Comparison of two means

Professor Dr Abubakir M. Saleh

Biostatistics NUR304

Fall semester

7th week

30/11/2023

Outline

1. Compare the mean of a sample with population mean.
   - e.g. mean hemoglobin of a sample & mean hemoglobin of population

2. Compare the mean of one sample with the mean of another independent sample.
   - e.g. mean hemoglobin of a sample of boys & mean hemoglobin of a sample of girls

3. Compare between the means of one sample but in 2 occasions.
   - e.g. mean hemoglobin of a sample of boys before treatment & mean hemoglobin of the same sample of boys after treatment

Objectives

At the end of this lecture, students should be able to :

- Identify applications of T-test.
- Use appropriate equation for different t-test applications.

1. Sample mean, and population mean

E.g.

- The mean hemoglobin of 10 persons was 8.95
- The mean hemoglobin of population is 12
- Is there a statistically significant difference between the two means (8.95 and 12)?
- Or the difference was due to chance?

1.Sample mean, and population mean

- The general steps of testing hypothesis must be followed:
- Assumption (normal distribution)
- Hypothesis
  - Ho: Sample mean = Population mean.
  - H_A: Sample mean ≠ Population mean.
- Level of significance 0.05
  - Degrees of freedom (df) = n – 1
- The statistics

$$t = \frac{\bar{X} - \mu}{SE}$$

Example

- The following data represents hemoglobin values in gm/dl for 10 patients:

10.5	9	6.5	8	11
7	7.5	8.5	9.5	12

Mean=8.95 SD= 1.80201

- Is the mean value for patients significantly differ from the mean value of general population

(12 gm/dl).

- Evaluate the role of chance.

Solution

- Mention all steps of testing hypothesis.

$$t = \frac{\bar{X} - \mu}{SE} \qquad t = \frac{8.95 - 12}{\frac{1.80201}{\sqrt{10}}} = -5.352$$

- 5.352 is the calculated value (neglect -)
- Then compare with tabulated value (standard t distribution table), for 9 df, and 5% level of significance. It is = 2.262
- The calculated value (5.352) > tabulated value (2.262).

  P<0.05
- Reject Ho and conclude that there is a statistically significant difference between the mean of sample and population mean, and this difference is not due t o chance.

Table A.2

t Distribution: Critical Values of *t*

		Significance level					
Degrees of freedom	*Two-tailed test:*	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850

2.Two independent samples

The following data represents weight in Kg for 10 males and 12 females.

Males:

80	75	95	55	60
70	75	72	80	65

Females:

60	70	50	85	45	60
80	65	70	62	77	82

2.Two independent samples, cont.

- Is there a statistically significant difference between the mean weight of males and females. Let alpha = 0.05
- To solve it follow the steps and use this equation.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^{\ 2} + (n_2 - 1)S_2^{\ 2}}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

S here means SD of mean

Results

- Mean1=72.7 Mean2=67.17
- SD1=11.33 SD2=12.56
- df = n1+n2-2=20

$$t = \frac{72.7 - 67.17}{\sqrt{\frac{(10-1)11.33^2 + (12-1)12.56^2}{10+12-2}\left(\frac{1}{10} + \frac{1}{12}\right)}}$$

- t = 1.074
- The tabulated t for alpha 0.05 is 2.086
- The calculated value (1.074) < tabulated value (2.086). P>0.05
- Then accept Ho and conclude that there is no significant difference between the 2 means. This difference may be due to chance.

TABLE A.2

t Distribution: Critical Values of *t*

		Significance level					
Degrees of freedom	*Two-tailed test:*	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850

3.One sample in two occasions

- Systolic blood pressure of 8 persons
- Mean before treatment = 196.9
- Mean after treatment = 138.8
- Is there a statistically significant improvement or this change is due to chance alone?

3.One sample in two occasions

- Mention steps of testing hypothesis.
- The df here = n – 1.
- $\bar{d}$ = $\text{Mean}_{\text{before}} - \text{Mean}_{\text{after}}$

$$t = \frac{\bar{d}}{\frac{sd}{\sqrt{n}}}$$

$$sd = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}$$

Example: Blood pressure of 8 patients, before & after treatment

BP before	BP after	d	d^2
180	140	40	1600
200	145	55	3025
230	150	80	6400
240	155	85	7225
170	120	50	2500
190	130	60	3600
200	140	60	3600
165	130	35	1225
Mean d=465/8=58.125		$\sum d=465$	$\sum d^2=29175$

Results and conclusion

$$sd = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}$$

$$sd = \sqrt{\frac{29175 - \frac{(465)2}{8}}{8-1}}$$

$$t = \frac{\bar{d}}{\frac{sd}{\sqrt{n}}}$$

$$t = \frac{196.9 - 138.8}{\frac{17.5}{\sqrt{8}}}$$

- t=9.387
- Tabulated t (df7), with level of significance 0.05, = 2.365
- Calculated (9.387) > tabulated (2.365)
- P<0.05
- We reject Ho and conclude that there is significant difference between BP readings before and after treatment.

Table A.2

t Distribution: Critical Values of *t*

		Significance level					
Degrees of freedom	*Two-tailed test:*	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850

Summary

- t test to compare between two means with three applications:
  - To compare the mean of a sample with population mean.
  - To compare the mean of one sample with the mean of another independent sample.
  - To compare between the means of one sample but in 2 occasions.

References

- Essential Medical Statistics, by Betty Kirkwood & Jonathan Sterne
  (Published by Blackwell)
  Statistics Without Tears, a Primer for Non-mathematicians, by Derek Rowntree
  (Published by Penguin)

Questions?