Comparing two Population Means & Proportions

Dr. M.M. Mweu,

Level II MBChB Biostatistics,

30 November, 2016

Comparing two means

- The objective here is to compare the mean value in two populations, or in two sub-populations by:
  - Calculating a **confidence interval** for the difference between 2 sample means which allows for sampling error in estimating the difference between the *true* means
  - **Testing the hypothesis** that the true means in the 2 populations are equal
- The underlying assumptions in these calculations are:
  - Variable of interest is normally distributed
  - Observations are independent i.e. random samples are chosen independently from the 2 pops of interest and there's no connection, for example, between 1st observation in one sample and the 1st observation in the other sample

Sampling distribution of the difference betwn 2 means

- Random samples of size n_1 and n_2 are taken from 2 pops of interest. The means and SDs of a quantitative variable x in the 2 pops and samples are:

	Pop 1	Sample 1	Pop 2	Sample 2
Mean	μ_1	$\overline{x_1}$	μ_2	$\overline{x_2}$
SD	σ_1	s_1	σ_2	s_2

- If random samples of a given size of the variable x were taken repeatedly in each of pop 1 & pop 2 and each time we measured ($\overline{x_1}$ - $\overline{x_2}$) we would find that:
  - The values of $\overline{x_1}$, s_1, $\overline{x_2}$, s_2 would vary from sample to sample
  - The values of ($\overline{x_1}$ - $\overline{x_2}$) would be distributed symmetrically (normal distr.) above and below the true population value (μ_1 -μ_2)
  - Values near (μ_1 -μ_2) would occur more frequently than values far from (μ_1 -μ_2)

Confidence interval for difference betwn 2 means

- Assuming a large sample size (n≥40) the 95% CI for the difference betwn 2 means is given by:

$$(\overline{x_1} - \overline{x_2}) \mp Z_{\alpha/2} * \sqrt{\frac{s^2{}_1}{n_1} + \frac{s^2{}_2}{n_2}}$$

NB: If there's no real difference betwn the 2 means the CI around the difference should include zero

Example:

- In a cohort study in northeast Brazil, the mental and psychomotor development of low birth weight (1500-2499g) infants born at ≥37 weeks gestation (at term), were compared to those of a control sample of infants born with appropriate birth weight (3000-3499g). Results for mental devpt at 12 months of age in samples of low and appropriate birth weight infants were as follows:

Confidence interval for difference betwn 2 means

Mental development score			
	n	Mean	SD
Appropriate b. weight (ABW)	84	115.1	9.44
Low b. weight (LBW)	84	108.1	11.50

- What can be said about the mental development scores of children in these 2 groups?

$$n_1 = 84, \quad \overline{x_1} = 115.1, \quad s_1 = 9.44$$

$$n_2 = 84, \quad \overline{x_2} = 108.1, \quad s_2 = 11.50$$

The difference in means is (115.1 – 108.1) = 7.0

Standard error is $\sqrt{(9.44^2/84 + 11.5^2/84)} = \sqrt{2.635} = 1.623$

Confidence interval for difference betwn 2 means

So the 95% CI for the difference (μ_1 -μ_2) is given by:

(115.1 – 108.1) $\mp$ 1.96$\sqrt{(9.44^2/84 + 11.5^2/84)}$ or 7.0 $\mp$ 1.96*1.623

which is: 3.82 to 10.18

So the data suggest that at age 12 months, ABW children have, <u>on average</u>, a mental development score betwn 3.8 and 10.2 points higher than LBW children

Significance test for comparison of 2 means

- To test the hypothesis about the difference (μ_1 -μ_2) between 2 pop means, we write:

$$H_o: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

After calculating the Z, the P-value may be read from tables of the normal distribution.

Example:

In the Brazilian study,

$(115.1 - 108.1)/\sqrt{(9.44^2/84 + 11.5^2/84)}$ or $7.0/1.623 = 4.31$

From the Z tables, we find that $P<0.001$. **In other words, there's strong evidence for a real difference betwn the 2 pop means**

Significance test for comparison of 2 means

- **NB:** Note the close relationship between the significance test and the confidence interval. The test will give a P-value less than 0.05 if the 95% CI excludes the hypothesised value (0) and vice versa. In the example, the 95% CI (3.82 to 10.2) excludes zero, and the P-value is less than 0.05.
- If sample sizes are small (n<40), and distribution of the individual values are approx. normal, a **t distribution** is used
- An additional assumption is that σ_1 and σ_2 are equal to a common value (**common variance**). So,

CI: $(\overline{x_1} - \overline{x_2}) \mp t^* \sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ *where* $S_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

The value of t is read from the tables of the t-distribution with $(n_1 + n_2 - 2)$ degrees of freedom

Significance test: $t = \frac{(\overline{x_1} - \overline{x_2})}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ **(Unpaired t-test)**

Significance test for comparison of 2 means

- However, if variances are different (**unequal variances**) then:
- First, test for equality of variances using an F test:

$$F_{(V_1,V_2)} = \frac{S^2{}_1}{S^2{}_2}$$

where the larger variance forms the numerator and $V_1 = n_1 - 1$ & $V_2 = n_2 - 1$ df

- If the F test-statistic calculated above is significant then perform:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s^2{}_1}{n_1} + \frac{s^2{}_2}{n_2}}}$$ (**Welch t-test**)

- **NB:** It is common to see t-tests used where sample sizes are large and adequate for use of z test.

Comparing two population means for paired data

- A **paired t-test** is used to compare 2 pop means when you have 2 samples in which observations in one sample are *paired* with observations in the other sample
- Examples of paired data:
  - Before-and-after observations on the same subjects (e.g. students diagnostic test results before and after a particular course)
  - A comparison of 2 different methods of measurement or 2 different treatments where the measurements/treatments are applied to the same subjects (e.g. antigen and antibody ELISA tests on blood samples)
  - In matched study designs e.g. in a matched clinical trial where we wish to test a new therapy for leg ulcers in sickle cell anaemia relative to an existing therapy. We form pairs of patients *matched for age, sex and severity* of ulcers and randomly allocate one member of the pair to new therapy, and the other to the existing therapy and compare the quantitative outcomes on the 2 treatments
- Observations in a pair are *not* independent, however different pairs are independent

Comparing two population means for paired data

- Let's assume 2 measurements (*x and y*) are taken on each sample drawn from different subjects. Steps involved in carrying out a paired t-test are:
  - Calculate the difference ($d_i = x_i - y_i$) betwn the 2 observations on each pair, making sure you distinguish betwn positive and negative differences
  - Calculate the mean difference, $\bar{d}$ (NB: the paired t-test assumes $d_i s$ are normally distributed, if not then *non-parametric tests* are used)
  - State the hypotheses:

$$H_0: \bar{D} = 0$$

$$H_a: \bar{D} \neq 0$$

  - Calculate the standard deviation of the differences, s_d, and use these to calculate the standard error of the mean difference, $SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$
  - Calculate the t-statistic which is given by $T = \frac{\bar{d}}{SE(\bar{d})}$. Under the H_0, this statistic follows a *t-distribution* with $n-1$ degrees of freedom
  - Use *t-tables* to compare your value of T to the critical t_{n-1} distribution. This will give the *P-value* for the paired t-test

Comparing two population means for paired data

Example

- During a nutritional survey, a quality control exercise was carried out to check the agreement betwn 2 observers in taking skinfold measurements. Both observers measured the same 15 adults, with the following results:

	Skinfold measurement (mm)		Difference
Subject	**Observer A**	**Observer B**	**A – B**
1	21.5	18.3	+3.2
2	25.0	21.5	+3.5
3	19.3	16.3	+3.0
4	33.9	32.3	+1.6
5	15.9	19.1	-3.2
6	39.9	34.6	+5.3
7	20.8	16.8	+4.0
8	33.2	31.0	+2.2
9	34.4	32.5	+1.9
10	20.5	18.6	+1.9
11	14.6	14.0	+0.6
12	15.8	15.5	+0.3
13	18.4	16.4	+2.0
14	25.5	19.0	+6.5
15	19.0	17.6	+1.4

Comparing two population means for paired data

Example

- Mean difference $\bar{d} = +2.28$ and standard deviation $s_d = 2.25$
- The T statistic will given by: $T = \frac{2.28}{\frac{2.25}{\sqrt{15}}} = \mathbf{3.92}\ with\ 14\ df$
- This gives $\mathbf{0.001 < P < 0.002}$, we reject H_0 and conclude that there's **strong evidence of a real difference between the two observers**.

Comparing two proportions

- As with means, to compare two proportions (percentages) we use:
  - Statistical test of significance
  - 95% CI for the difference in the 2 proportions

Statistical significance test:

Example:

In a clinical trial for advanced (metastatic) breast cancer, patients were randomly assigned to L-Pam or CMF. Tumour response was defined as a shrinkage of tumour surface area by at least a half for a minimum of 2 weeks:

		CMF	L-Pam	
Tumour response	Yes No	49 (52.7%) 44	18 (19.8%) 73	67 (36.4%) 117
Total patients		93	91	184

Comparing two proportions

Statistical significance test

- Is CMF better than L-Pam in shrinkage of the tumour?

$$H_o: \pi_1 = \pi_2$$

$$H_a: \pi_1 \neq \pi_2$$

$$Z = \frac{(p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad \textit{where } \bar{p} = \frac{x_1+x_2}{n_1+n_2}$$

- Here we make use of the fact that the difference between the 2 observed proportions has approximately a normal distribution (**normal approximation to the binomial distribution**): $n\pi \geq 5$; $n(1-\pi) \geq 5$

- $|Z| = \frac{52.7\% - 19.8\%}{\sqrt{36.4*63.6*\left(\frac{1}{93}+\frac{1}{91}\right)}} = \frac{32.9}{7.1} = 4.63$ P<0.001 (**Strong evidence that CMF patients had better response than L-Pam patients**)

Comparing two proportions

<u>Confidence interval:</u>

- The 95% CI for the difference betwn 2 proportions (percentages) is:

  Observed difference (p) $\pm$ 1.96*Standard error of difference

$$p \pm 1.96 * \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

- So in the previous example the 95% CI is given by:

$$32.9\% \pm 1.96 * \sqrt{\frac{52.7*(100-52.7)}{93} + \frac{19.8*(100-19.8)}{91}}$$

$$= 32.9\% \pm 13.04$$

= 19.86% to 45.94% (interval doesn't include zero)

So, we are 95% confident that the <u>true pop difference</u> in tumour responses between CMF & L-Pam is between 19.86% and 45.94%

NB: Standard error formula in the above calculation <u>doesn't assume</u> the H_o of the 2 proportions being equal (common variance).