Medical Statistics (MATH38071) - Exercise Sheet 3
(Analysis of Binary Outcome Measures)

Question 1 – 5 data from Critical Appraisal Exercise 1 [Quist-Paulsen P Gallefoss F Randomised controlled trial of smoking cessation intervention after admission for coronary heart disease BMJ 2003;327:1254.]

1. Using a z-test of proportions check the analysis in table 2 of the paper comparing the smoking cessation rates at 12 months.

   Solution: From the paper

Ceased Smoking	Intervention (T)	Control (C)	Total
Yes	57 (57%)	44(37%)	101
No	43	74	107
Total	100	118	218

The normal approximation to the binomial assumption is justified if $n_T p, n_T(1-p), n_C p, n_C(1-p) \geq 5$

where $p = \dfrac{r_T + r_C}{n_T + n_T}$ =101/218.

Since $p < (1-p)$ and $n_T < n_C$, $Min\left[n_T p, n_T(1-p), n_C p, n_C(1-p)\right] = n_T p = 100 \times \dfrac{101}{218} = 46.3$.

Since all expected frequencies are greater than 5, the normal approximation is justified.

$p_T - p_C = \dfrac{r_T}{n_C} - \dfrac{r_T}{n_C}$ =57/100-44/118=0.197.

$$SE_{null}\left[p_T - p_C\right] = \sqrt{p(1-p)\left(\frac{1}{n_T} + \frac{1}{n_C}\right)} = 0.0678$$

$$Z = \frac{|p_T - p_C|}{SE_{null}\left[p_T - p_C\right]} = \frac{0.1971186}{0.0677773} = 2.9083$$

$$p - value = 2\left(1 - \Phi(|Z|)\right) = 2\left(1 - \Phi(2.9083)\right)$$

From tables of the normal distribution $\Phi(2.908) \cong 0.99818$. Hence for a two-sided test

$p - value = 2(1 - 0.99818) = 0.00384$, which is statistical significant at a 1% level.

2. Calculate the point estimate and the 95% confidence interval of the treatment effect at 12 months.

Solution: The (1-α) confidence interval is given by $p_T - p_C \pm z_{\alpha/2} SE\left[p_T - p_C\right]$

From Q1 the point estimate $(p_T - p_C)$ is 0.197.

The non-null standard error for difference of proportions is

$$SE\left[p_T - p_C\right] = \sqrt{\frac{p_T\left(1-p_T\right)}{n_T} + \frac{p_C\left(1-p_T\right)}{n_C}} = 0.0665785$$

$$z_{\alpha/2} = \Phi^{-1}\left(1-\alpha/2\right) = \Phi^{-1}\left(0.975\right) = 1.96$$

For a 95% confidence interval is

$$p_T - p_C \pm z_{\alpha/2} SE\left[p_T - p_C\right] = 0.1971186 \pm 1.96 \times 0.0665785$$

So the 95% confidence interval is <u>0.067 to 0.328.</u>

3. Concisely summarize the results of this analysis in a narrative format giving (i) the rates for each treatment, (ii) the point estimate of the treatment effect and its 95% confidence interval and (iii) the p-value of the z-test.

Solution:

A concise summary could be

"The intervention significantly increased the rate of smoking cessation from 37%(44/118) in the control group to 57% (57/100) , which is an increase of 19.7% (95% c.i. 6.7% to 32.8 %,p=0.0038).

4. In the paper (page 3, para 4) a statistical analysis, sometimes called a sensitivity analysis, has been carried out by making assumptions regarding the missing data using all patients except those who have changed address or died. Death or change of address could in theory be due to the intervention and so be considered as adverse outcomes.
    a. Carry out a sensitivity analysis of the validated smoking cessation rate using **all randomised** patients assuming missing data are adverse outcomes. .

Solution:

Ceased Smoking	Intervention (T)	Control (C)	Total
Yes	57 (48%)	44(37%)	101
No	61	78	139
Total	118	122	240

$$Min\left[n_T p, n_T\left(1-p\right), n_C p, n_C\left(1-p\right)\right] = n_T p = 118 \times \frac{101}{240} = 49.7 .$$

Since all expected frequencies are greater than 5, the normal approximation for the binomial is justified.

$$Z = \frac{|p_T - p_C|}{SE_{null}\left[p_T - p_C\right]} = \frac{0.1223951}{0.0637443} = 1.92$$

$$\Phi(1.92) \cong 0.9719$$

Hence $p - value = 2(1 - 0.9719) = 0.056$.

This is not statistically significant at a 5%, but it is close to the 5% level and so might be considered as some evidence against the null hypothesis.

The (1-α) confidence interval is given by $p_T - p_C \pm z_{\alpha/2} SE\left[p_T - p_C\right]$

The non-null standard error for difference of proportions is $SE\left[p_T - p_C\right] = 0.0632948$

For a 95% confidence interval is

$$p_T - p_C \pm z_{\alpha/2} SE\left[p_T - p_C\right] = 0.1223951 \pm 1.96 \times 0.0632948$$

So the 95% confidence interval is -0.00166 to 0.246.

b. Considering the follow-up rates and the sensitivity analysis of part b, what conclusion do you draw from this analysis?

Solution:

Assuming all patients lost to follow-up did not cease smoking, which is a conservative sensitivity analysis, the intervention increased the rate of smoking cessation from 37%(44/122) in the control group to 48% (57/118) , a difference of 12.2% (95% c.i. -0.2% to 24.6%), but this was not quite statistically significant at a 5% level (p=0.056). Given that this conservative sensitivity analysis is almost statistically it might be reasonable to conclude that the intervention improved smoking cessation in this patient group. It should be noted nevertheless that the follow-up rate was lower in the intervention group (100/118=84.7%) than the control group (4/122=3.3%), which could indicate that some patients receiving the intervention felt harassed by the nurse follow-up or dislike fear arousal leading to drop-out from follow-up.

5. The number needed to treat (NNT) is a measure used to assess the effectiveness of a health-care intervention. The NNT is the number of patients who need to be treated to prevent one additional adverse outcome, in this case the number of patients who need to receive the smoking cessation intervention for one person to stop smoking.
   a. Write down an algebraic expression for NNT.

Solution: $NNT = \dfrac{1}{p_T - p_C}$ given in the notes.

   b. Calculate the NNT and its 95% confidence interval for the smoking cessation rates at 12 months calculated in Q2.

Solution: For this data

$$NNT = \frac{1}{0.197} = 5.073$$

The 95% confidence interval of NNT is 1/0.328 to 1/0.067 which is

3.05 to 15.01.

c. Calculate the NNT and its 95% confidence interval for the analysis in Q4.

Solution:

The point estimate (95% confidence interval) from Q4 are 0.122 (-0.00166 to 0.246). This includes zero so the confidence taking reciprocals the NNT is 1/0.122= 8.9196 and the confidence interval is 1/-0.00166= -602.4 and 1/0.246= 4.065. The NNT is 8.9 and the confidence limits are 602.4 and 4.065

d. What is the confidence interval for NNT from part c.

Solution:

The confidence interval for the rate difference in part c includes zero and so the NNT could be ±∞. A value of NNT less than zero. The confidence interval is therefore two region (-∞ to -602.4) and (4.1 to +∞) with the point estimate being in the second.

6. For the data from the SALK randomised trial calculate the point estimate and the 95% confidence of (i) the odds ratio (ii) the rate ratio comparing vaccine to placebo.

Solution (i) From notes confidence interval for odds ratio are calculated using

$$\exp\left[\log_e\left[\frac{r_T(n_C - r_C)}{(n_T - r_T)r_C}\right] \pm z_{\alpha/2}\sqrt{\frac{1}{r_T} + \frac{1}{n_T - r_T} + \frac{1}{r_C} + \frac{1}{n_C - r_C}}\right]$$

Substitution gives OR= (33x201,114)/(200,712x115)= 0.28753,

Data can be reorganised as a two by two table

Polio	Vaccine	Placebo	Total
Case	33 (0.000164)	115 (0.000571)	148 (0.000368)
Non Case	200,712	201,114	401,826
Total	200,745	201,229	401,974

To check assumptions for normal approximation consider $n_T p, n_T(1-p), n_C p, n_C(1-p) \geq 5$

The smallest is 200,745 x 0.000368 =73.9 so normal approximation.

$\log_e$(OR)= -1.24642 ,SE($\log_e$(OR)= 0.197506

Therefore 95% c.i. $\log_e$(OR) is -1.24642 ± 1.96 X 0.197506

which is(-1.63354,-0.85931).

Hence 95% c.i. for the odds ratio (OR) is (0.1952, 0.4235).