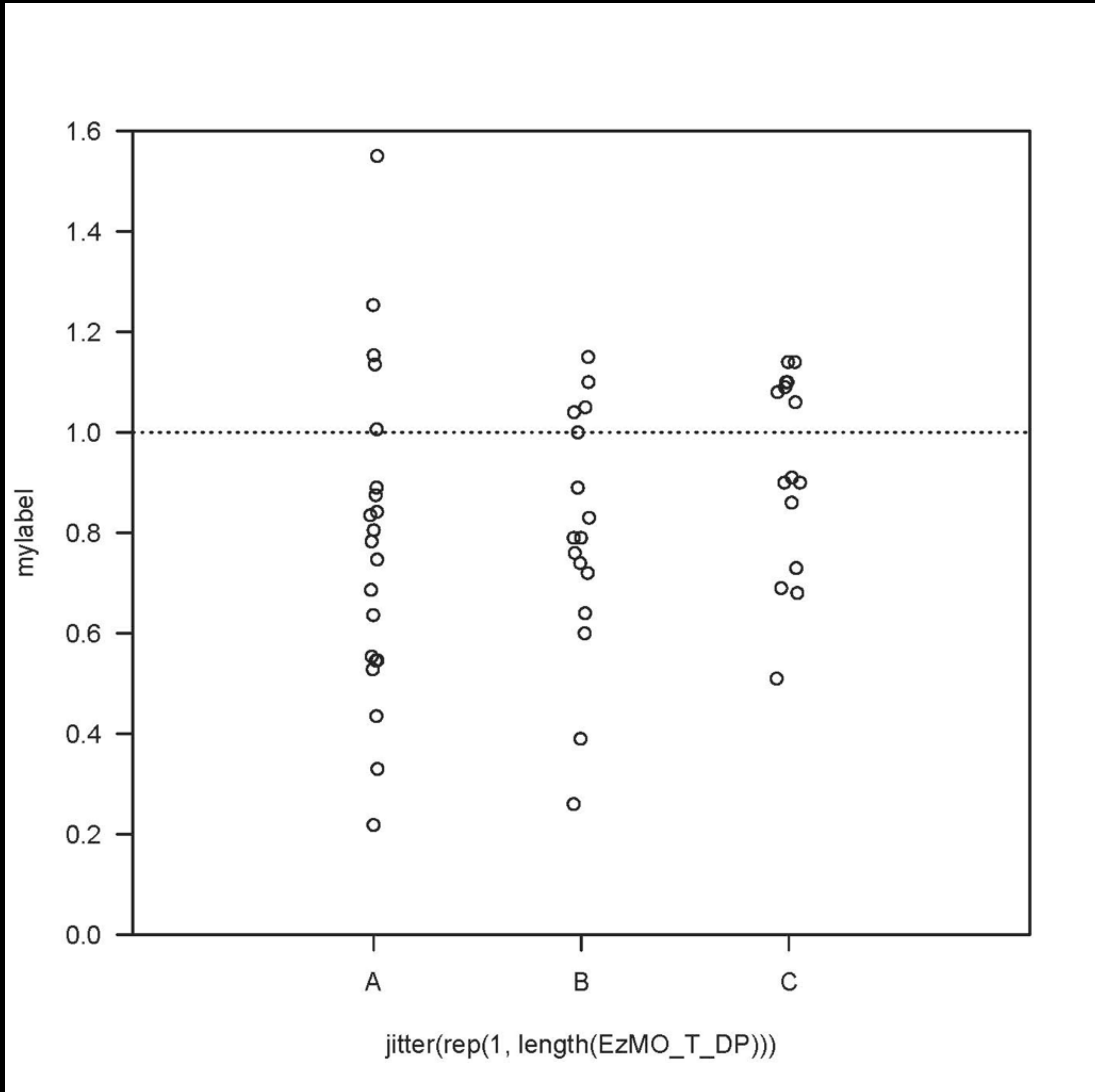
JIPP Lecture #4

Statistics Review

- Slides are at socialneuro.com\JIPP
- Follow me on Twitter: @socialneuro
- My email address: e.vanman@uq.edu.au

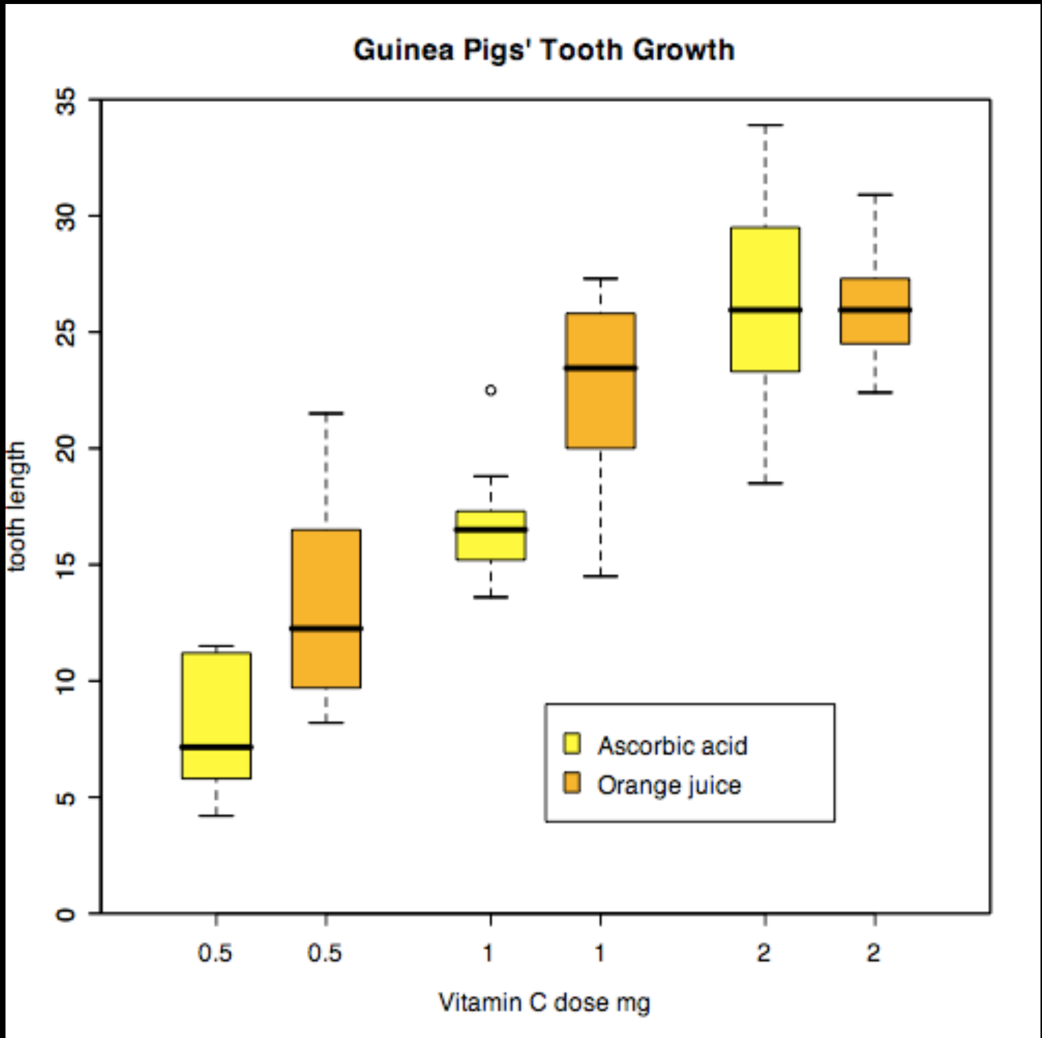
Three types of graphs that represent centre and spread of data

Dot plots

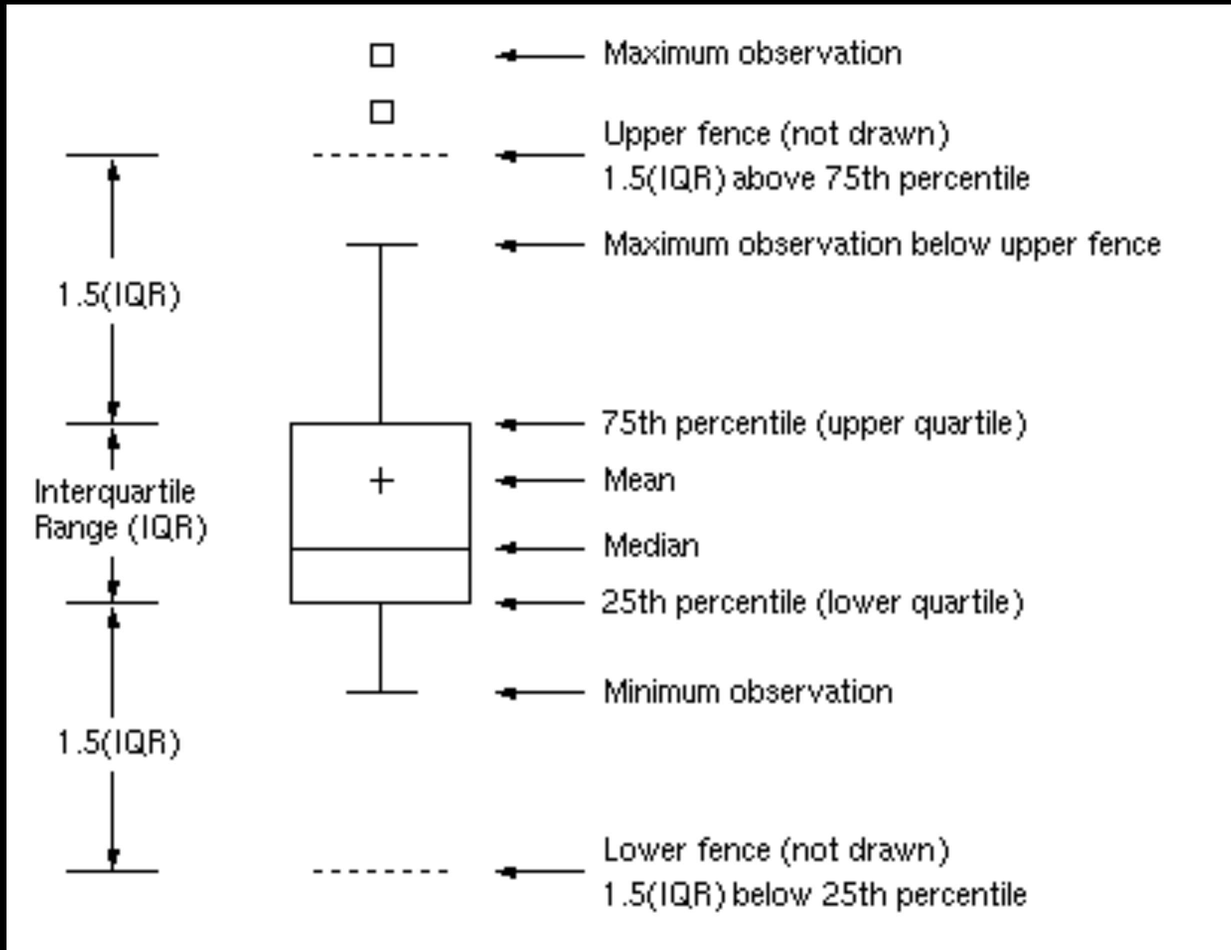


1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
mylabel
A
B
C
jitter(rep(1, length(EzMO_T_DP)))

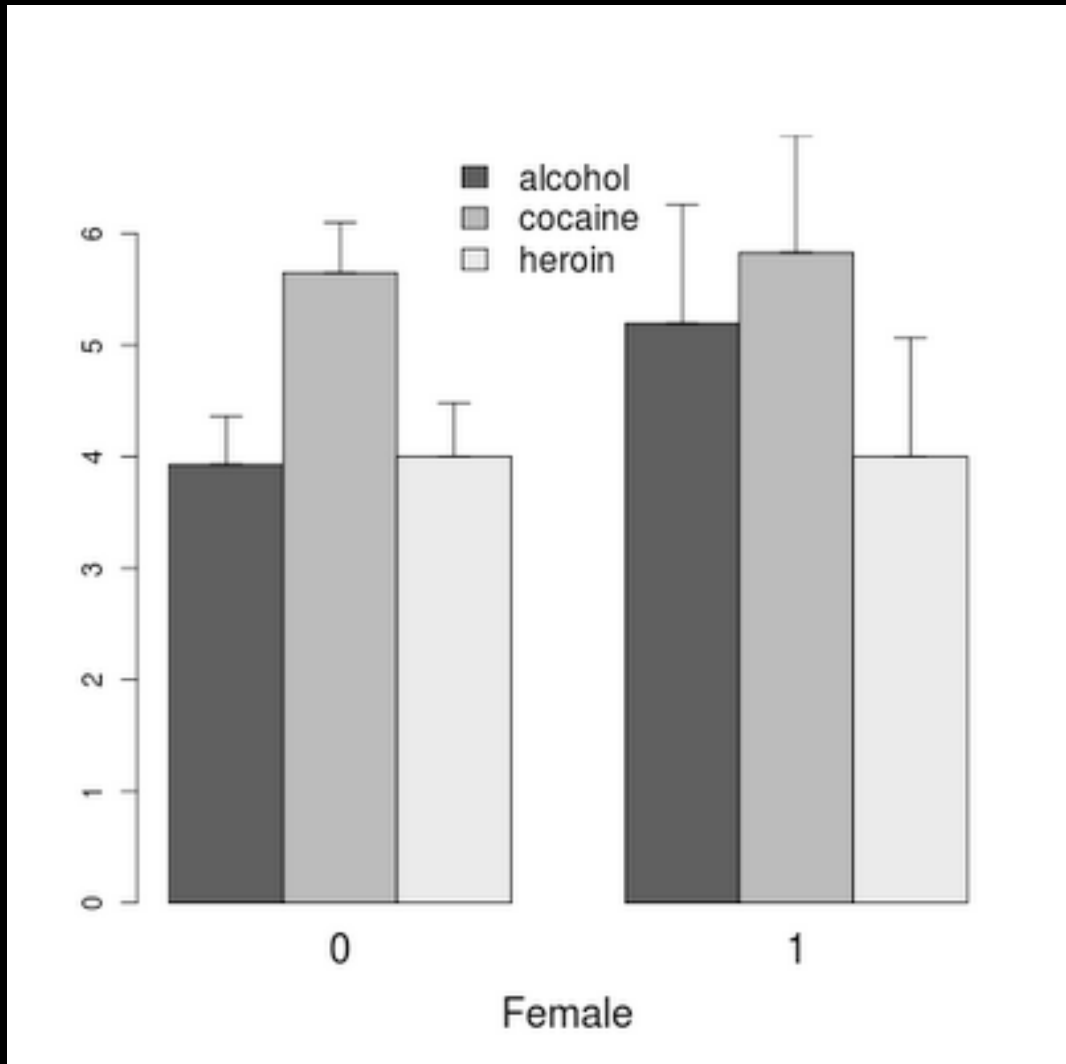
Box plots


Guinea Pigs' Tooth Growth
tooth length
0
5
10
15
20
25
30
35
0.5
0.5
1
1
2
2
Vitamin C dose mg
Ascorbic acid
Orange juice

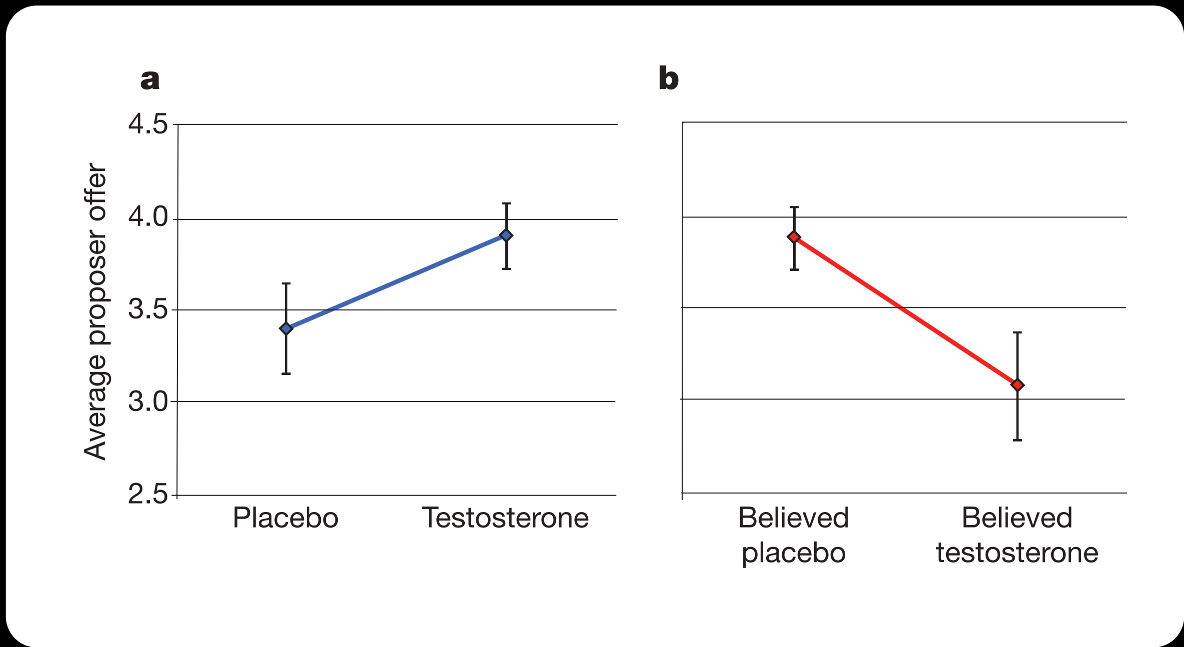
Box plots


Maximum observation
Upper fence (not drawn)
1.5(IQR) above 75th percentile
Maximum observation below upper fence
1.5(IQR)
75th percentile (upper quartile)
Mean
Interquartile
Range (IQR)
Median
25th percentile (lower quartile)
Minimum observation
1.5(IQR)
Lower fence (not drawn)
1.5(IQR) below 25th percentile

Barcharts

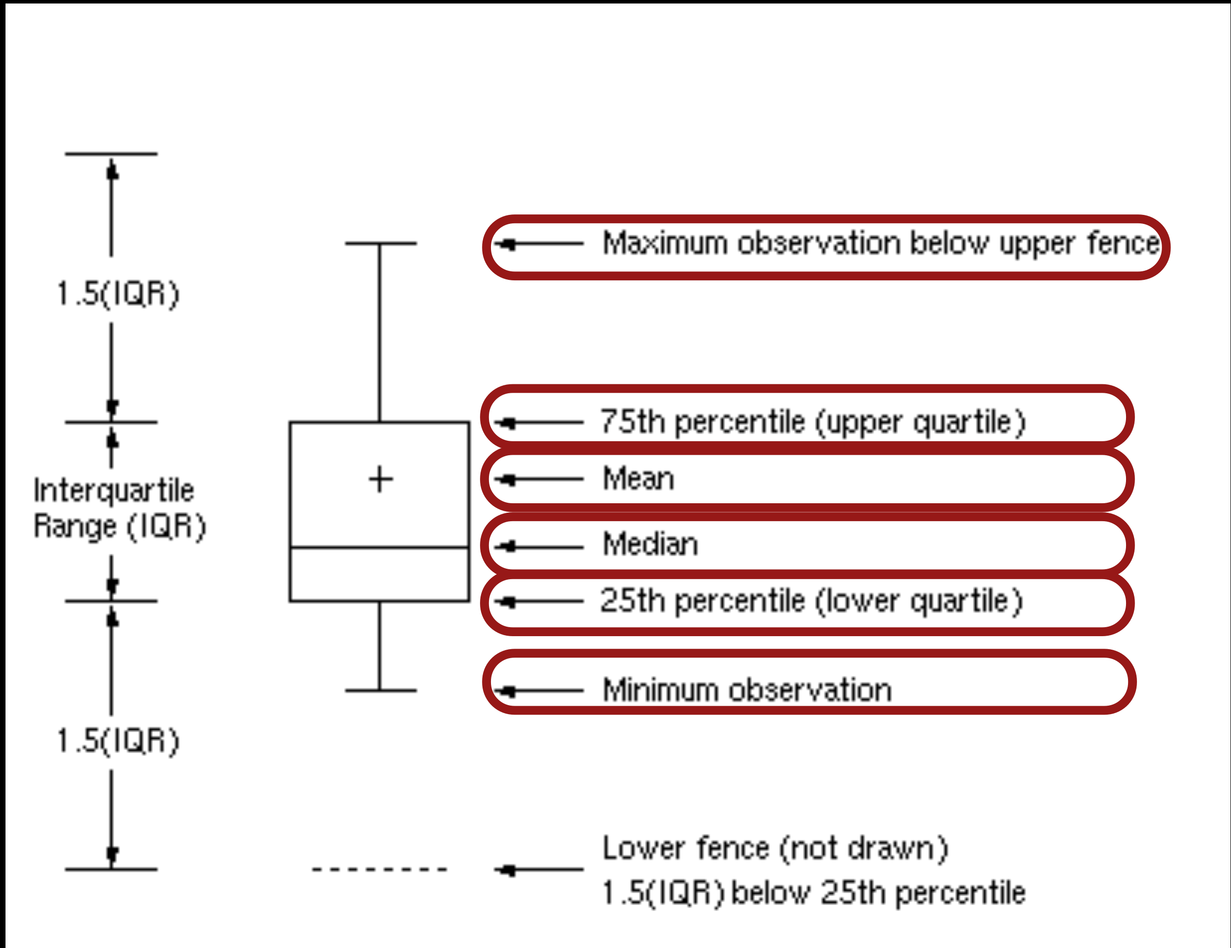


What about line graphs?

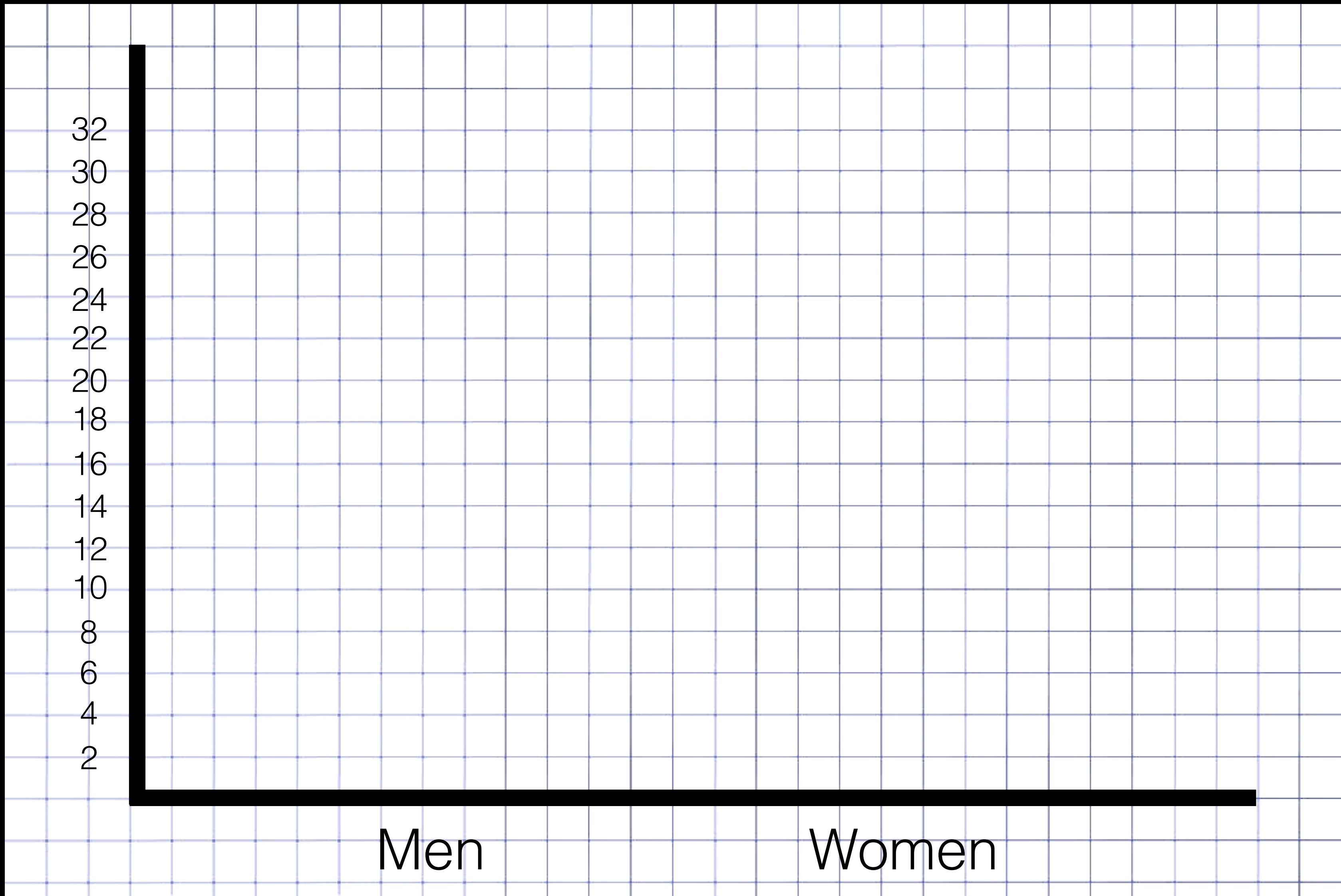


How to Make a Box and Whisker Plot

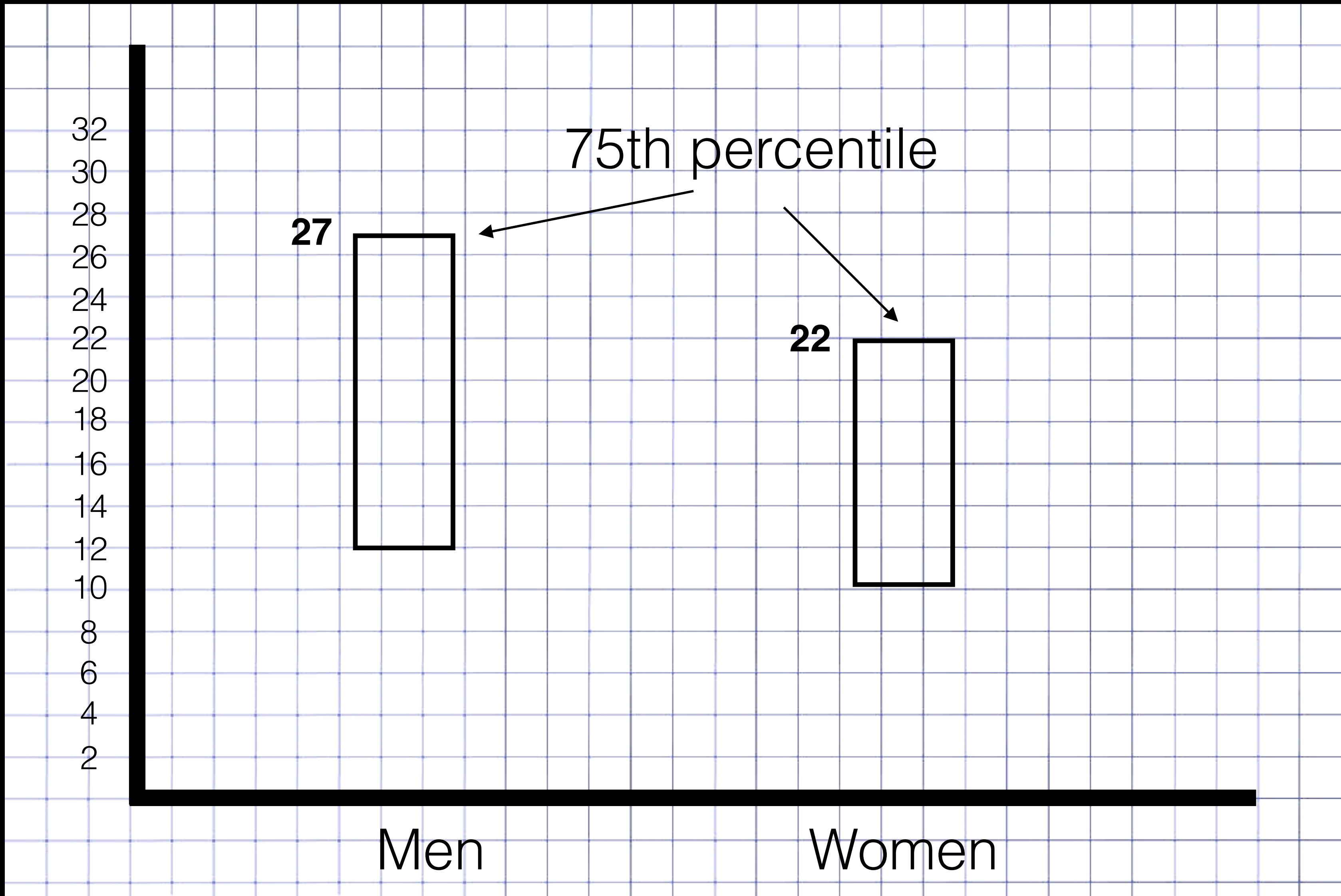
- You need the following six numbers:
  - 75th Percentile
  - 25th Percentile
  - Max
  - Min
  - Median
  - Mean


1.5(IQR)
Interquartile Range (IQR)
1.5(IQR)
Maximum observation below upper fence
75th percentile (upper quartile)
Mean
Median
25th percentile (lower quartile)
Minimum observation
Lower fence (not drawn)
1.5(IQR) below 25th percentile

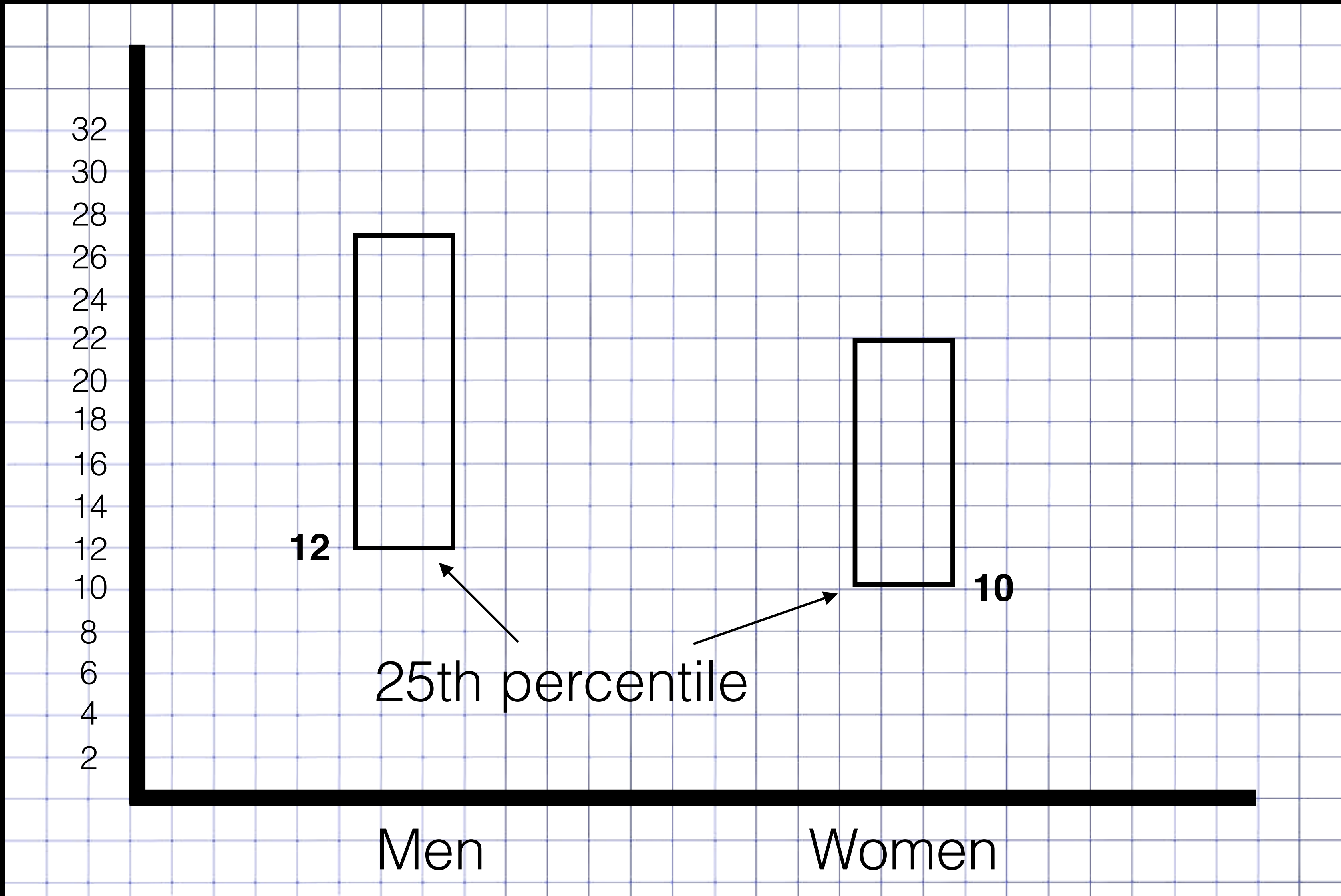
Drawing a Box Plot



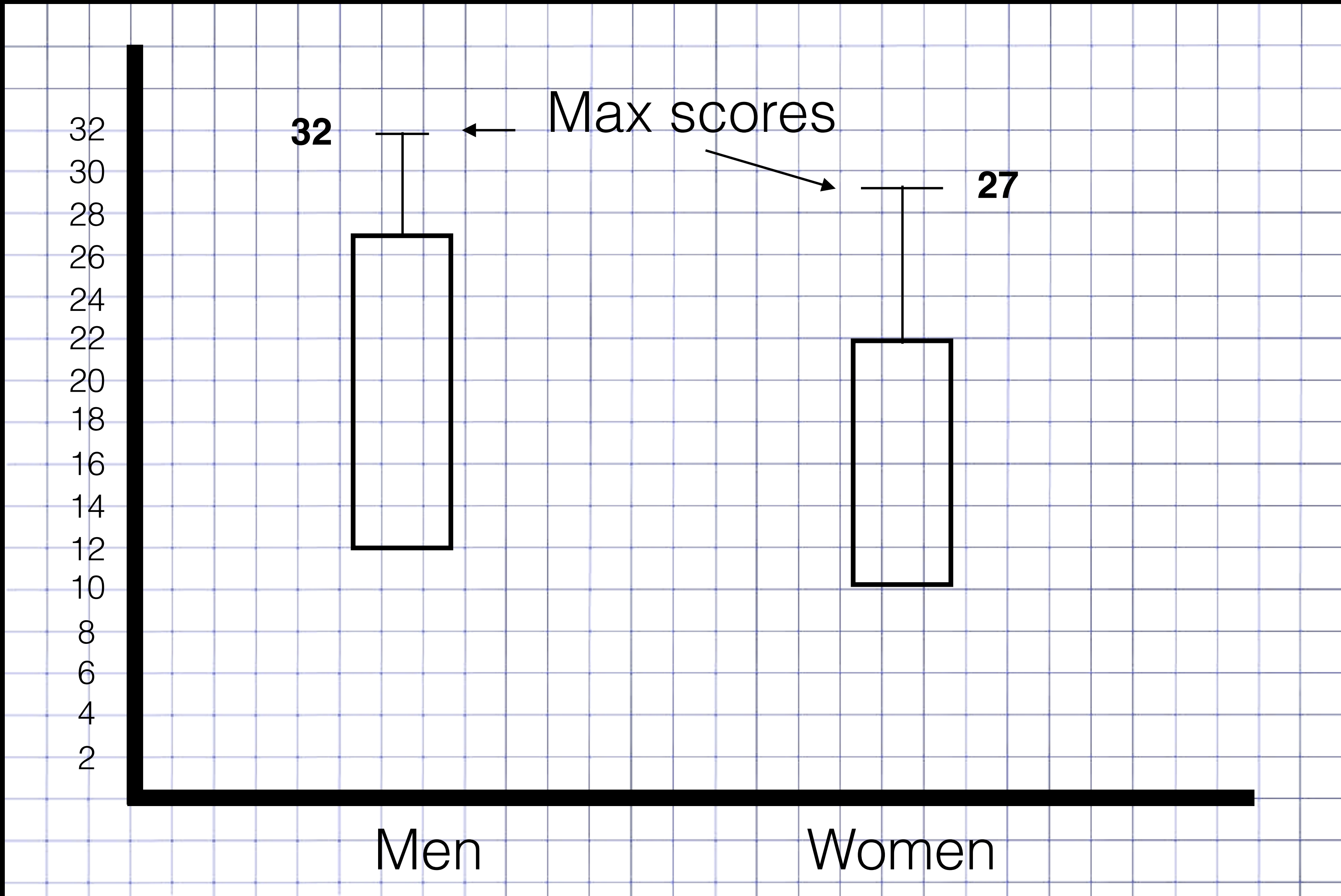
Drawing a Box Plot



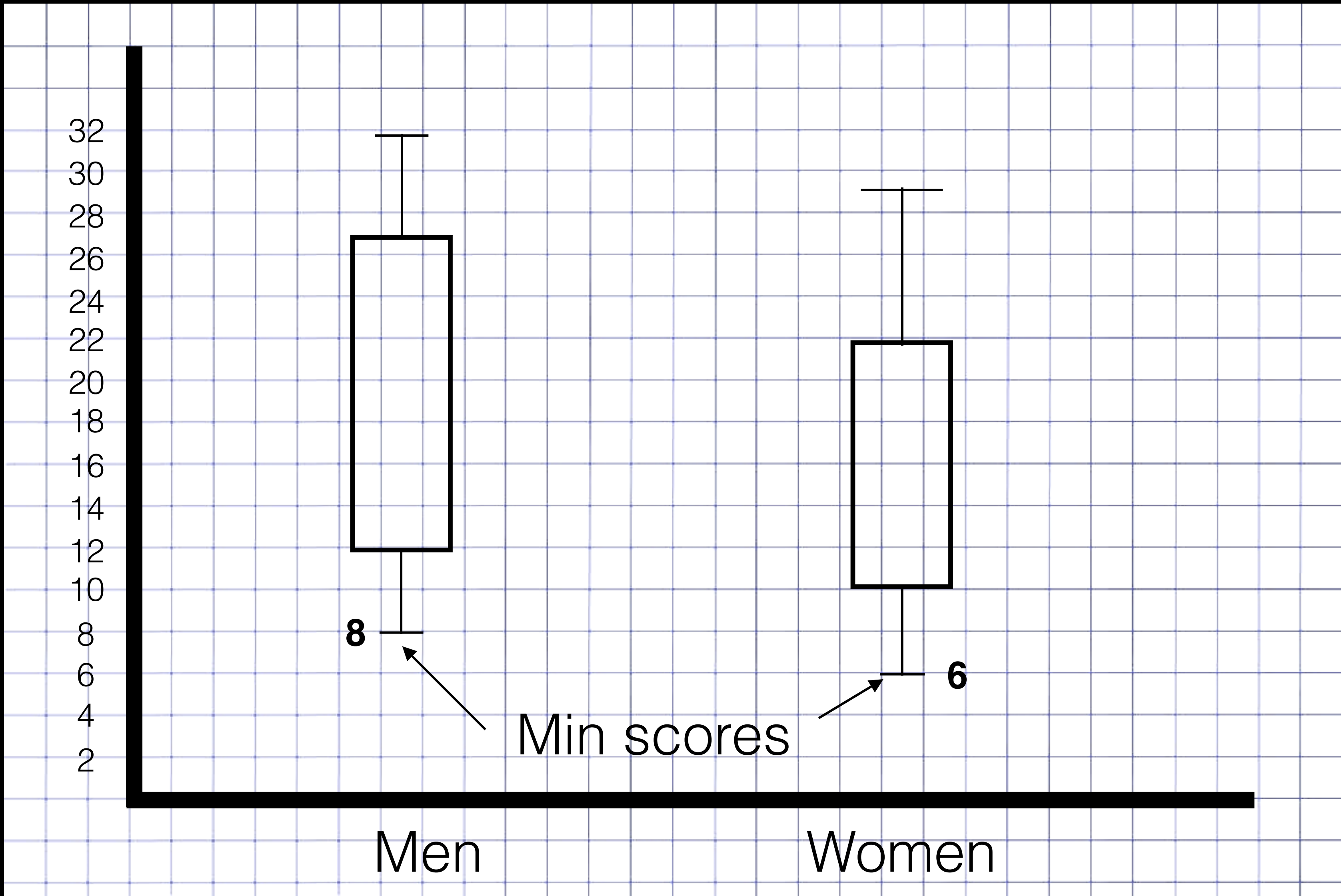
Drawing a Box Plot



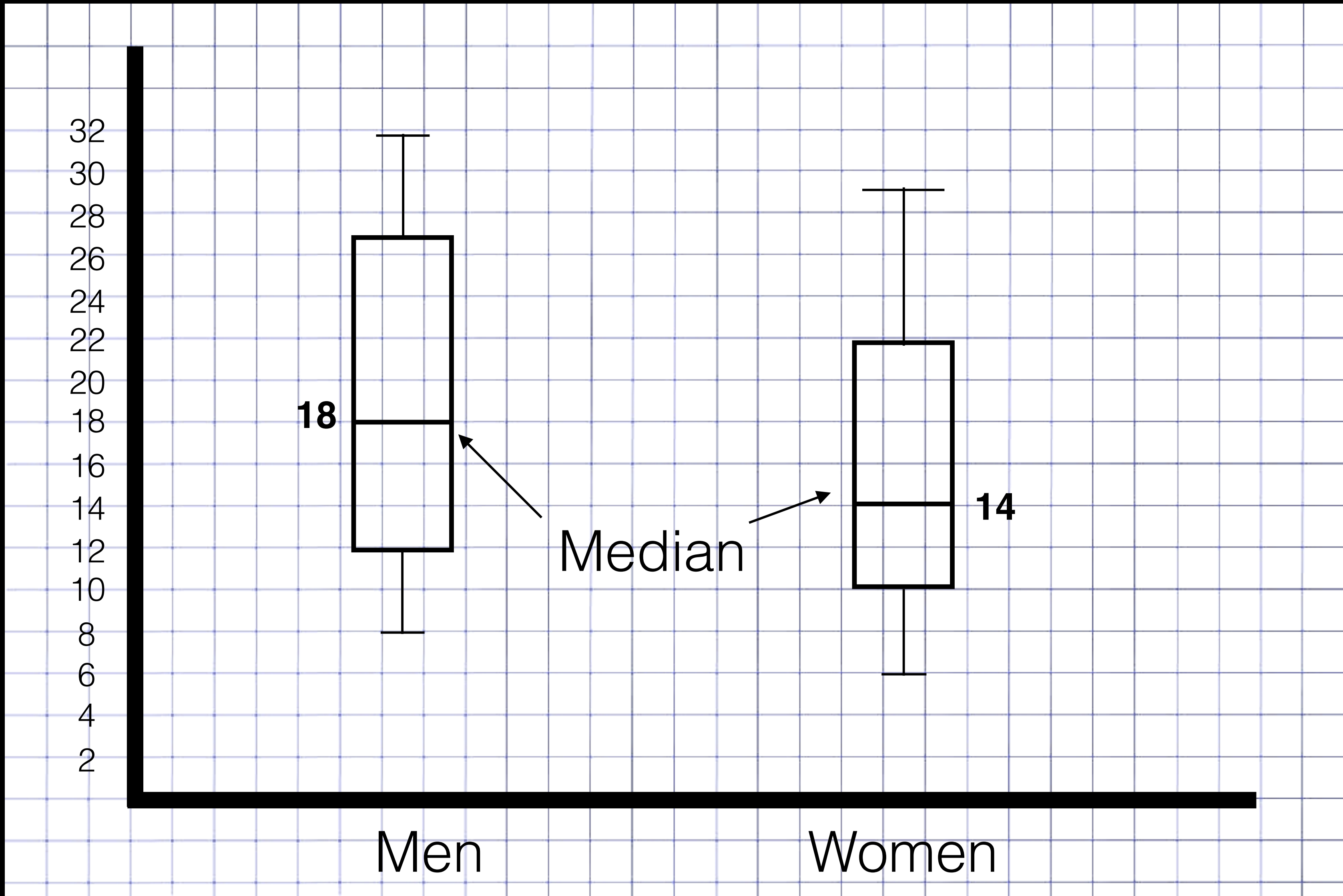
Drawing a Box Plot



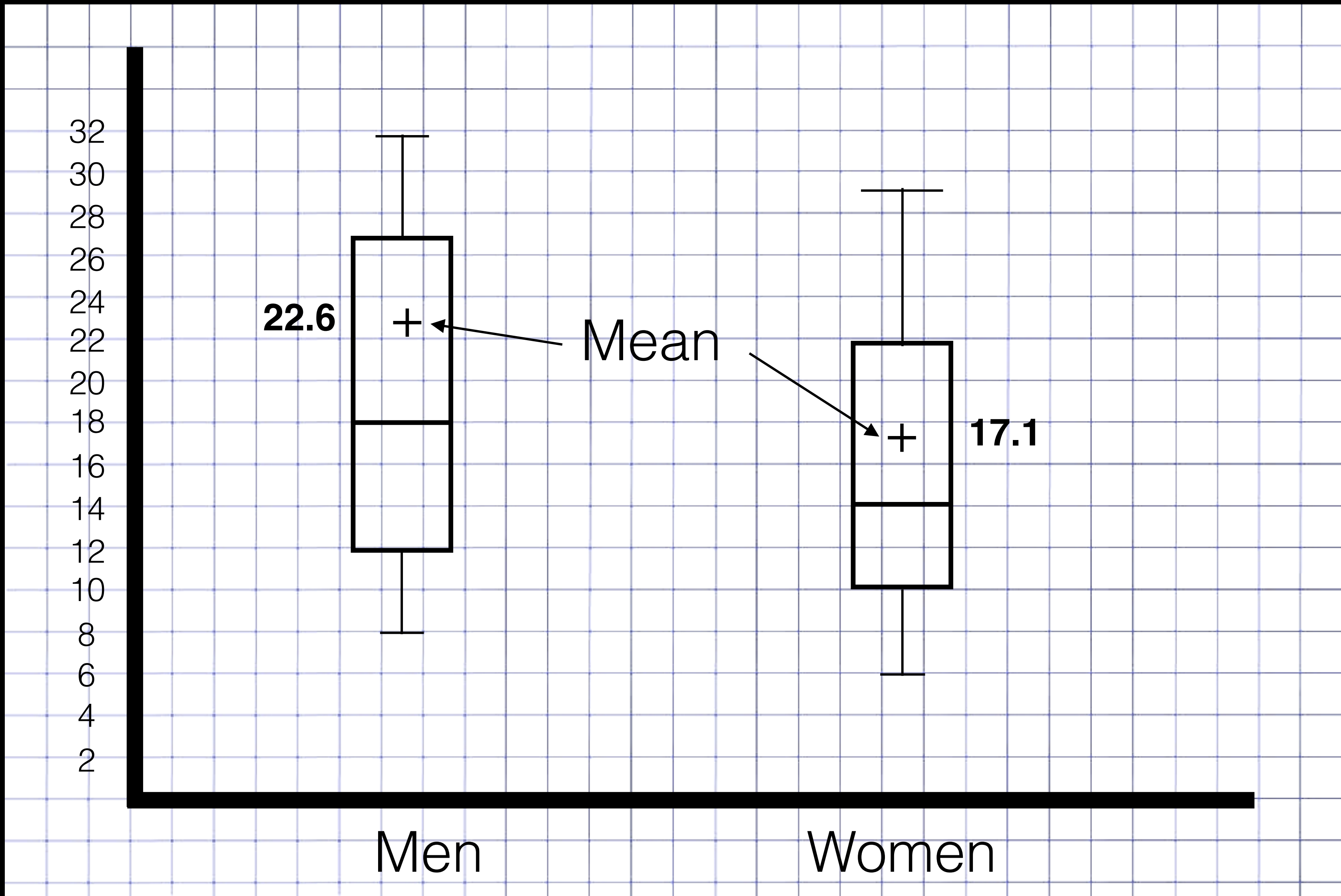
Drawing a Box Plot



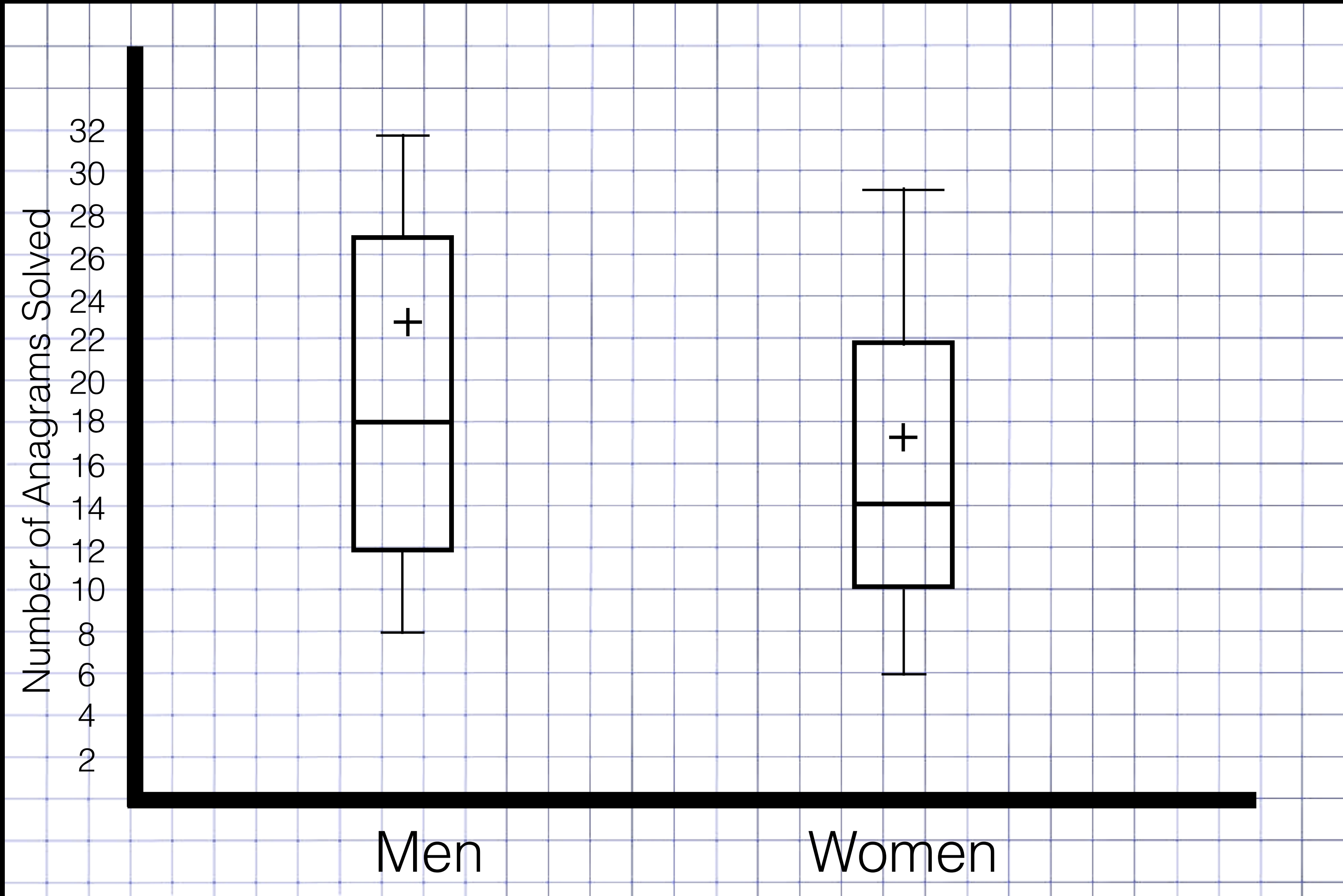
Drawing a Box Plot



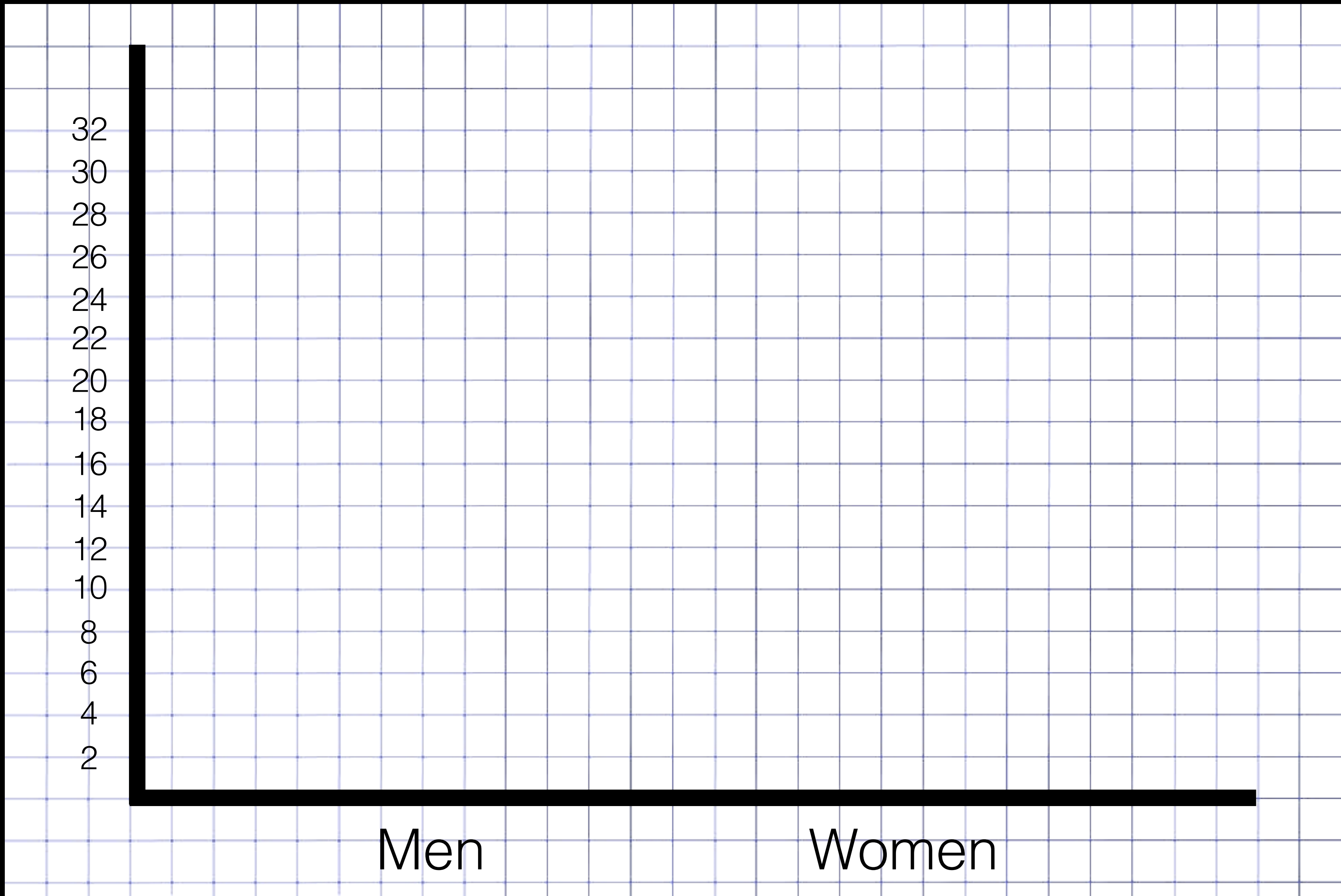
Drawing a Box Plot



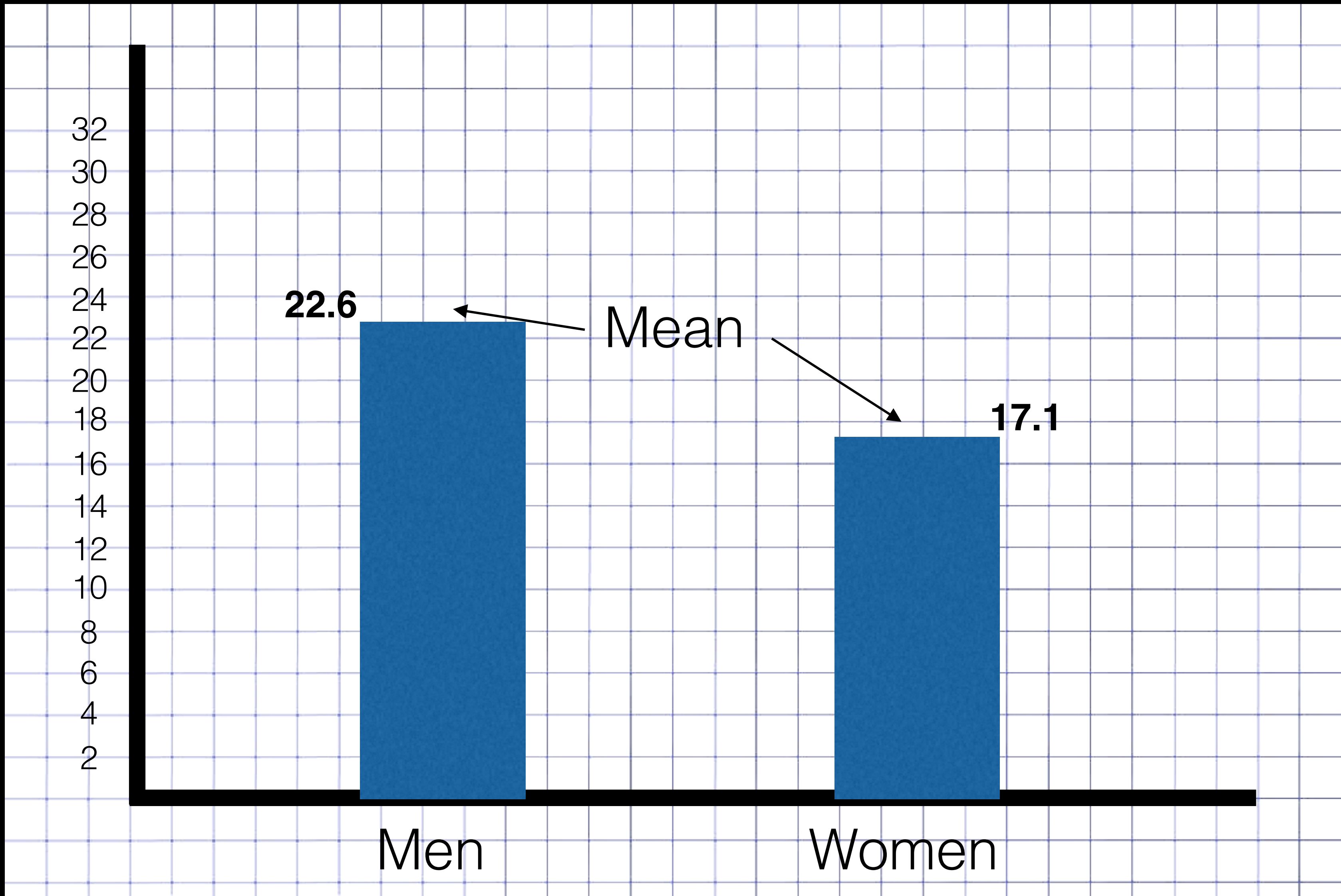
Drawing a Box Plot



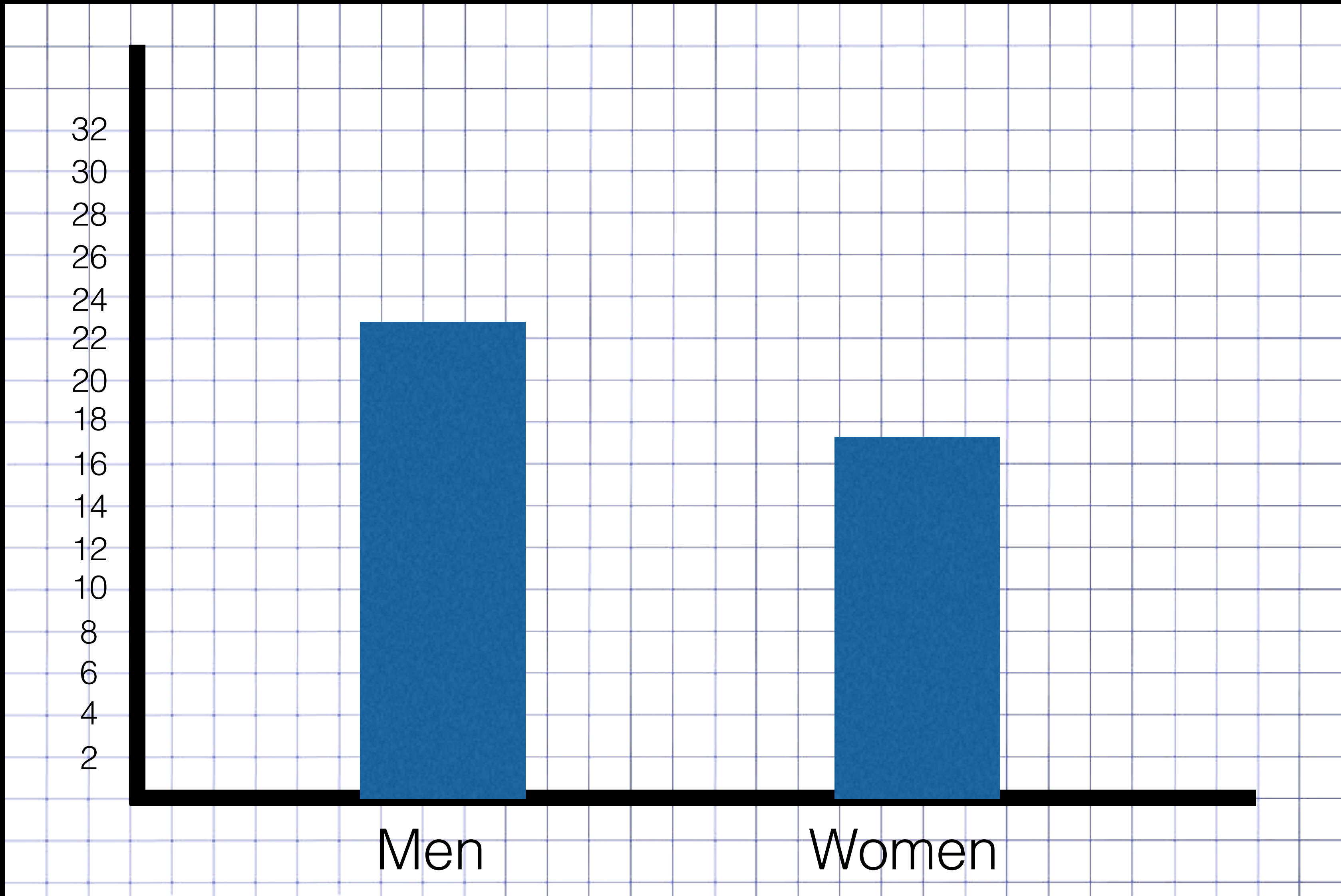
Drawing a Bar Chart (with Error Bars)



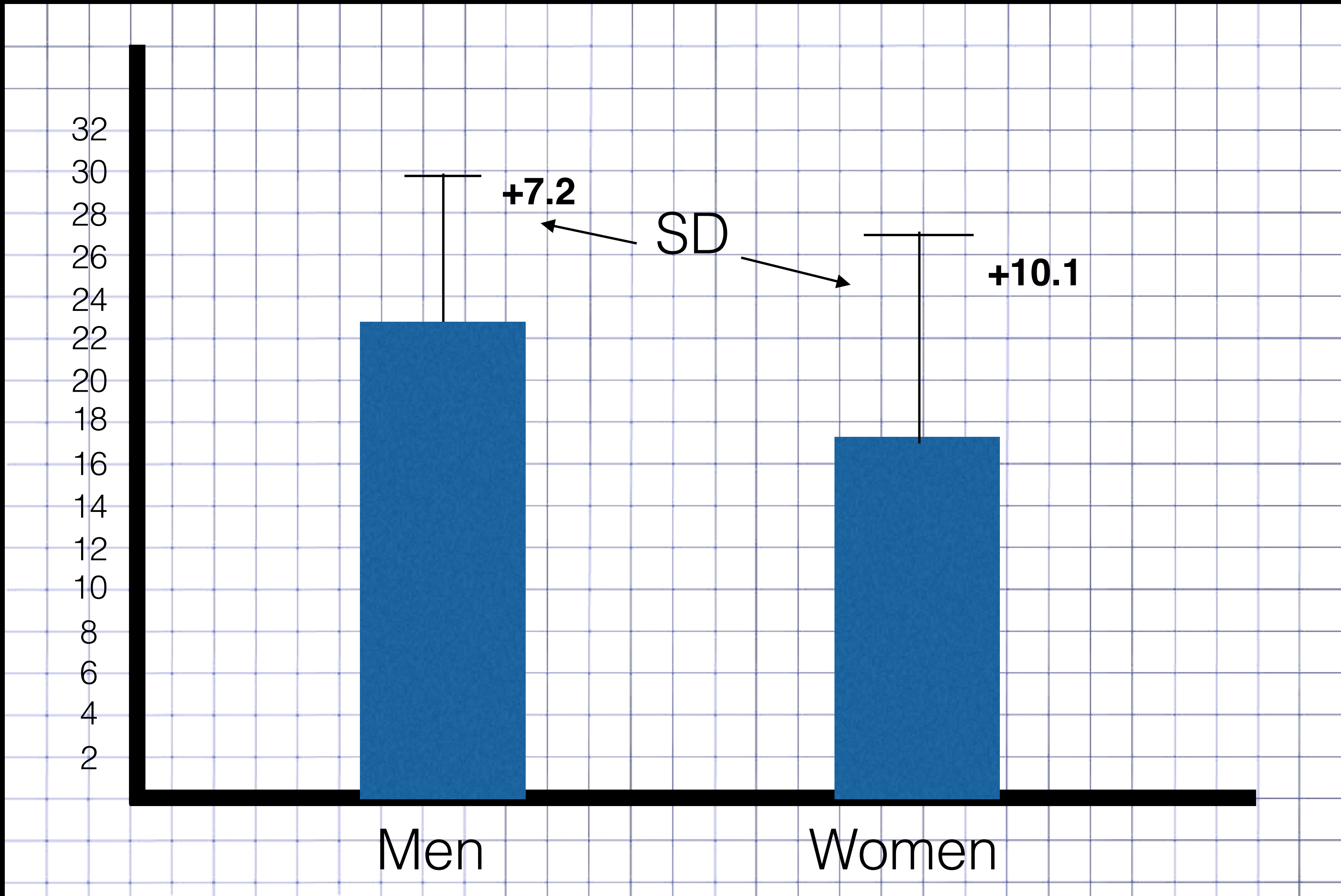
Drawing a Bar Chart (with Error Bars)



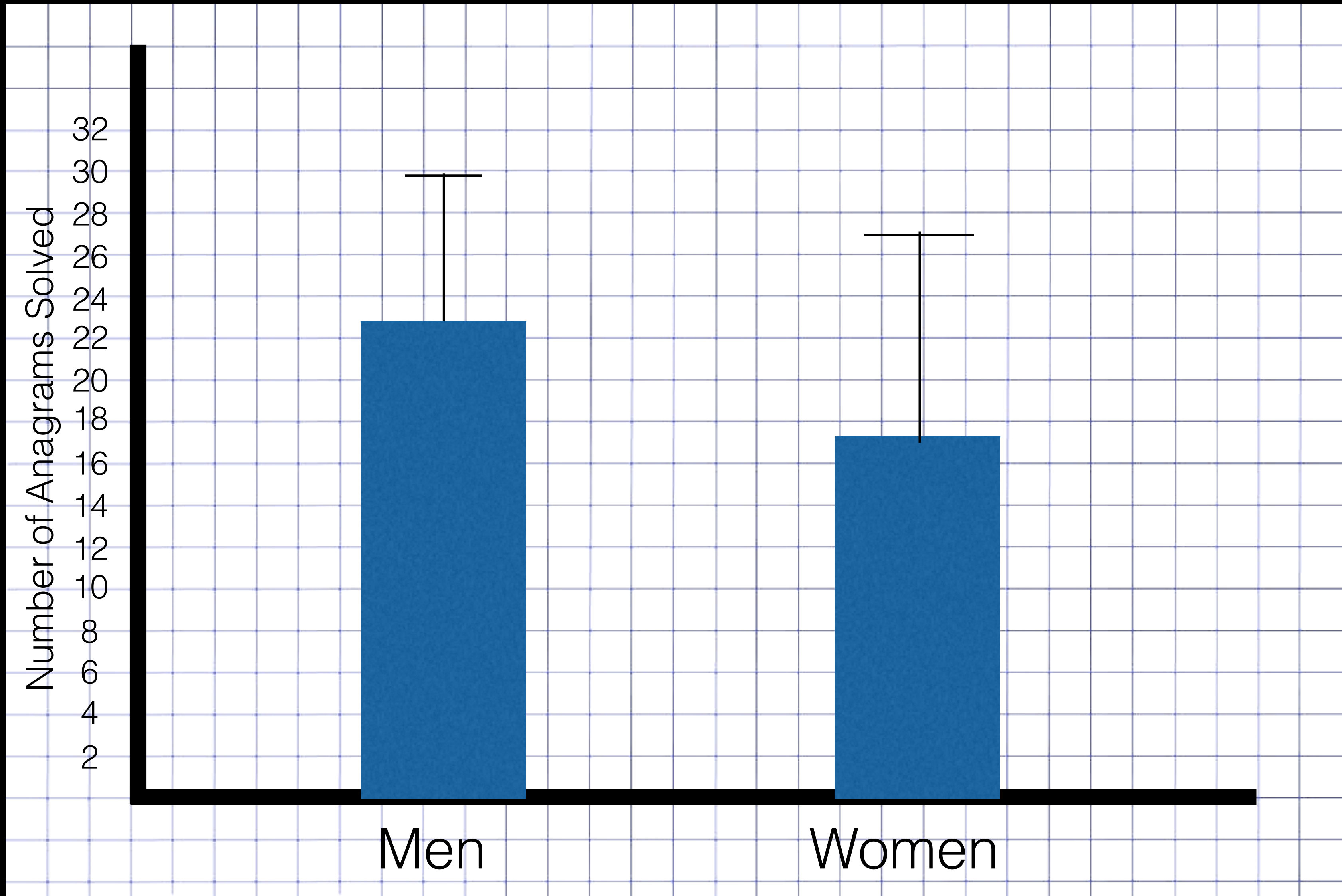
Drawing a Bar Chart (with Error Bars)



Drawing a Bar Chart (with Error Bars)



Drawing a Bar Chart (with Error Bars)

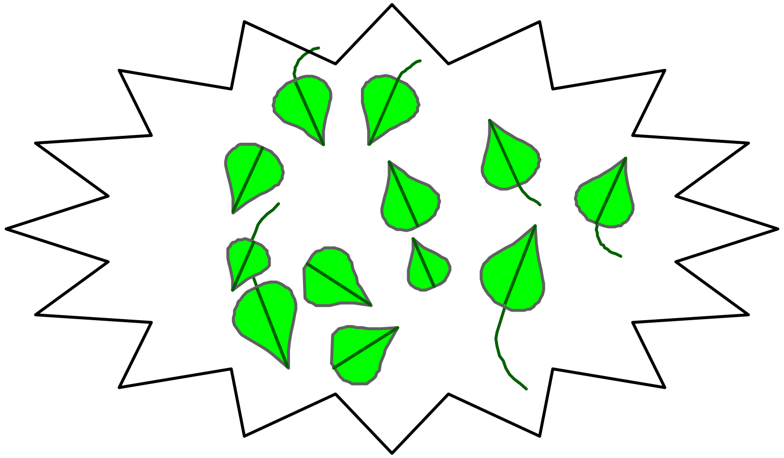


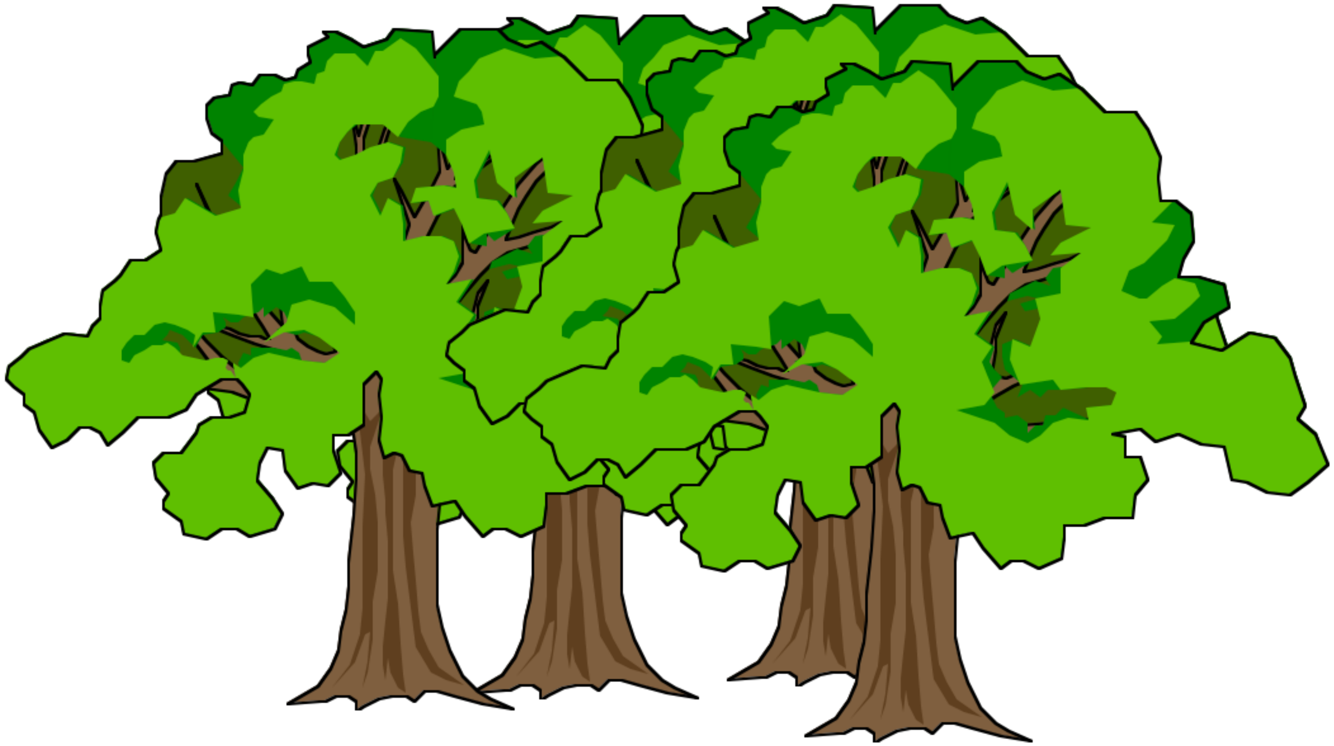
Hypothesis Testing and Sampling

- Hypothesis Testing relies on the concepts of *sampling*, *populations*, and *probability*.
- The basic logic: When we use a sample to make inferences about population parameters, we want to know whether our sample comes from a population that is best described by the null hypothesis.
- The null hypothesis is a statement which we usually are attempting to *reject* in our research.

Conceptually...

You have an “interesting” sample of leaves



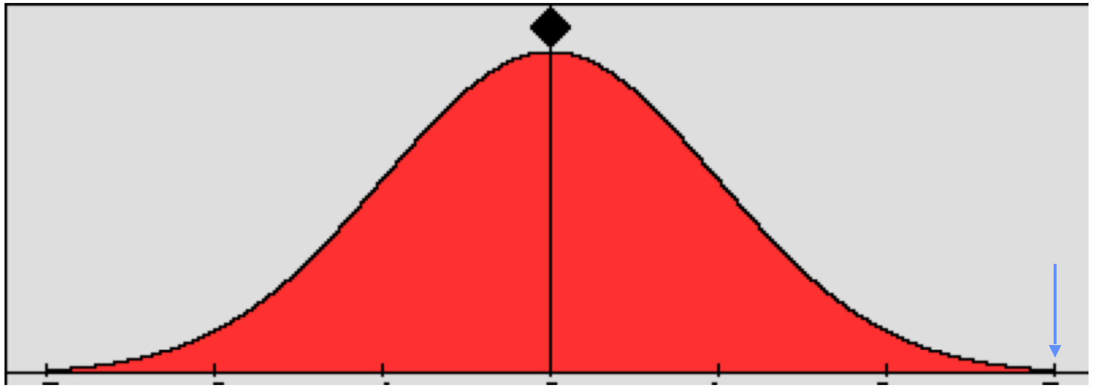


Ho: it comes from a standard tree

Alternative: it comes from this tree

An Example of Hypothesis Testing with Statistics

A group of Norwegian sailors has an average pulse of 98 beats per minute. Is this *significant* (that is, are they not healthy)?

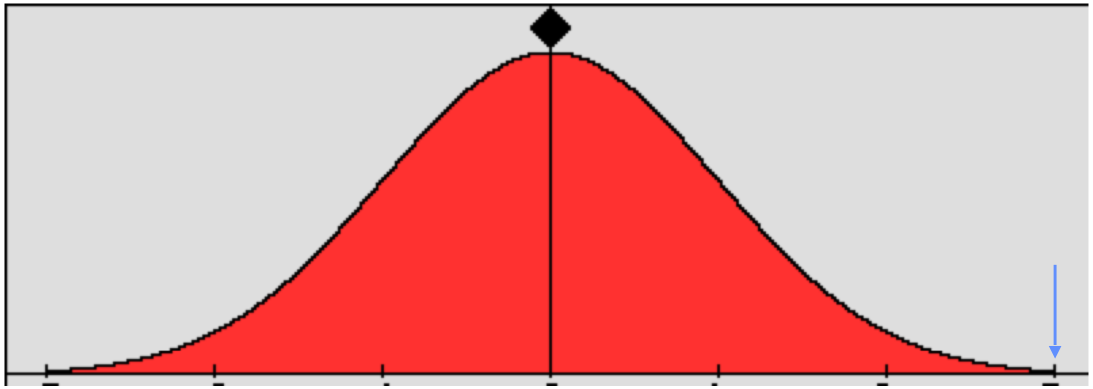


Null Hypothesis: μ=73

A "healthy" population has a mean pulse of 73 bpm, with an SD of 8. So, 98 beats per minute would be 3.125 standard deviations above the mean.

An Example of Hypothesis Testing with Statistics

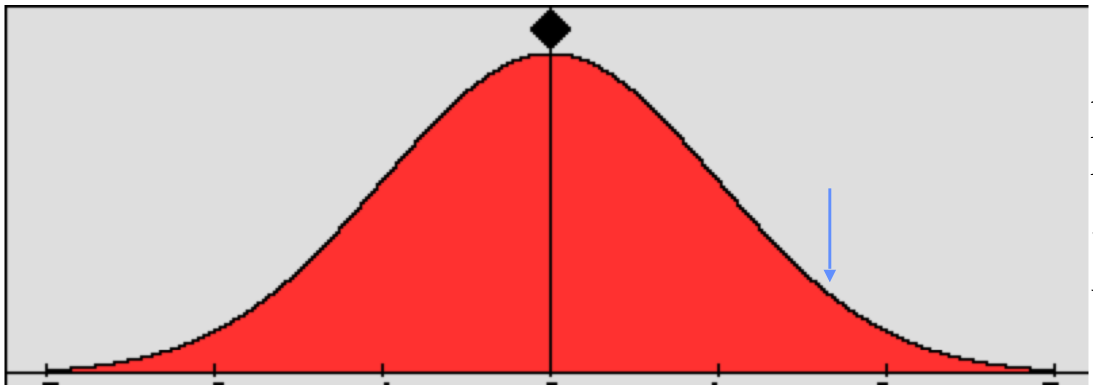
A group of Norwegian sailors has an average pulse of 98 beats per minute. Is this *significant* (that is, are they <u>not</u> healthy)?



Null Hypothesis: μ=73

A "healthy" population has a mean pulse of 73 bpm, with an SD of 8. So, 98 beats per minute would be 3.125 standard deviations above the mean.

--or--



Alternative Hypothesis: $\mu \neq 73$

An "unhealthy" population that has a higher mean pulse, with an SD of 8. So, 98 beats per minute would be closer to the mean.

Five Steps of Hypothesis Testing

1. Make some statements about the population parameters: a null hypothesis and an alternative (research) hypothesis.
2. Determine the population parameters assuming the null hypothesis is true.
3. Determine a “cut-off” point in the population at which the null hypothesis should be rejected.
4. Determine the probability of your sample statistic assuming the null hypothesis is true.
5. If that probability exceeds the “cut-off” point, reject the null hypothesis. If it doesn’t, retain it.

An Example: Hypothesizing about a Single Sample Score

- A certain depression questionnaire can have scores that range from 1 (*not at all depressed*) to 50 (*extremely depressed*). When given to a population of "normal" adults, μ=20 and $\sigma = 5$.
- Joe, an air traffic controller, completes the questionnaire, and gets a score of 42. Is this "normal?" Should he be considered at possible risk for depression?

Step #1. State the Null Hypothesis and Research Hypothesis

Example: Joe's depression score is 42 on a scale that is 1-50.

Ho: Joe comes from a population with $\mu=20$.

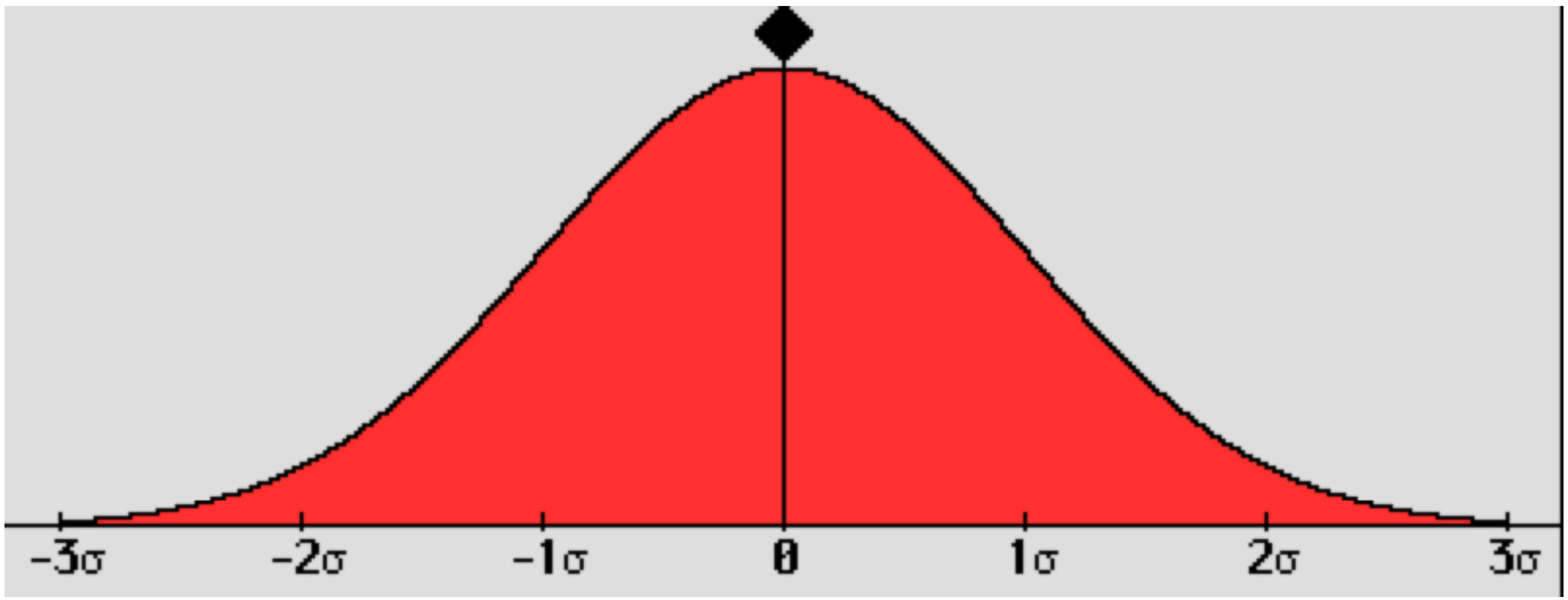
Ha: Joe comes from a population with $\mu\neq20$.

- NOTICE: the alternative (research) hypothesis is the logical complement to the null hypothesis (i.e., it should cover all instances not covered by Ho).

Step #2. Determine the Population's parameters

Example: Joe's depression score is 42 on a scale that is 1-50.

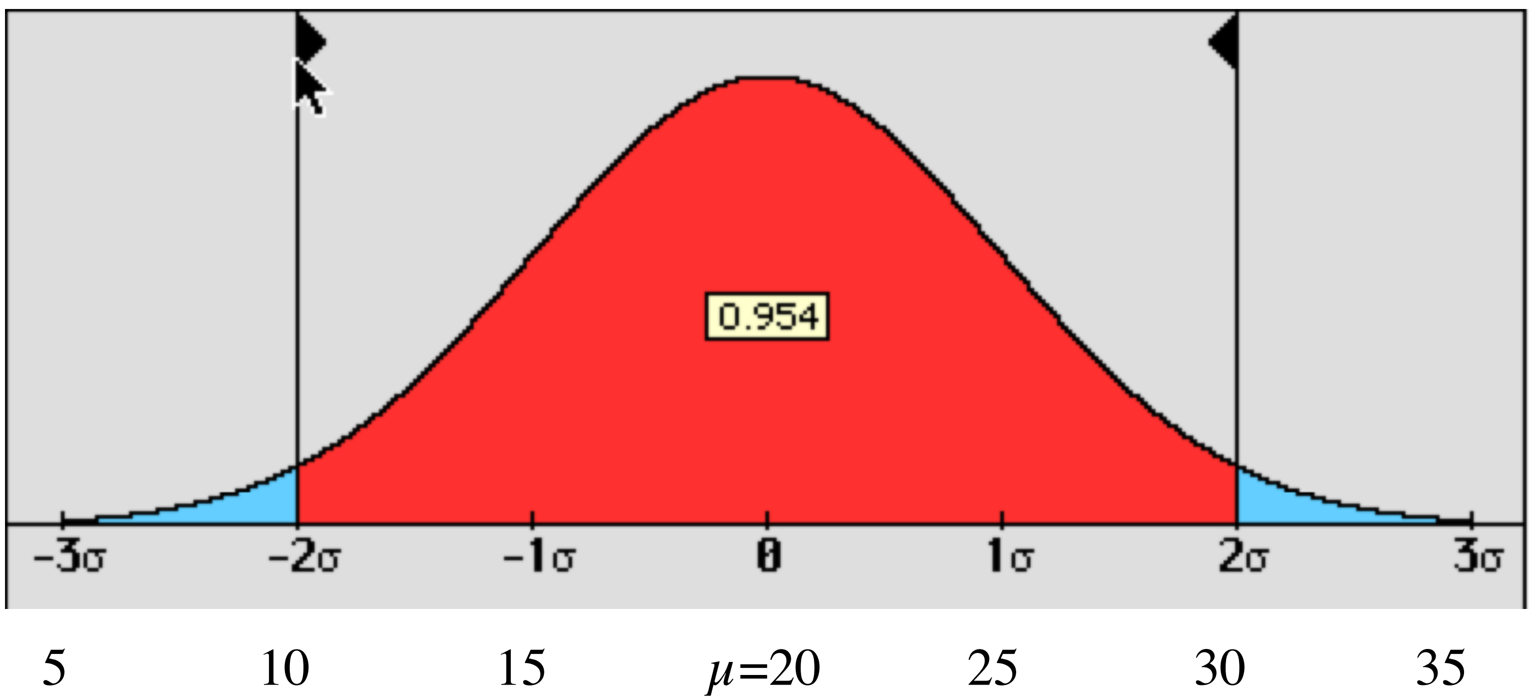
- If Ho is true, $\mu = 20$, and $\sigma = 5$



5 10 15 μ=20 25 30 35

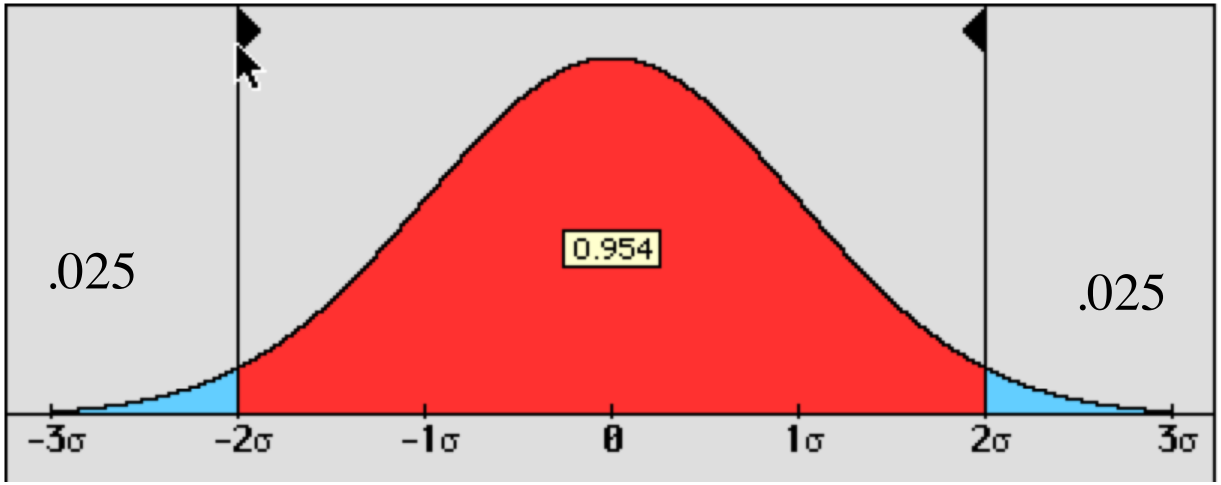
Step #3. Determine a "cut-off" statistic at which to reject Ho.

- We need to choose a "level of risk." Typically, researchers use p = .05 as the cut-off level.
- We then need to find the cut-off point, where a given parameter occurs less than .05 of the time, assuming Ho is true.



Step #3 (cont.) A Slight Detour: One-tailed or Two-tailed?

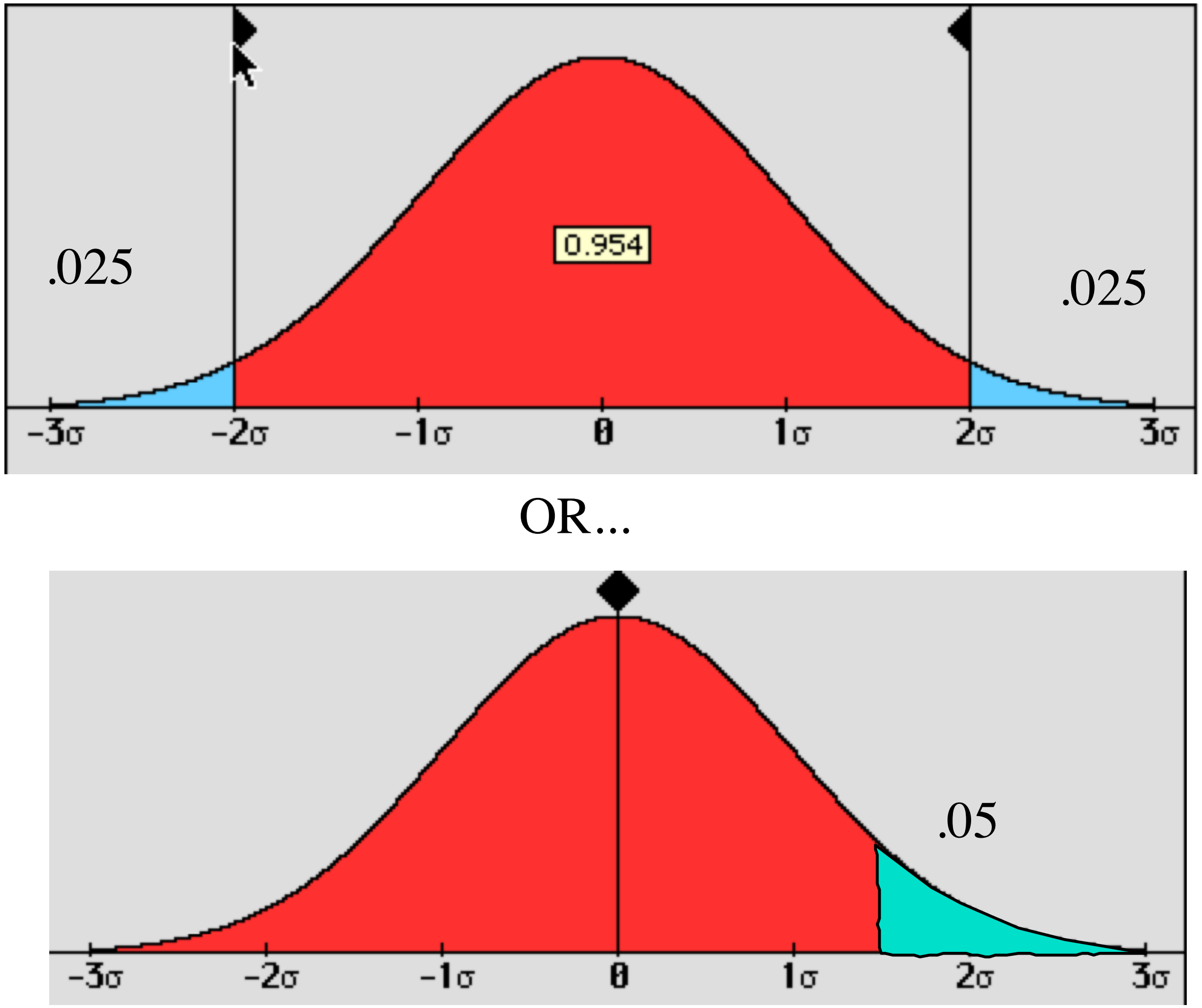
- We have the option of how to "divvy up" the .05 area:



"Two-tailed"

Step #3 (cont.) A Slight Detour: One-tailed or Two-tailed?

- We have the option of how to "divvy up" the .05 area:

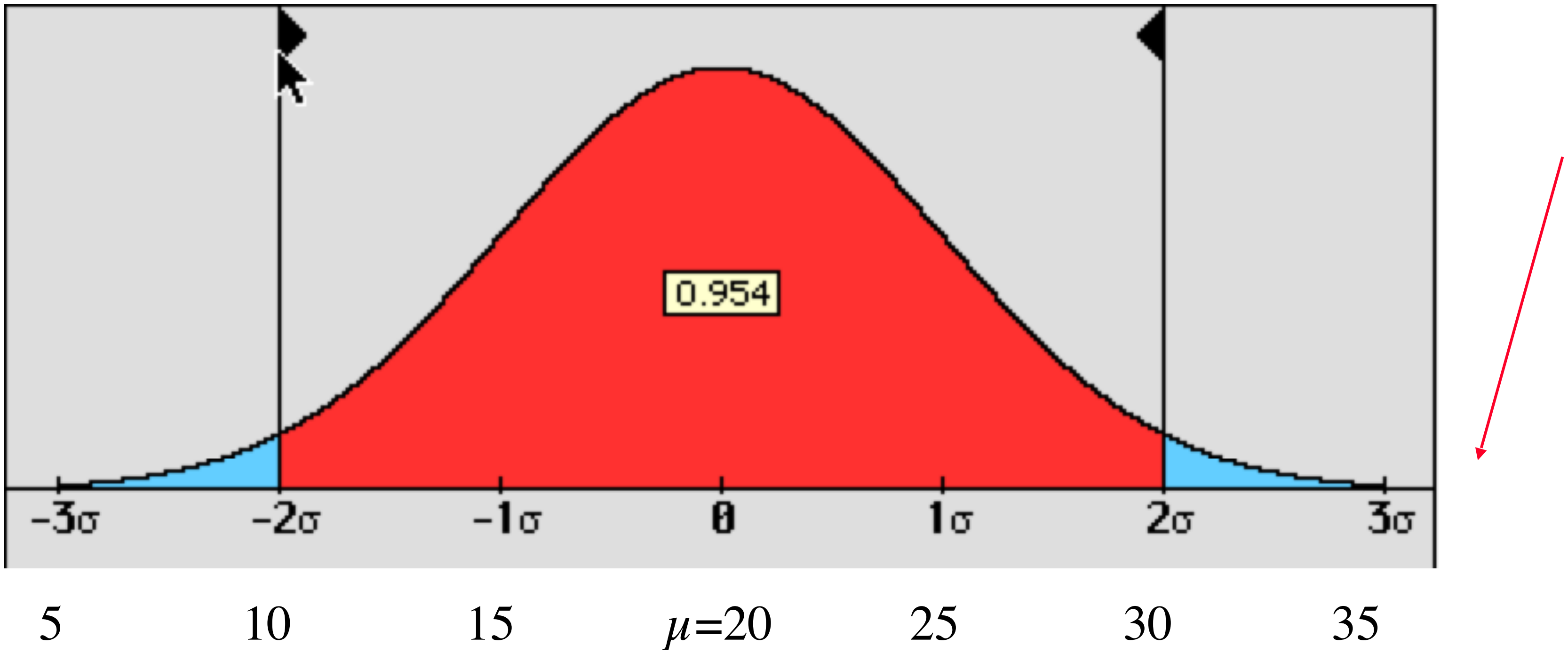


"One-tailed"

Step #4. Determining the probability of your sample score if Ho is true.

Example: Joe's depression score is 42 on a scale that is 1-50.

- If Ho is true, $\mu=20$, $\sigma=5$.
- Joe's score is z= (X-Mean)/SD = (42-20)/5 = 4.4



Step #5. Reject or Retain Ho, depending on the probability of your sample statistic

- Because Joe's Z-score exceeds the "cut-off," reject Ho, and adopt the alternative hypothesis.
- Joe therefore comes from a population where $\mu \neq 20$.
- We can say "Joe's score significantly differs from the normal population of depression scores, $p < .05$"

Choosing between H_0 and H_1

In choosing between the two hypotheses, we assume H_0 to be true.

↓

We test statistically the probability of obtaining our observed results if H_0 is actually true

If the probability is low Reject H_0	If the probability is high Retain H_0
↓	↓
Infer observed difference is genuine	Infer observed difference is due to chance

How Unlikely does our observation have to be?

- Conventionally
  - $p<0.05$
  - 5%
  - 1 in 20
- More conservative criteria
  - $p<0.01$
  - 1%
  - 1 in 100

Now, More Generally Speaking...

- What about any errors that we could make “just by chance?”

Errors in Hypothesis Testing

The Researcher concludes…

But reality is…

	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Errors in Hypothesis Testing

The Researcher concludes…

But reality is…

	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Type I Error

- The probability of rejecting Ho when Ho is true.
- Known as alpha (α)
- Same as significance level!
- When we say p<.05, we're stating that the Type I error is less than 5%.

Errors in Hypothesis Testing

The Researcher concludes…

But reality is…

	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Type II Error

- The probability of retaining Ho when Ho is false.
- Known as Beta or β

The Relationship of Type I and Type II Errors

- Generally, as you decrease α, you increase β
- Generally, as you increase α, you decrease β
- The choice of an alpha of .05 or .01 is based on a compromise of the two types of Error

Not Making an Error

The Researcher concludes…

But reality is…

	Reject	Retain
Reject	GOOD! (1-β)	Type II Error (β)
Retain	Type I Error (α)	GOOD

Power

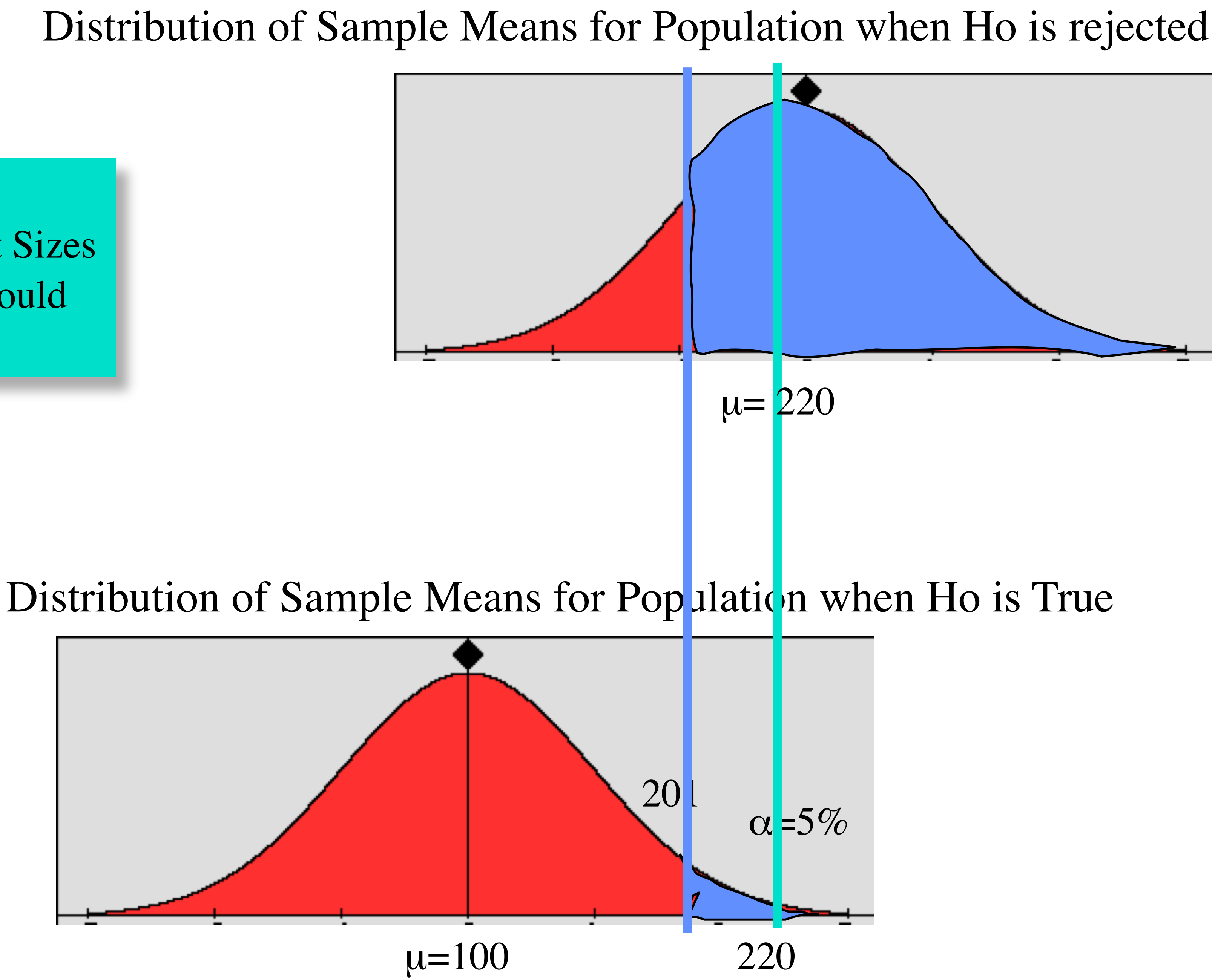
- Power is the probability that a researcher will (correctly) reject Ho when Ho is false

How do we maximize Power?

- Increase the effect size
- Increase the sample size
- Increase alpha
- Use one-tailed tests

A “Visual” Look at Increasing Power
Way #1

Increase Power
By Getting big Effect Sizes
(e.g., our pop mean could
be much higher)

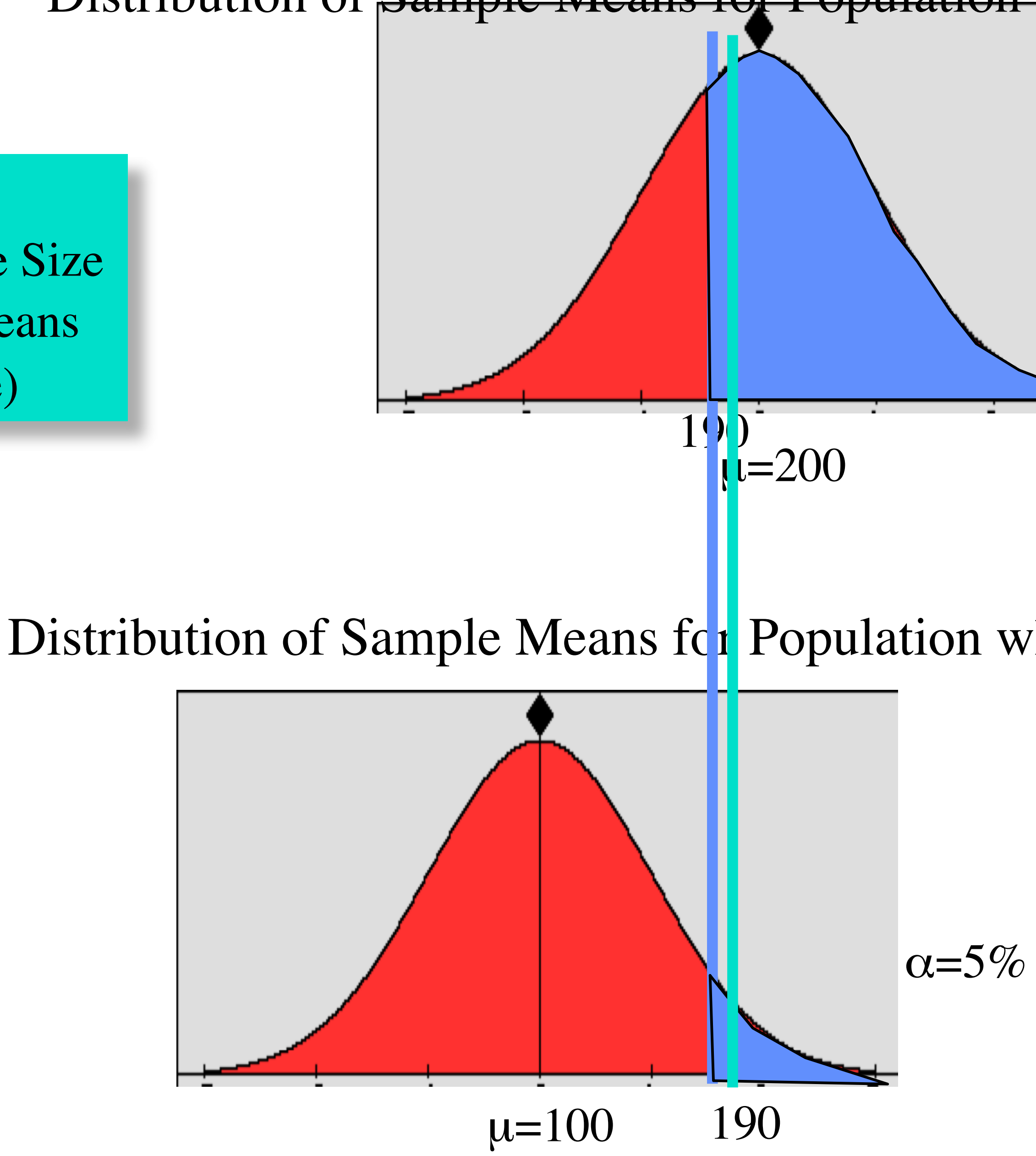


A "Visual" Look at Increasing Power

Way #2

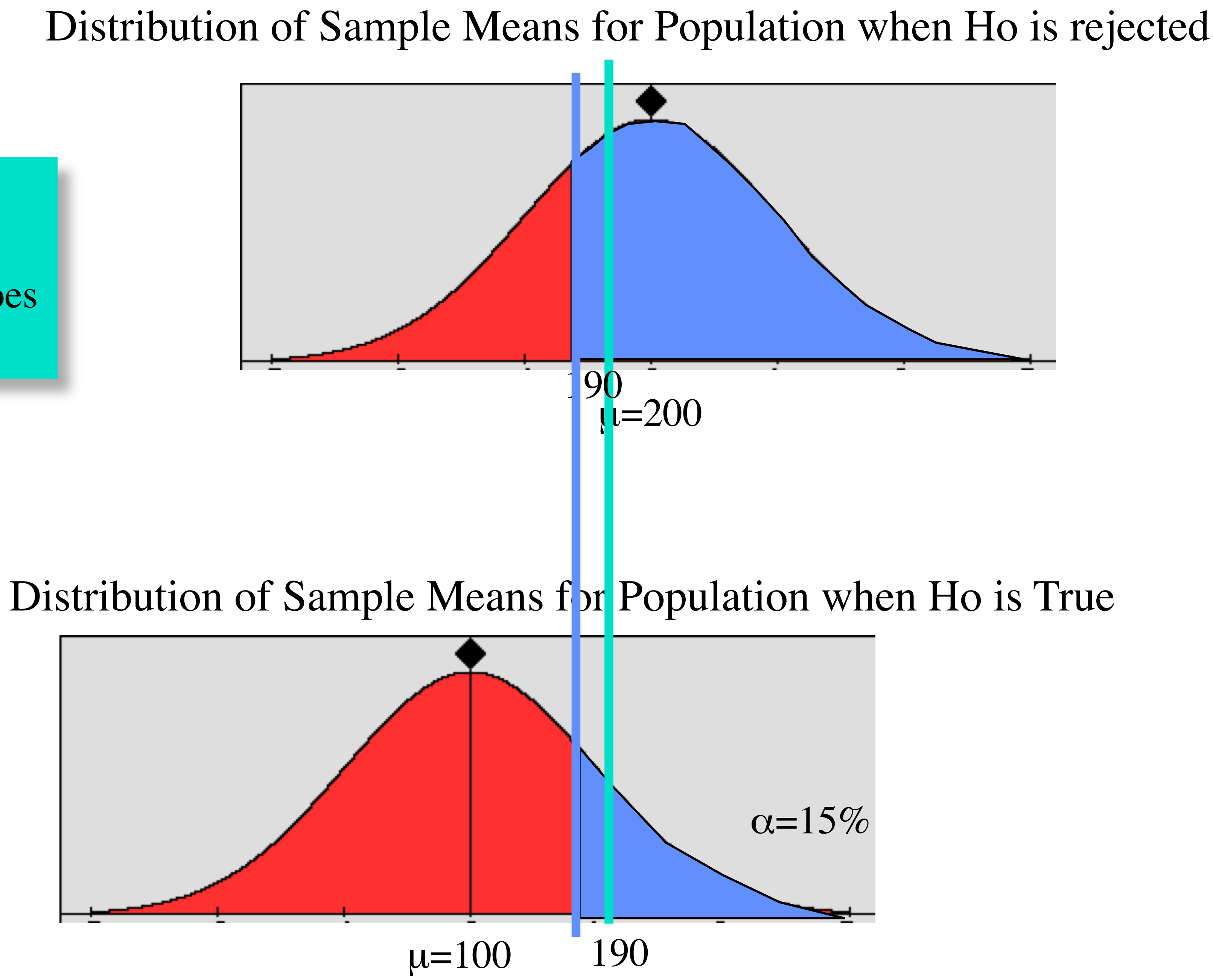
Increase Power
By Increasing Sample Size
(the distribution of means
becomes less variable)

Distribution of Sample Means for Population when Ho is rejected



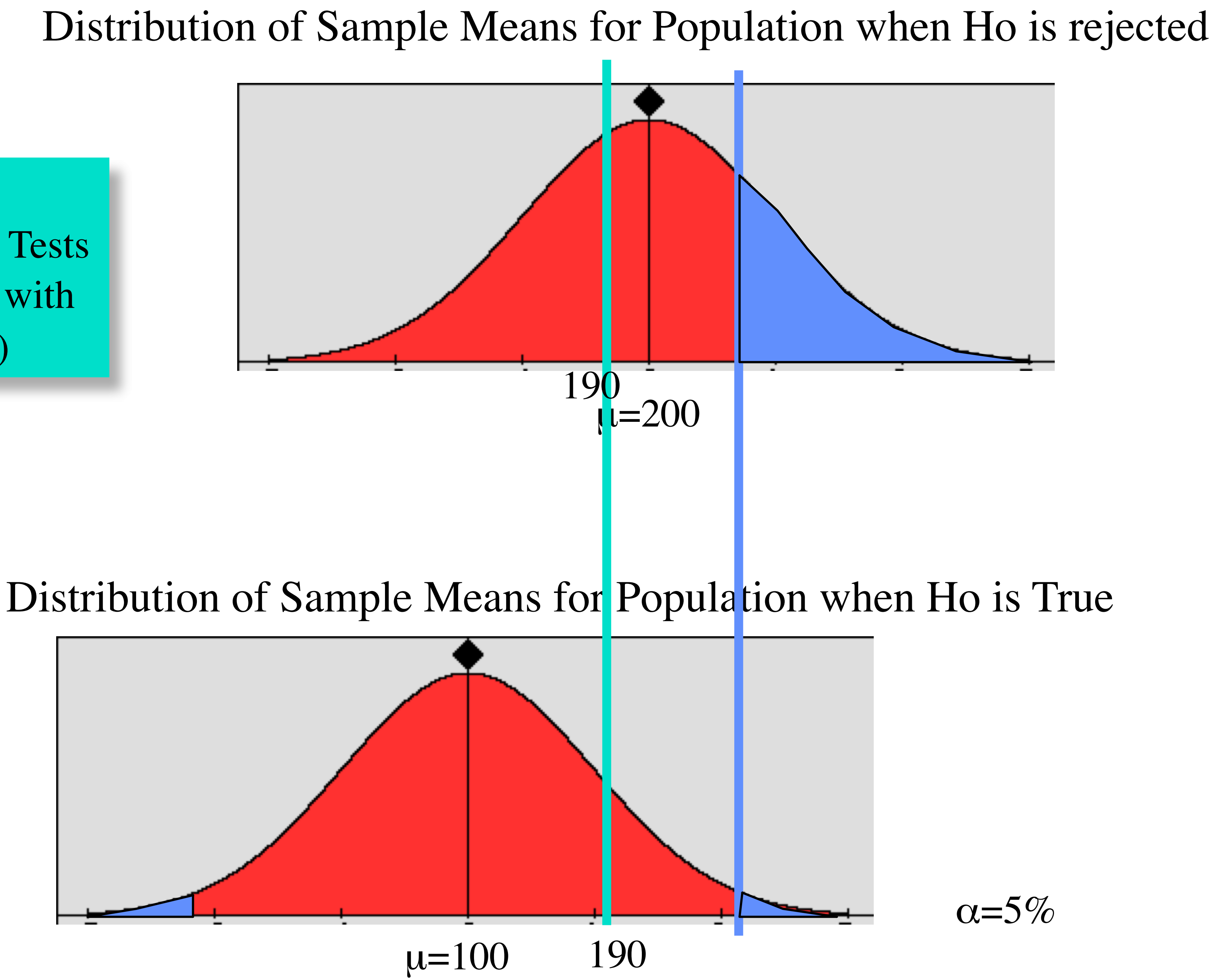
A "Visual" Look at Increasing Power
Way #3

Increase Power
By Increasing Alpha
(as α increases, so does power)



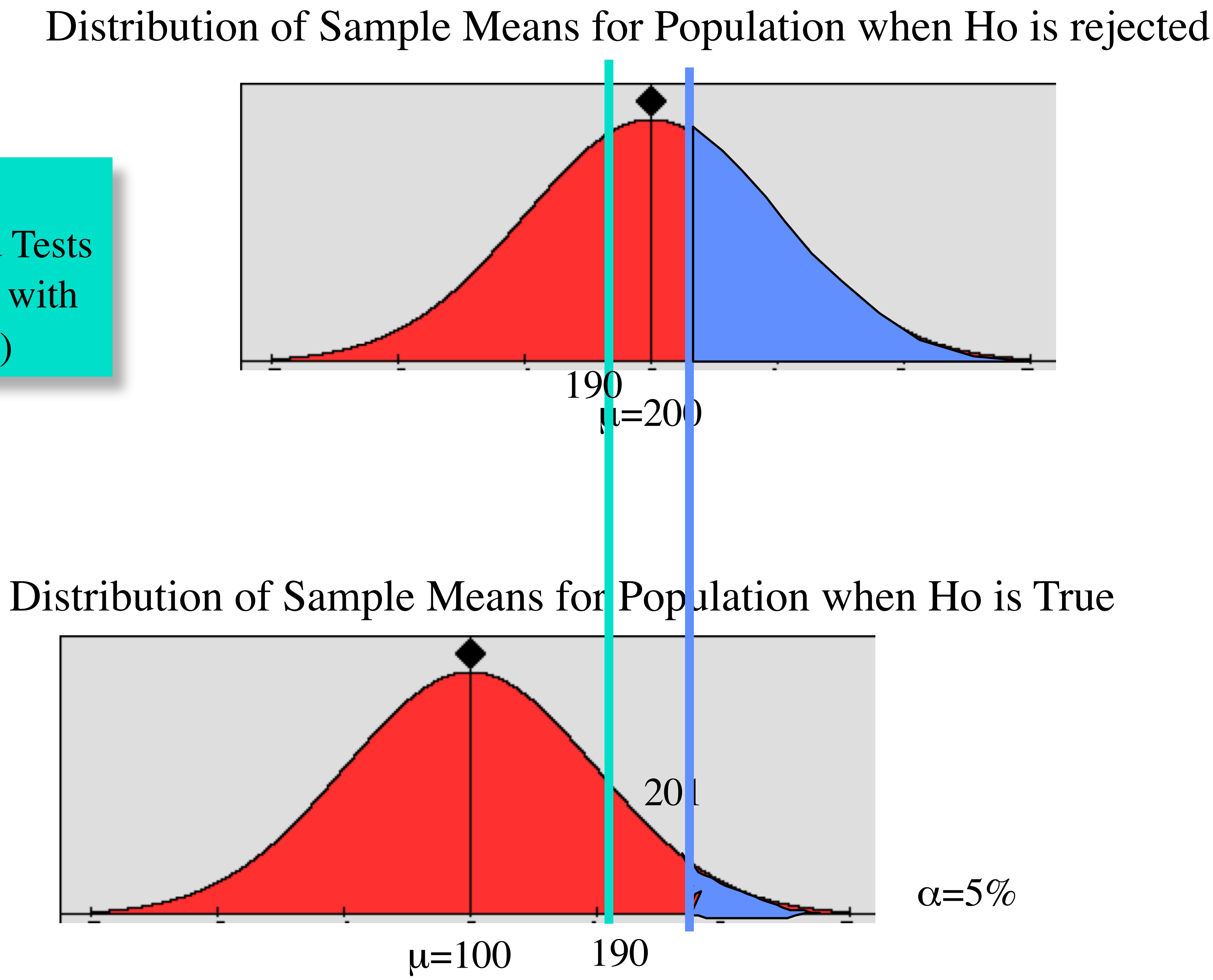
A “Visual” Look at Increasing Power
Way #4

Increase Power
By Using One-Tailed Tests
(i.e., we could get by with
a lower sample mean)



A "Visual" Look at Increasing Power
Way #4

Increase Power
By Using One-Tailed Tests
(i.e., we could get by with
a lower sample mean)



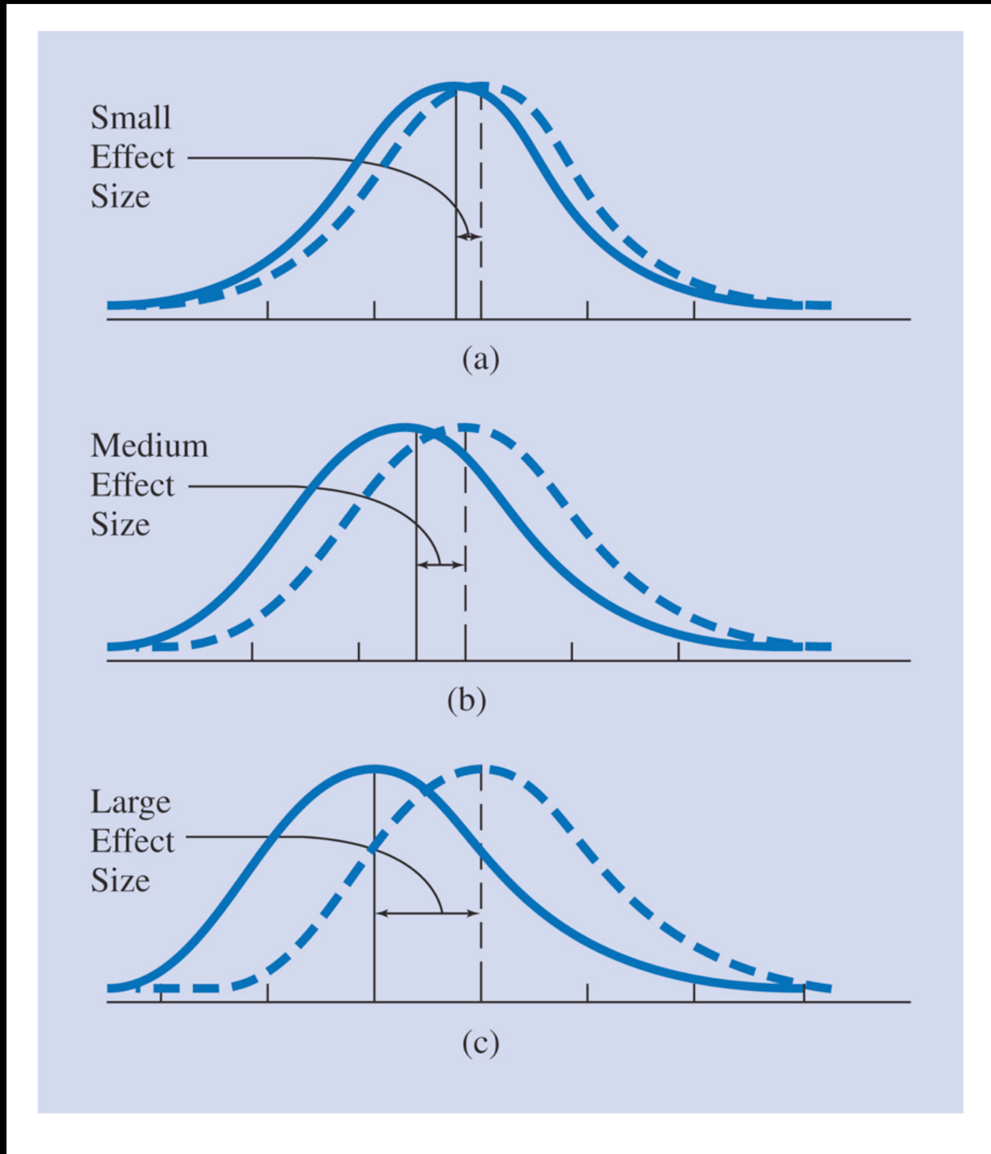
Four Pillars of Research Methodology

- Hypothesis Testing
- Sample Size
- Power
- Effect Size

Effect Size

- a standardised measure of difference (lack of overlap) between populations.
- A common effect size measures is *d*
- $d = (\mu_1 - \mu_2) / \sigma$
- It can be positive or negative

The Bigger the Effect Size, the Bigger the Difference Between Pop Means



Cohen's (1998) Guidelines

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Table 6-2 Summary of Cohen's Effect Size Conventions for Mean Differences

Verbal Description	Effect Size (d)
Small	.20
Medium	.50
Large	.80

Putting It All Together

- You’re planning to conduct an experiment with two groups: a treatment and a control group
- How many people should be in each group?

Researchers typically overestimate their study's power, and underestimate the sample size need to obtain .80 power with a small effect size.

Research Article

Psychological Science
2016, Vol. 27(8) 1069–1077

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797616647519
pss.sagepub.com

Researchers' Intuitions About Power in Psychological Research

Marjan Bakker[1], Chris H. J. Hartgerink[1], Jelte M. Wicherts[1], and Han L. J. van der Maas[2]

[1]Department of Methodology and Statistics, Tilburg School of Social and Behavioral Sciences, Tilburg University, and [2]Department of Psychology, Psychological Methods, University of Amsterdam

Abstract

Many psychology studies are statistically underpowered. In part, this may be because many researchers rely on intuition, rules of thumb, and prior practice (along with practical considerations) to determine the number of subjects to test. In Study 1, we surveyed 291 published research psychologists and found large discrepancies between their reports of their preferred amount of power and the actual power of their studies (calculated from their reported typical cell size, typical effect size, and acceptable alpha). Furthermore, in Study 2, 89% of the 214 respondents overestimated the power of specific research designs with a small expected effect size, and 95% underestimated the sample size needed to obtain .80 power for detecting a small effect. Neither researchers' experience nor their knowledge predicted the bias in their self-reported power intuitions. Because many respondents reported that they based their sample sizes on rules of thumb or common practice in the field, we recommend that researchers conduct and report formal power analyses for their studies.

How do researchers typically determine the sample size?

- According to Bakker et al. (2016), the answers are:
  - Practical constraints
  - Some rule of thumb (e.g., 20 per cell)
  - Based on common practice in their field
  - Want as many as possible

How many participants do you actually need for a two-condition study?

Table 2. Results From Study 2: Respondents' Estimates of the Required Sample Size and the True Required Sample Size to Reach a Power of .8

Required sample size	d = 0.20 (small ES)	d = 0.50 (medium ES)	d = 0.80 (large ES)
True	788	128	52
Estimated	216 [196, 236]	124 [114, 134]	77 [72, 83]

Note: The table presents the 20% trimmed means of the sample-size estimates, with 95% confidence intervals inside brackets. ES = effect size.

Bottom line: You should be aiming for at least 60 participants per condition, assuming a medium ES and power of .80

Eric Vanman

School of Psychology

University of Queensland

twitter: @socialneuro

UQ Social Neuroscience Lab: www.socialneuro.com