UNIVERSITY OF ABERDEEN

Independent samples t-test

David McLernon

Research Fellow in Medical Statistics

Issues to consider when choosing a statistical test

- Comparing groups of continuous data or categorical data?
- Number of groups for comparison
- Paired or independent groups?
- Parametric vs non-parametric tests

Issues to consider when choosing a statistical test

- Comparing groups of continuous data or categorical data?
- Number of groups for comparison
- Paired or independent groups?
- Parametric vs non-parametric tests

Issues to consider when choosing a statistical test

- Comparing groups of continuous data or categorical data?
- Number of groups for comparison **=2**
- Paired or independent groups?
- Parametric vs non-parametric tests

Issues to consider when choosing a statistical test

- Comparing groups of continuous data or categorical data?
- Number of groups for comparison **=2**
- Paired or independent groups?
- Parametric vs non-parametric tests

Issues to consider when choosing a statistical test

- Comparing groups of continuous data or categorical data?
- Number of groups for comparison **=2**
- Paired or independent groups?
- Parametric vs non-parametric tests

Assumptions of the independent groups t-test

1. Distribution: Outcome variable is normally distributed in both groups

Assumptions of the independent groups t-test

1. Distribution: Outcome variable is normally distributed in both groups

2. Comparison is between two independent groups

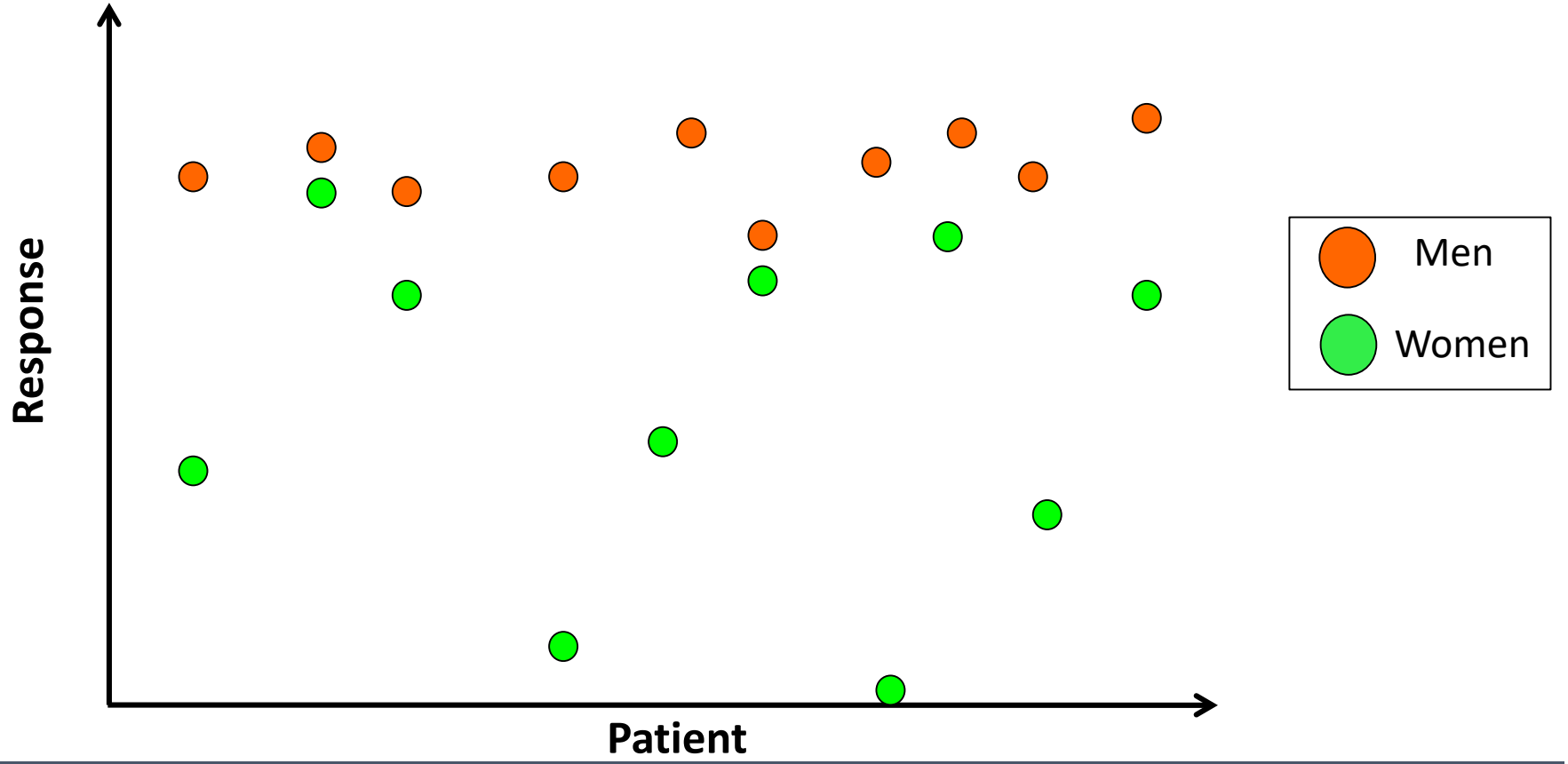
Assumptions of the independent groups t-test

1. Distribution: Outcome variable is normally distributed in both groups

2. Comparison is between two independent groups

3. Variances are assumed to be equal

Unequal variances



Assumptions of the independent groups t-test

1. Distribution: Outcome variable is normally distributed in both groups

   CHECK BY PLOTTING A HISTOGRAM FOR EACH GROUP

2. Comparison is between two independent groups

3. Variances are assumed to be equal

Assumptions of the independent groups t-test

1. Distribution: Outcome variable is normally distributed in both groups

2. Comparison is between two independent groups

   DIFFERENT INDIVIDUALS IN EACH GROUP

3. Variances are assumed to be equal

Assumptions of the independent groups t-test

1. Distribution: Outcome variable is normally distributed in both groups

2. Comparison is between two independent groups

3. Variances are assumed to be equal

USE LEVENE'S TEST:

- H_o: There is no difference between the variance of group 1 and the variance of group 2 in the population

  (i.e. $H_o: \sigma_1^2 = \sigma_2^2$)

Example of independent samples t-test

- Is there a difference between the mean total cholesterol of males and females?

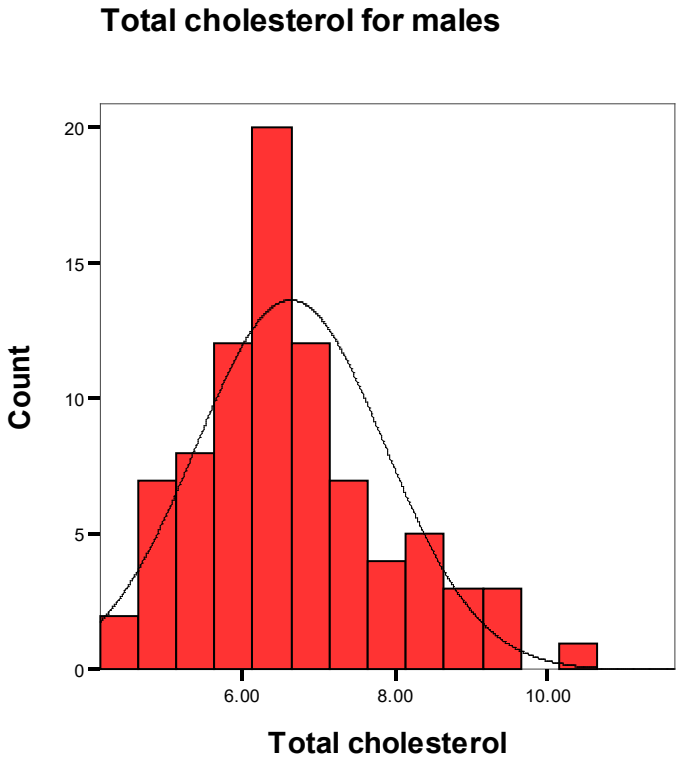
- The null hypothesis is that the group means are the same. That is:

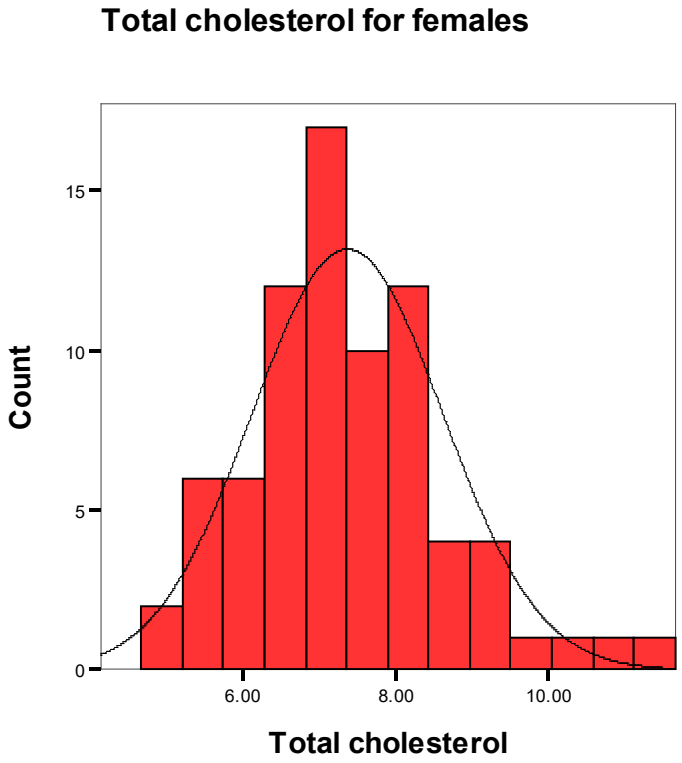
$$H_0: \mu_{Males} = \mu_{Females}$$

- The alternative hypothesis:

$$H_1: \mu_{Males} \neq \mu_{Females}$$

Normality assumption for example





Example of independent samples t-test

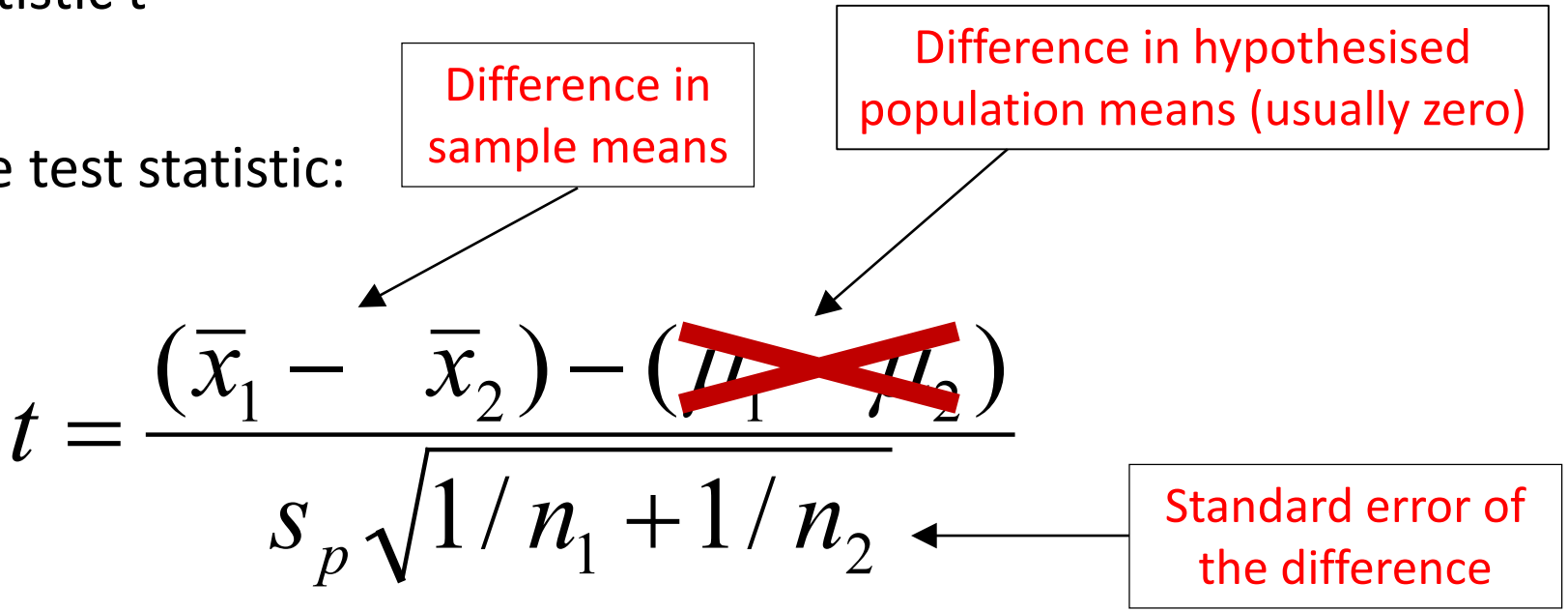
Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506

Test statistic

- In order to perform the test, we must again construct a test statistic t

- The test statistic:



Test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{1/n_1 + 1/n_2}}$$

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506

Test statistic

$$t = \frac{(7.36 - 6.63) - 0}{s_p \sqrt{1/77 + 1/84}}$$

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506

Test statistic

$$t = \frac{(7.36 - 6.63) - 0}{s_p \sqrt{1/77 + 1/84}}$$

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506

s_p is the pooled SD i.e. the average of the 2 sample variances

$$s_p = \sqrt{\frac{(df_1)s^2{}_1 + (df_2)s^2{}_2}{df_1 + df_2}}$$

Test statistic

$$t = \frac{(7.36 - 6.63) - 0}{s_p \sqrt{1/77 + 1/84}}$$

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506

s_p is the pooled SD i.e. the average of the 2 sample variances

$$s_p = \sqrt{\frac{(76)1.2575^2 + (83)1.2379^2}{76 + 83}} = 1.247$$

Test statistic

$$t = \frac{(7.362 - 6.626) - 0}{1.247\sqrt{1/77 + 1/84}}$$

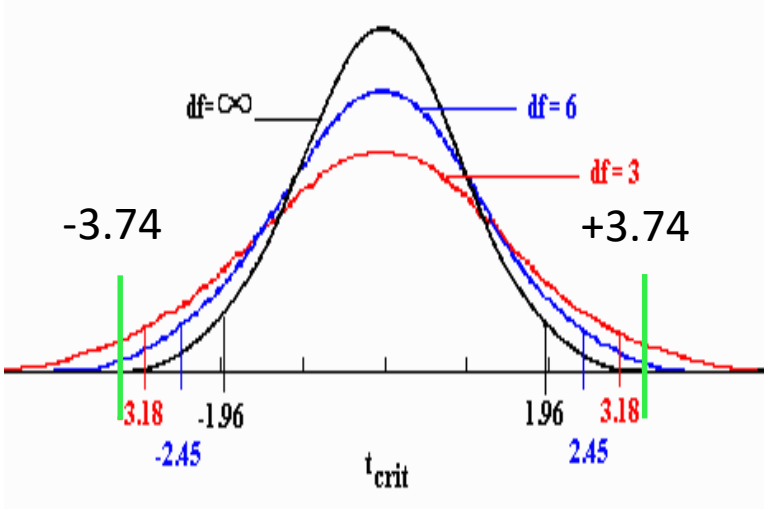
Therefore,

$$t = 3.74$$

with 77+84-2=159 df

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506



http://www.psychstat.missouristate.edu/introbook/sbk24m.htm

Output from example

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Total cholesterol (mmol/l)	Equal variances assumed	.011	.918	3.740	159	.000	.73602	.19679	.34736	1.12467
	Equal variances not assumed			3.738	157.322	.000	.73602	.19692	.34706	1.12497

Output from example

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Total cholesterol (mmol/l)	Female	77	7.3622	1.25750	.14331
	Male	84	6.6262	1.23788	.13506

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Total cholesterol (mmol/l)	Equal variances assumed	.011	.918	3.740	159	.000	.73602	.19679	.34736	1.12467
	Equal variances not assumed			3.738	157.322	.000	.73602	.19692	.34706	1.12497

Therefore assumption of equal variances in each group holds

Test statistic, t

df=no. of males + no. of females - 2

P<0.05

Mean difference = mean for males – mean for females

Conclusions

- Female total cholesterol levels were on average 0.736mmol/l higher than male total cholesterol.
- The value of t=3.74 is highly significant at the 5% level.
- Therefore, we reject the null hypothesis and conclude that female total cholesterol levels are significantly higher than male total cholesterol levels in the population.
- In the wider population there is a 95% chance that the actual mean difference is between 0.347 and 1.125mmol/l.

Summary

- Independent samples t-test is a parametric test that is used when we want to compare the mean of a continuous measurement in two independent groups.
- The assumptions:
  - The continuous or outcome measure must be approximately normally distributed in each group
    - Tested by drawing a histogram
  - The variance of the outcome measure is assumed equal in both groups
    - Tested using Levene's test.

Copyright Statement

- Unless otherwise stated all content Copyright University of Aberdeen.
- The University of Aberdeen subscribes to the Copyright Licensing Agency's Higher Education Photocopying and Scanning Licence. You may access, download and print out a copy of any material included under the terms of this licence.
- Any digital or print copy supplied to or made by you are for use in connection with this Course of Study. You may retain such copies after the end of the course, but strictly for your own personal use.
- All copies (including electronic copies) shall include this Copyright Notice and shall be destroyed and/or deleted if and when required by the University of Aberdeen.
- Except as provided by copyright law, no further copying, storage or distribution (including by e-mail) is permitted without the consent of the copyright holder.
- The author has moral rights in the work. No distortion, mutilation, or other modifications of the work, or any other derogatory treatment prejudicial to the honour or reputation of the author is permitted.