normal curves satisfy emp. rule <u>exactly</u>



Negative value: -1.00

from table: .1587

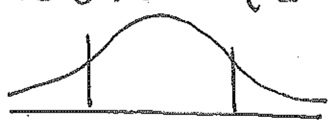
↑ table gives decimal = 15.87% ≈ 16%

BC curve is symmetric, area to left -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 · 16%) = 68%



Positive value: +1

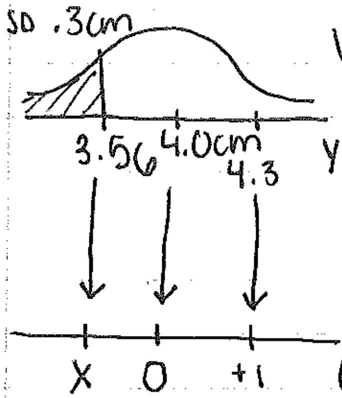
from table: .8413 = 84%

← 16%   16% →

84% - 16% = 68%

should be equal!

ex



Wing Length

Y (in raw units)

RAW ↓ Standard

(Standard units z)

$$X = \frac{3.56\,cm - 4.0\,cm}{0.3\,cm} = \frac{\# - mean}{SD} =$$

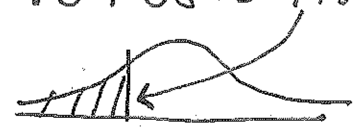
converting to stnd units
$\frac{Y - \bar{Y}}{S}$

$= -1.47 = X$

note: units should cancel, no units

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

Using negative chart,
-1.4 .07 → .0708 ≈ 7%



Section 2: Experimental Design

L-67

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

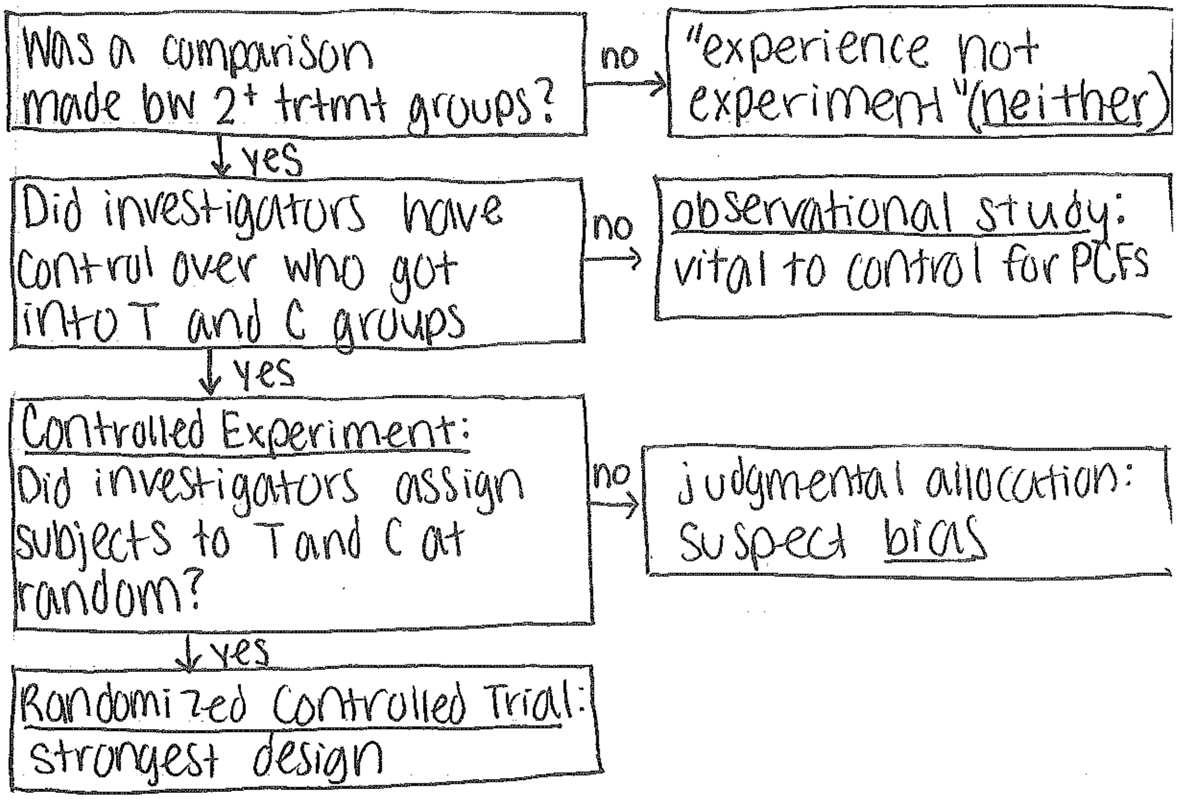
"Is this difference practically significant?"
Comparing Data ① absolute
② percentage

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

To encourage similarity: randomize

→ <u>Randomized Controlled Trial (RCT) Experiment</u>

Flowchart for Classifying Experimental Designs



Bias: systematic tendency to over/underestimate the truth

Back to the cortex experiment

- $\underline{Y}$ outcome : cortex weight
- $\underline{X}$ treatment : T vs C 1/0
- Z potential confounding factor (PCF) : genetic background (cortex weight)

Positively Associated: $u \uparrow$, $v \uparrow$ on avg (+ vice vers
Negatively Associated: $u \uparrow$, $v \downarrow$ " "

<u>PCF</u>: in an experimental setting w/ treatment variable X and outcome variable Y, any <u>third variable Z may plausibly be associated both w/ X and Y</u>

ex. Y (out come) : cortex weight X (trtmnt) T: enrich, C: depriv
Z (genetics)
as $Z \uparrow$ $Y \uparrow$ ✓
as X $C \to T$, $Z \uparrow \downarrow$ ✓

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

How to defeat PCFs? Hold them <u>CONSTANT</u> in T/C comparison.