Sample size estimation

Wei Xuan

Ingham Institute
Applied Medical Research

Why do we need a sample size estimation

- To make sure the study have adequate power to show a difference to be significant
- Un-sufficient sample size not able to show an important difference
- Un-necessarily large sample size will waste resources
- Statistically significant difference may not have clinical importance

Factors affecting sample size estimation

- Magnitude of the difference (between the two comparison groups)
- Standard deviation of the outcome measures
- Pre-specified significant level of the statistical test (default value=0.05) and
- Pre-specified power of the statistical test (default value=80%)

Example 1: sample size estimation for comparison on two mean values

- To compare average height between boys and girls at 12-14 year's of age
- Mean height of boys=147cm
- Mean height of girls=152cm
- Standard Deviation (SD) = 8cm
- How many samples we need to show 5 cm difference to be significant

Example 1 (continue): sample size estimation for comparison on two mean values

- We need 41 in each group (82 in total)

Ingham Institute
Applied Medical Research

Sample size requirement in relation to magnitude of difference and SD

mean1 (cm)	mean2 (cm)	Difference of mean values (cm)	SD (cm)	Sample size required in each group (n1)
145	152	**7**	8	21
147	152	**5**	8	41
149	152	**3**	8	112

mean1 (cm)	mean2 (cm)	Difference of mean values (cm)	SD (cm)	Sample size required in each group (n1)
147	152	5	**5**	16
147	152	5	**8**	41
147	152	5	**10**	63

- The smaller the difference, the larger the sample size
- The larger the SD, the larger the sample size required

Ingham Institute
Applied Medical Research

Example 2: 6MWT

- A research group want to investigate whether hospital inpatients after knee replacement would have better outcome, measured by six minutes walking test (6MWT), compared to patients going home after knee replacement

- What information we need for a sample size calculation?

Example 2 (continue): 6MWT

- Mean distance 6MWT = 400 m (inpatient group)
- Mean distance 6MWT = 330 m (home group)
- SD=100 m
- Assume (by default): alpha=0.05, power=80%
- To show the difference between 400m and 330m to be statistically significant

- How many sample do we need?

Ingham Institute
Applied Medical Research

Online calculator for sample size

- https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html

Inference for Means: Comparing Two Independent Samples

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values for mu1 (mean of population 1), mu2 (mean c
for each sample). You may also modify α (type I error rate) and the power, if relevant. After making your entries, hit the

- (•) Calculate Sample Size (for specified Power)
- ( ) Calculate Power (for specified Sample Size)

Enter a value for mu1: 400

Enter a value for mu2: 330

Enter a value for sigma: 100

- ( ) 1 Sided Test
- (•) 2 Sided Test

Enter a value for α (default is .05): .05

Enter a value for desired power (default is .80): .80

The sample size (for each sample separately) is: 33

Calculate

- need 33 in each group (66 in total)

Ingham Institute
Applied Medical Research

Example 3: sample size estimation for a RCT study

- The research question is: whether a new drug to be more efficient as a treatment of hypertension compared to the current standard drug. The sitting diastolic blood pressure (SDBP) will be measured at baseline and then 3 month later. Change of SDBP will be used as primary outcome measure

Example 3 (continuous) : RCT study

- mean change of SDBP in new treatment group =18 mm Hg
- Mean change in the control group=14 mm Hg
- SD=9 mm Hg
- Alpha=0.05, power=80%
- 1:1 ratio between two randomisation groups

Using online calculator:

Inference for Means: Comparing Two Independent Samples

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values for mu1 (mean of population 1), mu2 (
for each sample). You may also modify α (type I error rate) and the power, if relevant. After making your entries,

- (●) Calculate Sample Size (for specified Power)
- ( ) Calculate Power (for specified Sample Size)

Enter a value for mu1: 18

Enter a value for mu2: 14

Enter a value for sigma: 9

- ( ) 1 Sided Test
- (●) 2 Sided Test

Enter a value for α (default is .05): .05

Enter a value for desired power (default is .80): .80

The sample size (for each sample separately) is: 80

Calculate

- The study will need to recruit 80 sample in each randomisation group (160 in total)

Statistical significant versus clinical important

- The difference between two (treatment) groups should have clinical meaning
- Clinical benefit should be based on findings from the literature or historical knowledge
- Minimal clinically important difference (MCID) https://en.wikipedia.org/wiki/Minimal_clinically_important_difference
- e.g. 0.5 point difference in QOL to be clinically meaningful
- e.g. 70 meter difference in 6MWT

Example 4: sample size calculation for comparing two proportions

- To compare the prevalence of wheeze in the last 12 month between 8-10 year's old boys and girls
- Boys tends to have higher prevalence of recent wheeze in this age range than girls
- To show a 24% of wheeze in boys to be significantly different to a 9% of wheeze in girls, how many samples (per group) do we need?

Ingham Institute
Applied Medical Research

Example 4 (continue): comparison of two proportions

- P_1 (in boys)=24%
- P_2 (in girls)=9%
- Assume significant level=0.05, power=80%
- How many sample do we need?

- Answer: need 93 in each group

Example 5: ‘translate’ original research question:

- To investigate that the sub-therapeutic levels occurs with higher frequency where there is a suspicion of LOR by biochemical or clinical parameters (and are more common than in those patients on maintenance therapy without suspicion of LOR)
- To do sample size estimation, we need to translate this research question into a hypothesis testing question

Ingham Institute
Applied Medical Research

Example5 (continue): comparing two proportions

- To compare the occurrence of sub-therapeutic levels to be higher in suspicion of LOR patients compared to no-suspicion of LOR patients
- p_1 =proportion of 'sub-level' = 15% in group 1
  p_2=proportion of 'sub-level' = 7% in group 2
- Assume suspicion of LOR and no-suspicion of LOR to be 50:50 in patients

Online calculator for sample size calculation

- https://www.stat.ubc.ca/~rollin/stats/ssize/b2.htmll

Inference for Proportions: Comparing Two Independent Samples

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values (as decimal fractions) for p1 (proportion in p
sample). You may also modify α (type I error rate) and the power, if relevant. After making your entries, hit the **calcula**

- ◉ Calculate Sample Size (for specified Power)
- ○ Calculate Power (for specified Sample Size)

Enter a value for p1: 0.15

Enter a value for p2: 0.07

- ○ 1 Sided Test
- ◉ 2 Sided Test

Enter a value for α (default is .05): .05

Enter a value for desired power (default is .80): .80

The sample size (for each sample separately) is: 239

Calculate

- need 239 in each group (478 in total)

Ingham Institute
Applied Medical Research

Where to find information on mean difference and SD?

- From literature, other studies with the same measurements
- Studies with the same outcome measure but with slightly different study population
- Conducting a pilot study

What if SD not known

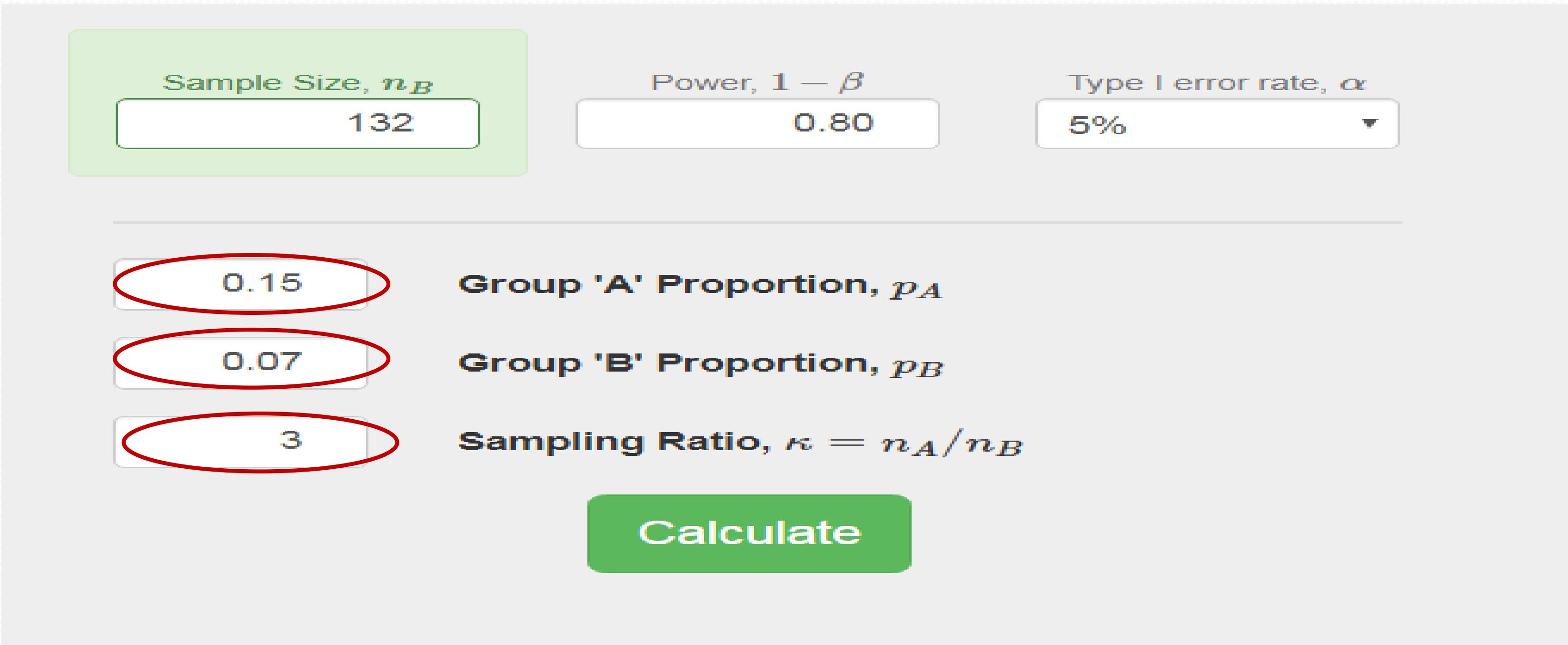
- 3 SD on either side of the mean value will cover 99.7% of the sample
- SD ~ (max – min) / 6 = range / 6

Two group comparison with un-equal sample per group

- Look at example 5 again (suspicion of LOR example)
- If with 1:1 ratio, we need 239 in each group (478 in total)
- What if there are 75% who would have suspicion of LOR (25% non-suspicion)
- Sample size calculation assuming 3:1 ratio

Sample size calculation for 3:1 ratio

http://powerandsamplesize.com/Calculators/Compare-2-Proportions/2-Sample-Equality



- We need 132 in non-LOR group, and 3x132=396 in LOR group, i.e.528 in total for 3:1 ratio

What if there are drop outs in follow up studies

- Conducting a sample size calculation (assume no drop out first)
- Inflate the sample size by the possible drop out rate (say 15%)
- For a 15% drop out, multiply sample size by 1.15 approximate

Ingham Institute Applied Medical Research

Online calculator for sample size calculation

- http://powerandsamplesize.com/Knowledge/
- http://powerandsamplesize.com/Calculators/
- https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html

Thank you

Introduction to Statistical Analysis
Basic concepts and hypothesis testing

Joseph Descallar
Biostatistician
Ingham Institute

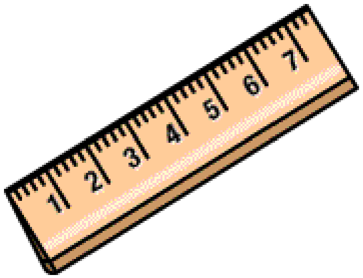
Outline

- **Types of Data**
- **Descriptive Statistics**
  - Numerical summaries
  - Graphical Summaries
- **Hypothesis Testing**
  - Null and alternative hypothesis

TYPES OF DATA

Continuous data

Data that can take on **any** value (within a range)



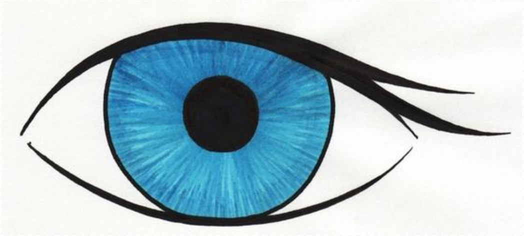
Also referred to as **Scale Data**

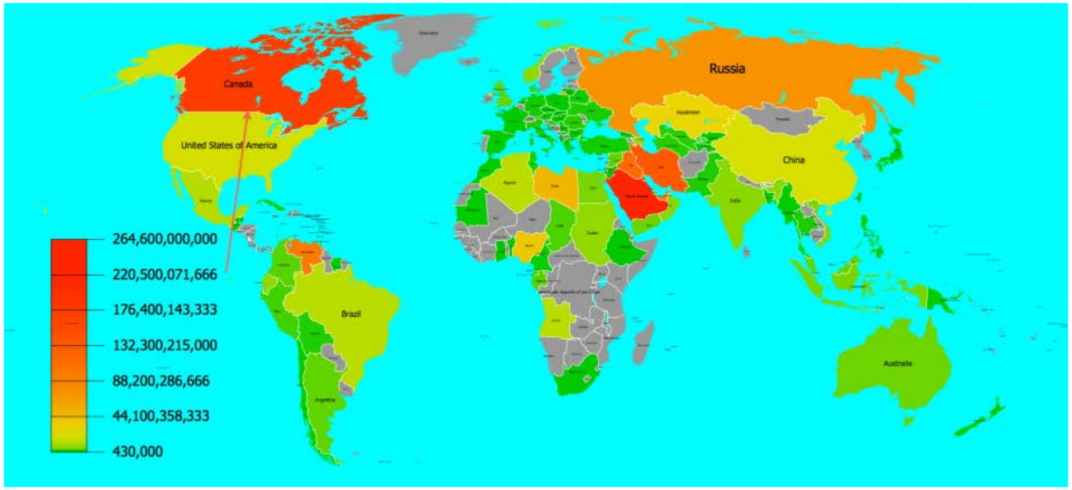
Binary data

Categorical Data
with **2** categories

Nominal data

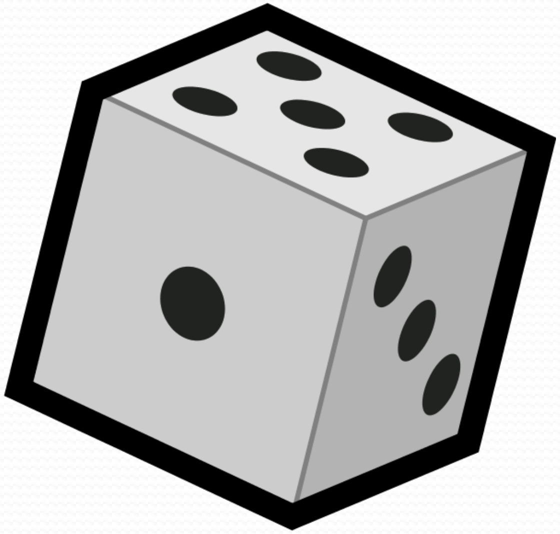
Categorical Data with **more than 2** categories





Ordinal data

Categorical Data with **ordered** categories



DESCRIPTIVE STATISTICS

Continuous data summaries

Measures of Central Tendency

- Mean (Average)
- Median (50th Percentile)

Measures of Spread

- Standard Deviation
- Lower and upper quartiles (25th and 75th Percentile respectively)
- Interquartile Range (75th Percentile – 25th Percentile)
- Range (Maximum - Minimum)

Standard Deviation (SD)

- Useful for normally distributed data (more on this later)
- Larger standard deviation indicates more variability

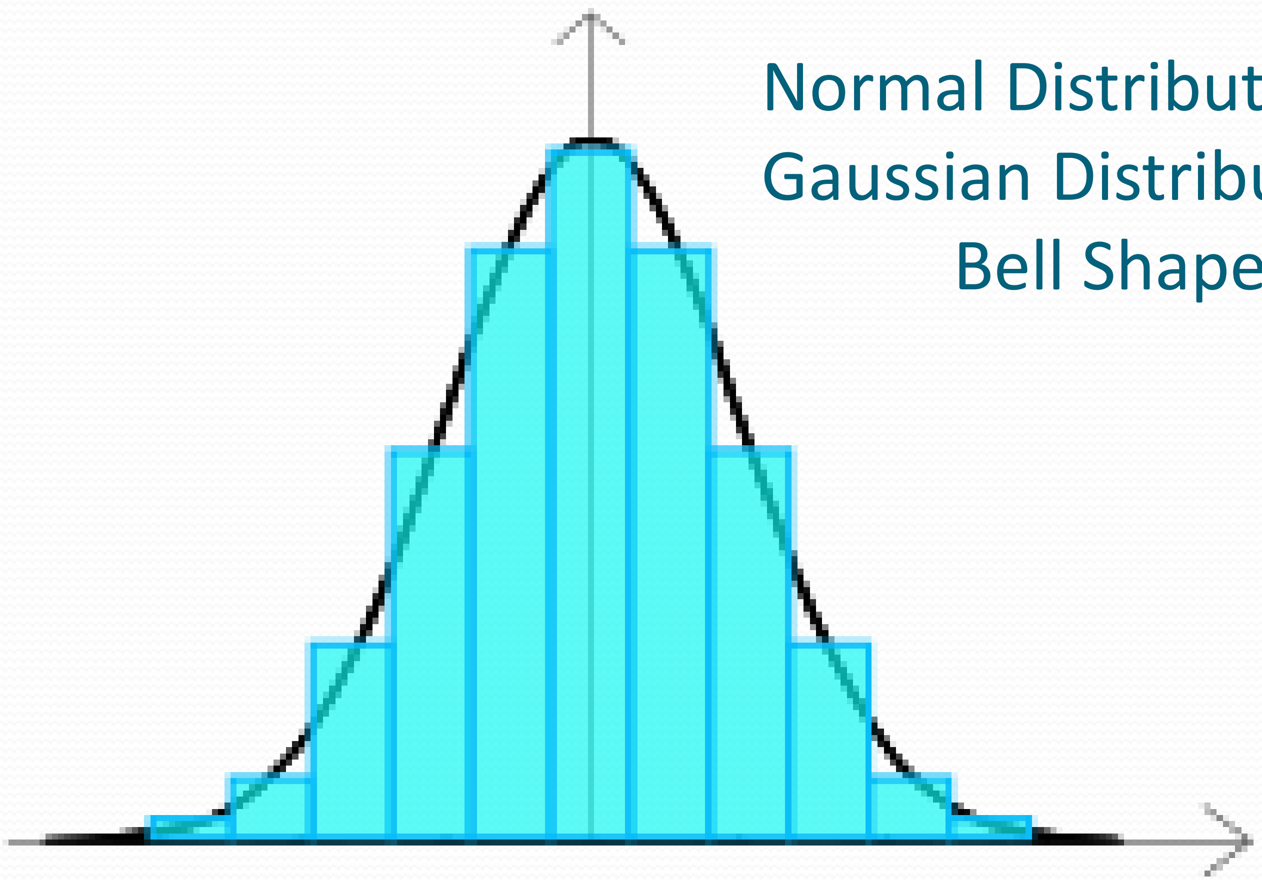
Gender	N	Mean Height	Standard Deviation
Females	150	145 cm	**10 cm**
Males	150	150 cm	**5 cm**

≈ 68% of your data fall within 1 standard deviation

≈ 95% of your data fall within 2 standard deviations

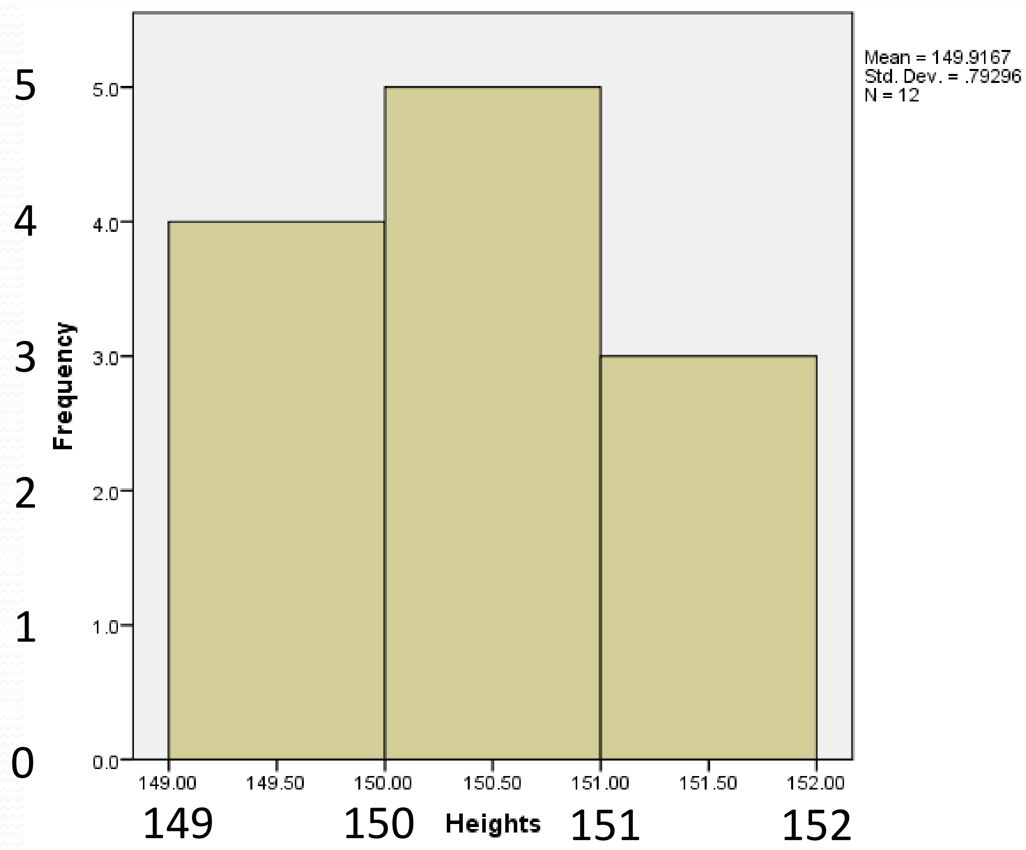
≈ 99.7% of your data fall within 3 standard deviations

Histogram



Normal Distribution
Gaussian Distribution
Bell Shape curve

Histogram

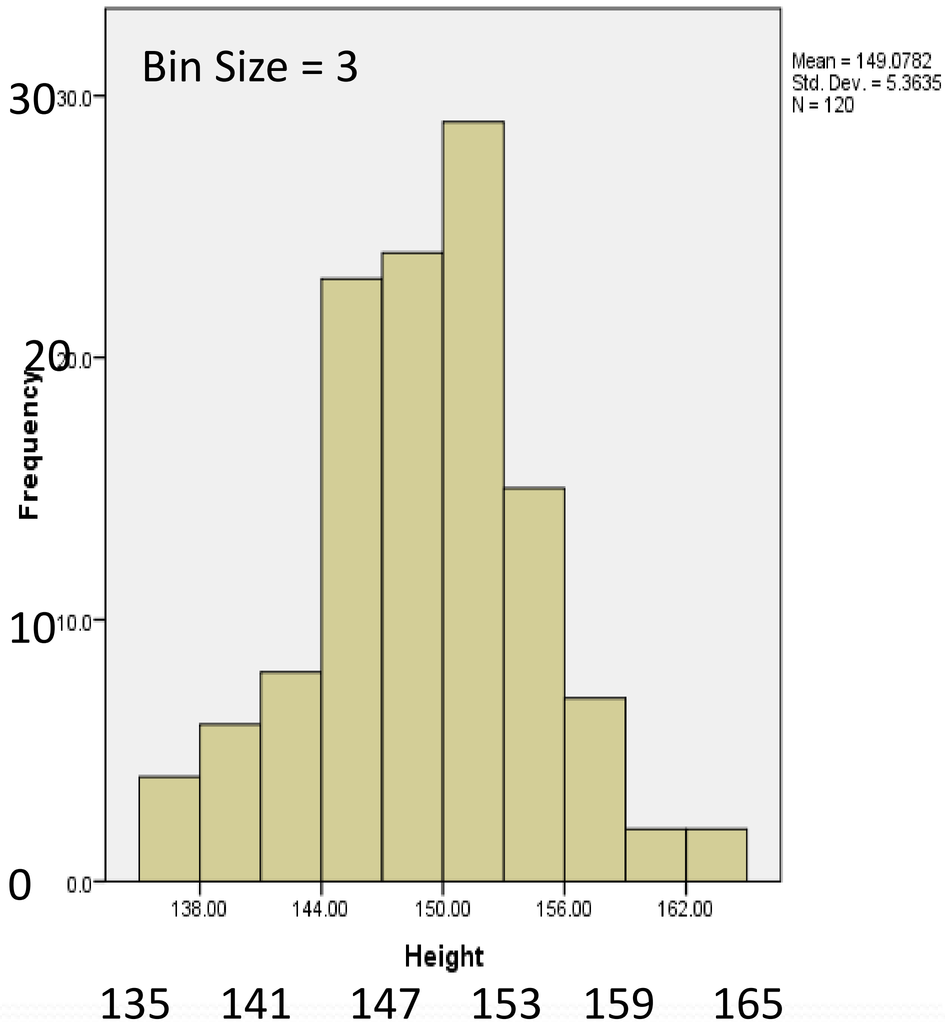


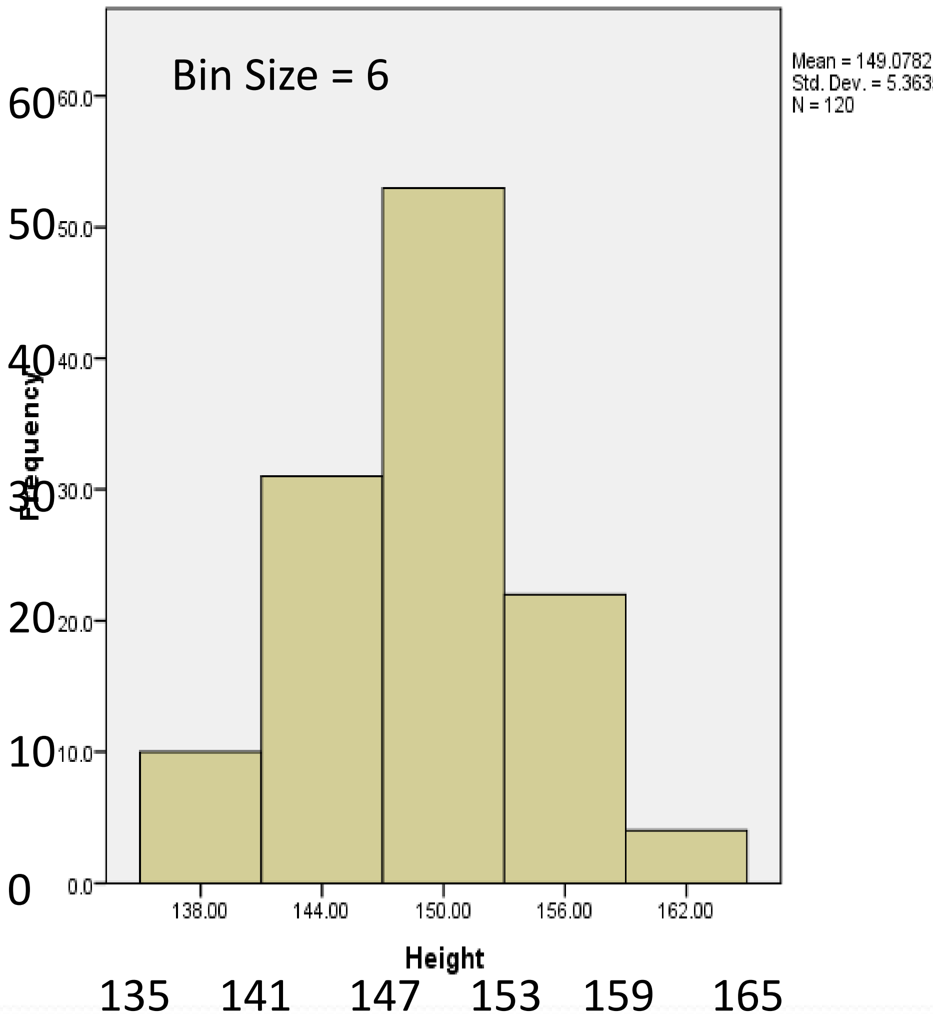
4 people with a height between 149 and 150 cm’s

5 people with height between 150 and 151 cm’s

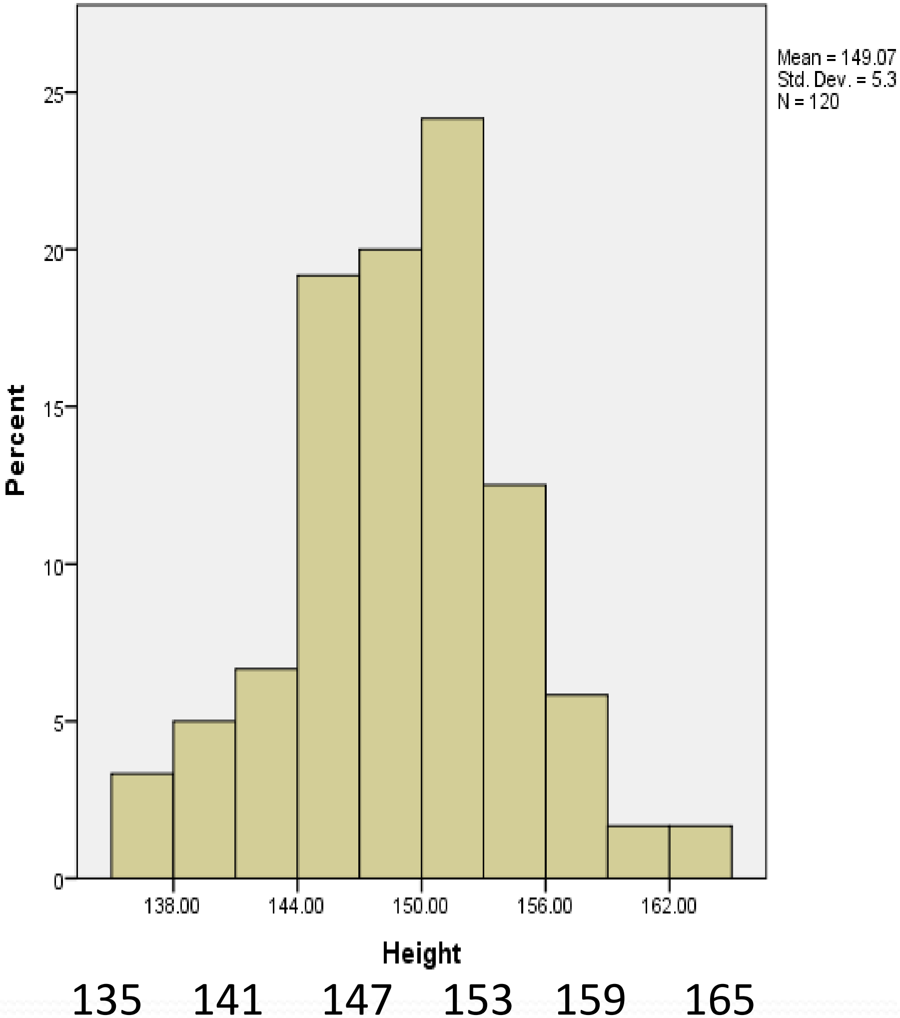
3 people with height between 151 and 152 cm’s

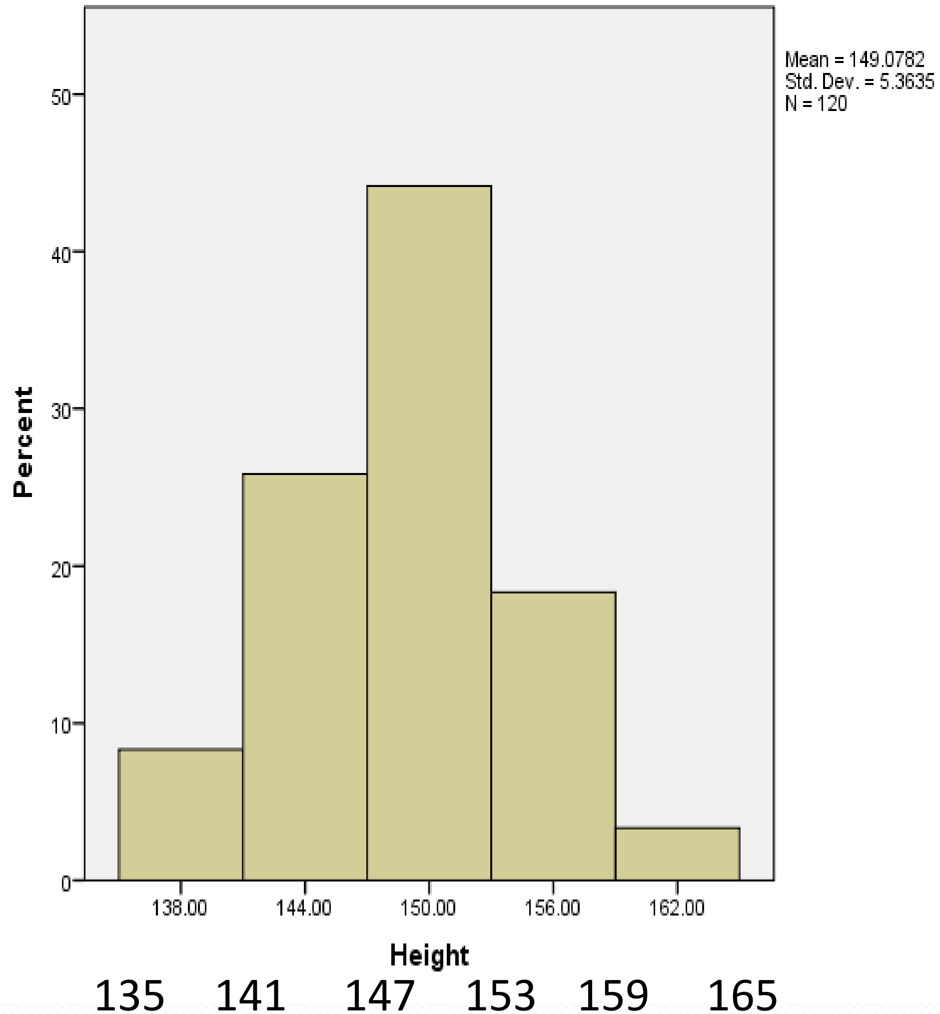
Histogram (Bin Size)



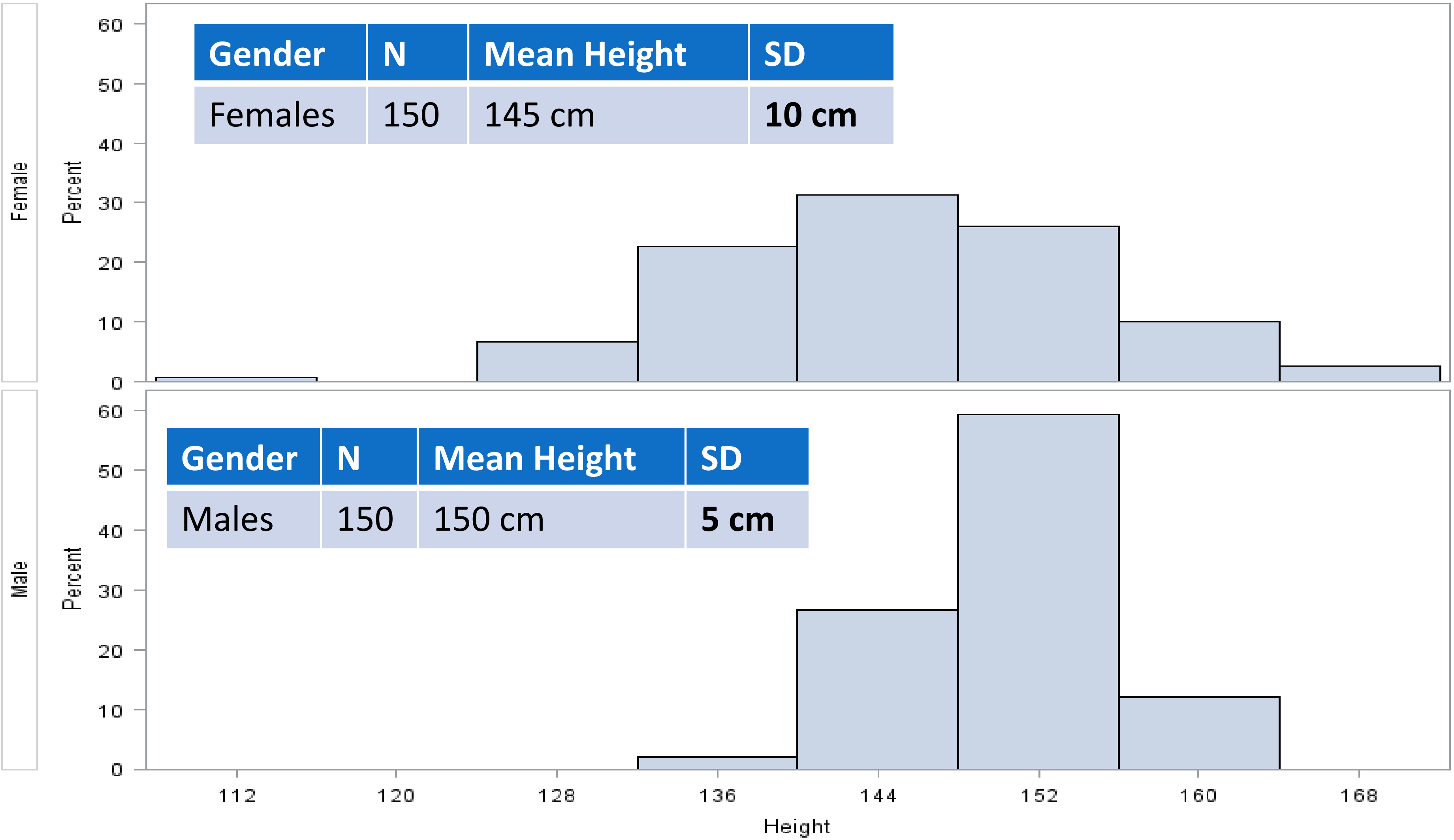


Histogram (Y-axis)





Histogram



Gender	N	Mean Height	SD
Females	150	145 cm	**10 cm**

Gender	N	Mean Height	SD
Males	150	150 cm	**5 cm**

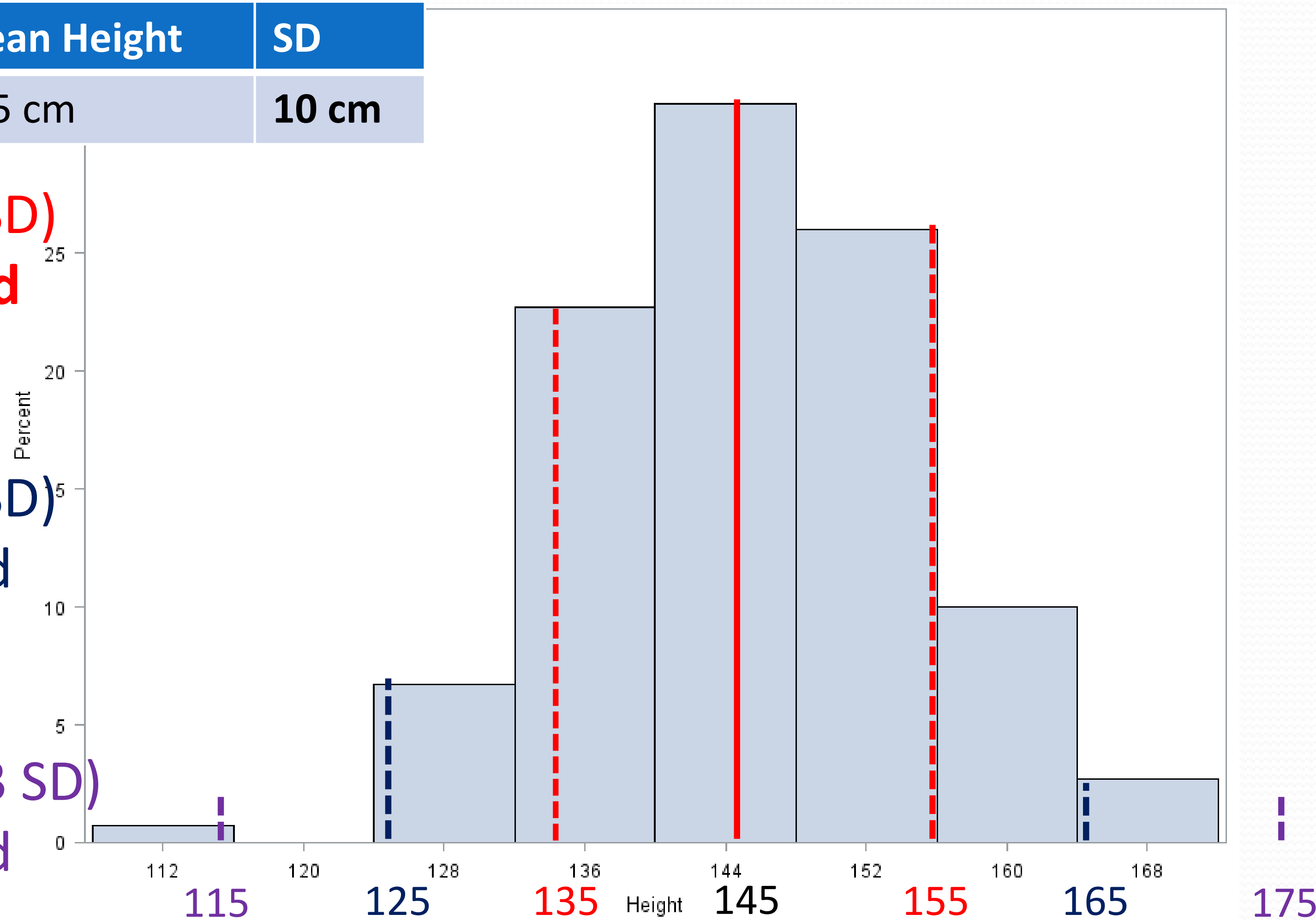
Histogram and SDs

Gender	N	Mean Height	SD
Females	150	145 cm	**10 cm**

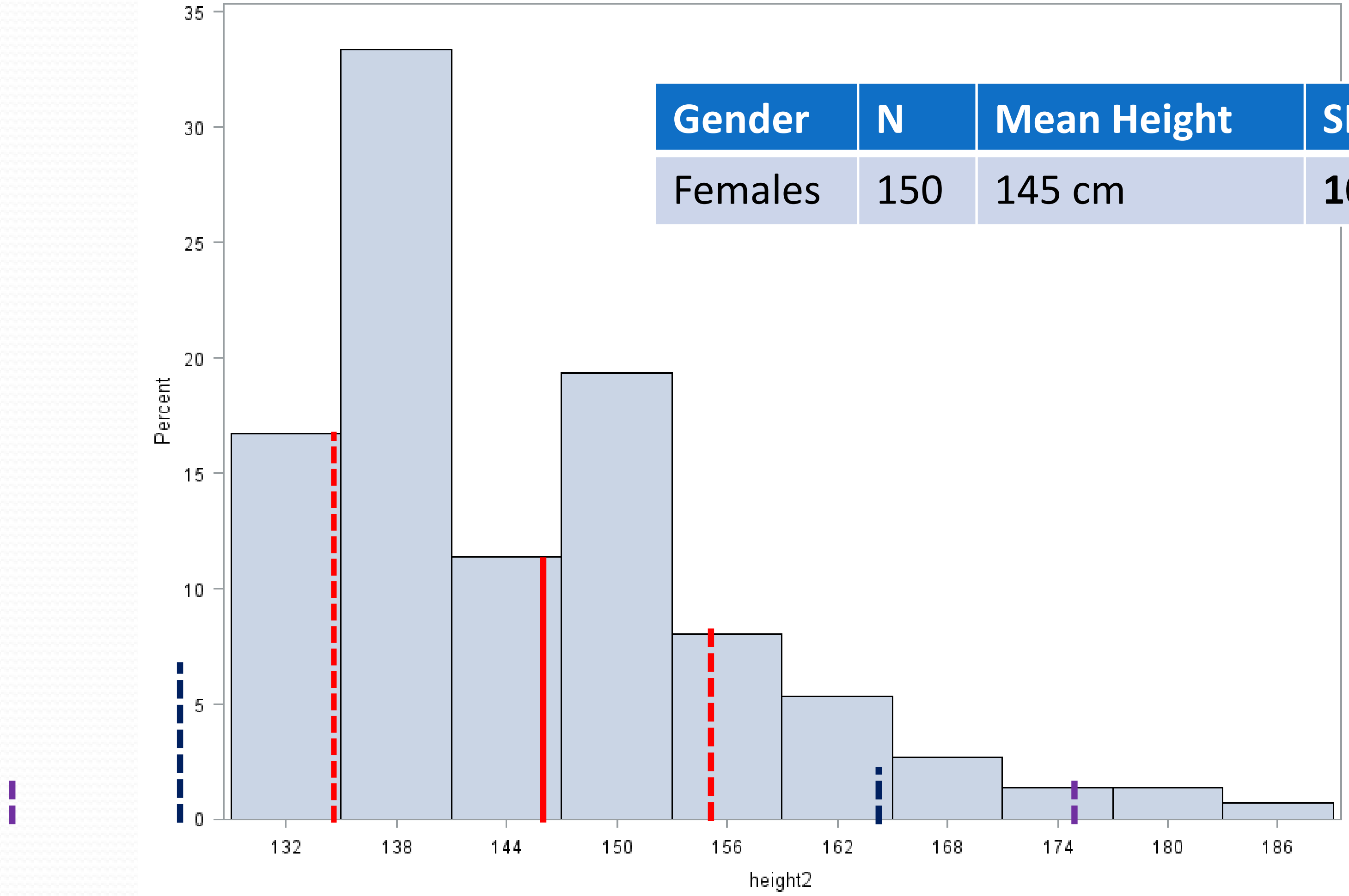
≈ 68% of data (1 SD)
(Between 135 and 155)

≈ 95% of data (2 SD)
(Between 125 and 165)

≈ 99.7% of data (3 SD)
(Between 115 and 175)



Skewed Data



Gender	N	Mean Height	SD
Females	150	145 cm	**10 cm**

Range

Students Heights		Students Heights	
150 cm	Range 8 cm	150 cm	Range 35 cm
151 cm		151 cm	
152 cm		152 cm	
153 cm		153 cm	
154 cm		154 cm	
155 cm		155 cm	
156 cm		156 cm	
157 cm		157 cm	
158 cm		205 cm	

Percentiles

Students Heights

150 cm
151 cm
152 cm
153 cm
154 cm
155 cm
156 cm
157 cm
180 cm

25th Percentile

Point where **at least** 25% of total data falls below

50th Percentile

Median

Middle observation

75th Percentile

Point where **at least** 75% of total data falls below

Interquartile Range

Students Heights

150 cm
151 cm
152 cm
153 cm
154 cm
155 cm
156 cm
157 cm
180 cm

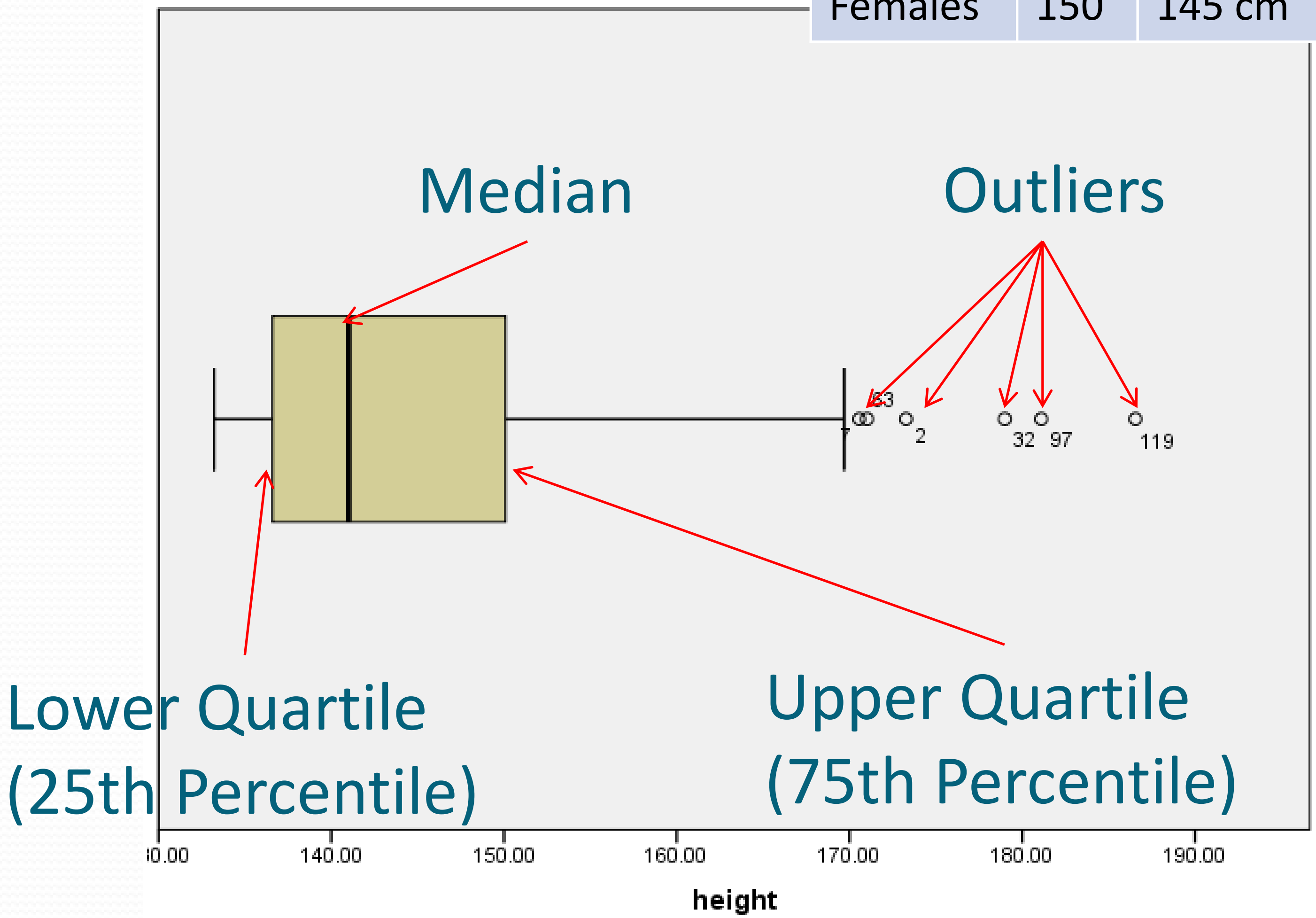
IQR represents of where the middle 50% of your data is

IQR = 156 – 152 = 4 cm

Larger IQR suggests more spread/variability

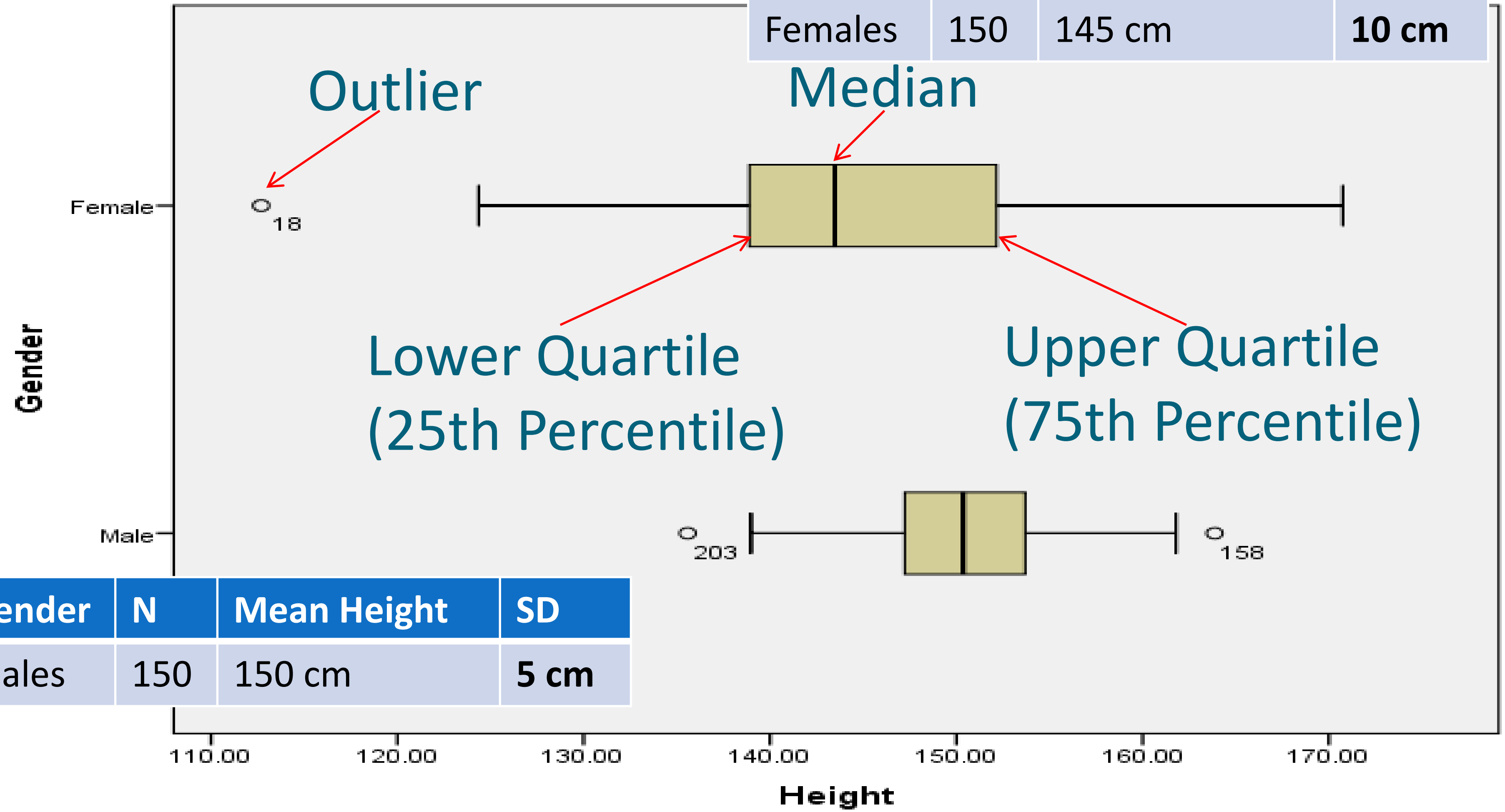
Boxplot

Gender	N	Mean Height	SD
Females	150	145 cm	**10 cm**



Boxplot

Gender	N	Mean Height	SD
Females	150	145 cm	**10 cm**



Gender	N	Mean Height	SD
Males	150	150 cm	**5 cm**

Continuous Data Summary

Measure of Central Tendency	Measure of Spread
Mean	Standard Deviation
Median (50th Percentile)	Lower Quartile Upper Quartile Interquartile Range

At minimum display either the Mean with Standard Deviation together, or Median (preferable when data is skewed) with percentiles

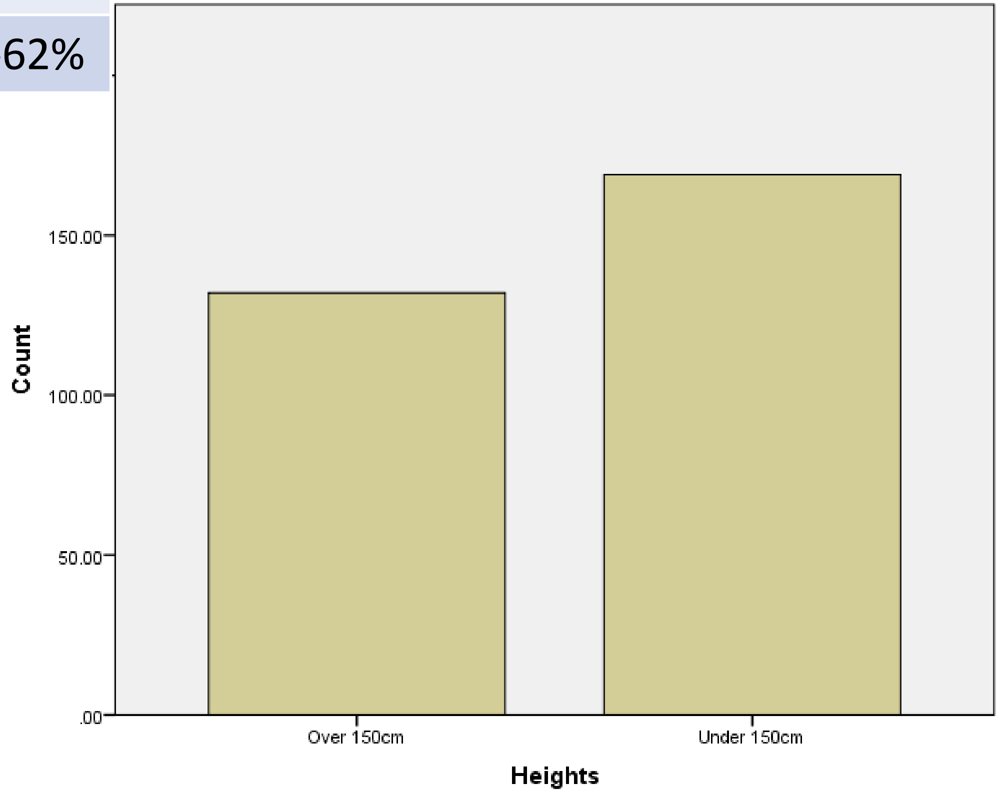
Categorical Data

- **Numerical Summaries**
  - N / counts
  - Percentages / proportions
  - Mode – Most frequent category
  - Confidence Intervals

- **Graphical summaries**
  - Bar Charts
  - Clustered Bar Charts

One Categorical Variable

	Students Heights		
	n	%	95% CI
Over 150cm	132	43.9%	38%-50%
Under 150cm	169	56.1%	50%-62%



Two Categorical Variables

- **Table Structure (2x2)**

	Outcome			
	Level 1		Level 2	
	n	%	n	%
Predictor 1				
Level 1	a	Row %	b	Row %
Level 2	c	Row %	d	Row %

Compare these % against each other

Compare these % against each other

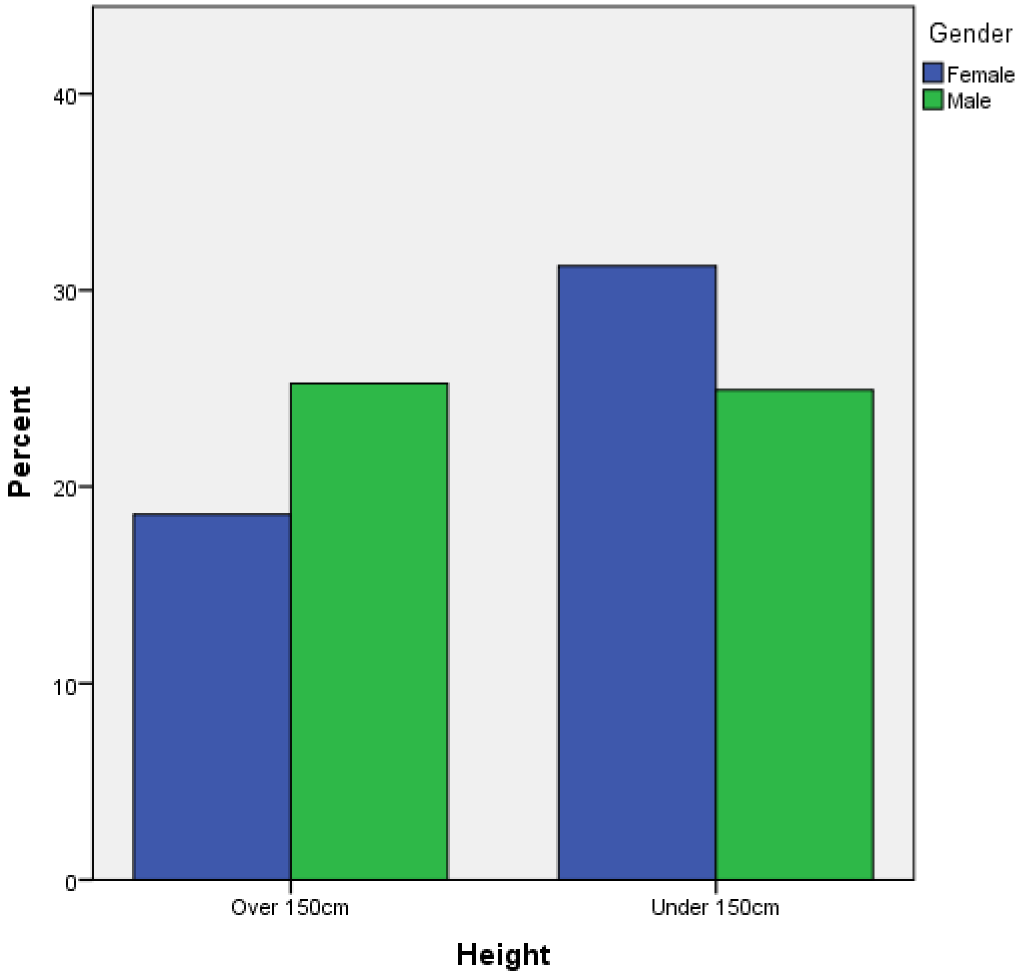
2x2 Tables

	Students Heights			
	Over 150cm		Under 150cm	
	n	%	n	%
Gender				
Female	56	37.3	94	62.7
Male	76	50.3	75	49.7

Males are more likely to be over 150cm

Females are more likely to be less than 150cm

Clustered Bar Charts



Other graphs?

HYPOTHESIS TESTING

Hypothesis testing

Research question

Do knee surgery patients that undergo a newly created advanced (but expensive) rehabilitation protocol after surgery, experience an improved quality of life (QOL) compared with patients that undergo standard rehab protocol 3 months after surgery?

Hypothesis testing

Two group comparison with a continuous outcome measure.

What is the specific outcome measure?

Trying to prove a difference or equality in the outcome between the groups?

In which direction are you trying to prove difference/equality in the outcome measure between the groups?

What is the magnitude of difference or equality?

Hypothesis testing

Two group comparison with a continuous outcome measure.

What is the specific outcome measure?
QOL?? → Oxford Knee Score (0-48) (Lower score indicates worse knee arthritis)
Trying to prove a difference or equality in the outcome between the groups?
Difference
In which direction are you trying to prove difference/equality in the outcome measure between the groups?
Increase in OKS for the advanced group.
What is the magnitude of difference or equality?
Minimum clinically important difference for OKS = 5 (Clement et al.)

Hypothesis testing

Null hypothesis: The OKS is not higher in the advanced group by at least 5 units compared to the standard care group.

Alternative hypothesis: The OKS is higher by at least 5 units for the advanced group compared with the standard care group.

Hypothesis testing

Research question

Is there a difference in the number of circulating tumour cells (CTC) between patients with stage 2 and stage 3 cancers?

Hypothesis testing

Two group comparison with a continuous outcome measure.

What is the specific outcome measure?
Circulating tumour cells (CTC)
Trying to prove a difference or equality in the outcome between the groups?
Difference
In which direction are you trying to prove difference/equality in the outcome measure between the groups?
Not specified in research question, though could expect to observe patients with stage 3 to have more CTCs than patients with stage 2 cancer.
What is the magnitude of difference or equality?
?? – Need to review literature to determine what is clinical difference in CTCs

Hypothesis testing

Null hypothesis: There is no difference in the CTCs between stage 2 and 3 cancers.
Alternative hypothesis: There is a difference in the CTCs between stage 2 and 3 cancers.

Null hypothesis: Patients with stage 3 cancers do not have more CTCs than patients with stage 2 cancers.
Alternative hypothesis: Patients with stage 3 cancers have more CTCs than patients with stage 2 cancers.

Hypothesis testing

Research question

Do knee surgery patients that undergo a new rehabilitation program which is cheaper and requires less resources experience similar QOL outcomes 3 months after surgery compared with current standard care?

Hypothesis testing

Two group comparison with a continuous outcome measure.

What is the specific outcome measure?
Not specified – go with Oxford Knee score (OKS) again
Trying to prove a difference or equality in the outcome between the groups?
Equality
In which direction are you trying to prove difference/equality in the outcome measure between the groups?
Not specified, though could argue that only interested in outcome that is clinically worse.
What is the magnitude of difference or equality?
MCID OKS = 5 (Clement et al.)

Hypothesis testing

Null hypothesis: OKS is at least 5 units lower in the new rehab program than compared with standard care.

Alternative hypothesis: OKS is no less than 5 units lower in the new rehab program compared with standard care.

Hypothesis testing

Translate your research question into a set of null and alternative hypothesis.

The way the hypotheses are set up have an influence on sample size, study design, and statistical analysis (topics of future seminars).

These can be extended to more groups and different types of outcomes.

References

Clement, N. D., et al. (2014). "The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty." Knee Surg Sports Traumatol Arthrosc 22(8): 1933-1939.