Medical Statistics (MATH38071) Solutions Exercise Sheet 7

(Equivalence and Non-Inferiority Trials)

1. Quinn et al [Critical Appraisal Exercise 2] compared conservative treatment with suturing for the treatment of minor hand wounds. The primary outcome was *cosmetic appearance* assessed by a 100mm visual analogue scale with 0 and 100 representing the worst and best outcomes respectively. The analysis of this presented in the paper is tabulated below.

	Treatment						*Mean difference*	
	Suture			*Conservative*				
	Mean (mm)	*s.d. (mm)*	*N*	*Mean (mm)*	*s.d. (mm)*	*n*	τ *(mm)*	*95% c.i.*
Cosmetic Appearance	83	10.0	41	80	11.3	40	-3	(-8 to 1)

(i) Suppose that a difference of 5 mm on the visual analogue scale is clinically important. If τ is the treatment effect, test the hypothesis $H_0 : |\tau| \geq \tau_E$ vs $H_1 : |\tau| < \tau_E$ with $\tau_E = 5$ using a 5% significance level.

Solution

To test $H_0 : |\tau| \geq \tau_E$ against the alternative hypothesis $H_1 : |\tau| < \tau_E$ one uses the $(1-2\alpha)$ confidence interval, given by $\overline{Y}_T - \overline{Y}_C \pm t_\alpha(\nu) SE\left[\overline{Y}_T - \overline{Y}_C\right]$ where $SE\left[\overline{Y}_T - \overline{Y}_C\right] = s\sqrt{1/n_T + 1/n_C}$ and

$$s = \sqrt{\frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T + n_C - 2}} \ .$$

$$s = \sqrt{\frac{40 \times 10^2 + 39 \times 11.3^2}{79}} = 10.662$$

$$SE\left[\overline{Y}_T - \overline{Y}_C\right] = 10.66\sqrt{1/41 + 1/40} = 2.369$$

Taking α=0.05, the 90% confidence interval is given by

$$\overline{Y}_T - \overline{Y}_C \pm t_\alpha(\nu) SE\left[\overline{Y}_T - \overline{Y}_C\right] = 80 - 83 \pm t_{0.05}(79) \times 2.369$$

From tables $t_{0.05}(80) = 1.6641$, $t_{0.05}(75) = 1.6654$. By linear interpolation $t_{0.05}(75) = 1.6644$ to 4 dp. Hence the 90% confidence interval is (-6.94 to 0.94). From the question the range of equivalence was $(-5, 5)$. Since the lower limit is less than –5 one cannot reject the null hypothesis at a 5% level.

(ii) Comment on the results of (i).

Solution

The paper concludes that the two procedures gave similar aesthetic outcome. This analysis suggests that there was insufficient evidence to do this if we use the suggested range of equivalence. The study was probably underpowered if the range $(-5,5)$ is used.

(iii) Using the pooled within-group standard deviation calculated in (i), estimate the sample size required to have 90% power to reject $H_0 : |\tau| \geq \tau_E$ vs $H_1 : |\tau| < \tau_E$ with $\tau_E = 5$ for a 5% level test assuming that $\tau = 0$ under the alternative hypothesis.

Solution

These are hypotheses for an equivalence trial. From the notes the formula for sample size is

$n = \frac{2\sigma^2}{\tau_E^2}\left(z_\alpha + z_{\beta/2}\right)^2$ where $(1-2\alpha)$ is the coverage of the confidence interval , $(1-\beta)$ is the power, σ is the within group standard deviation and τ_E is the limit of equivalence $H_0 : |\tau| \geq \tau_E$.

From the question $\alpha = 0.05$. Therefore from tables $z_\alpha = z_{0.05} = 1.645$.

From the question the power required was 90%, that is (1-β)=0.9. From tables $z_{\beta/2} = z_{0.05} = 1.645$.

From above σ = 10.662.

From question use $\tau_E = 5$.

Substitution give $n = \frac{2\sigma^2}{\tau_E^2}\left(z_\alpha + z_{\beta/2}\right)^2 = \frac{2 \times 10.662^2}{25}\left(1.645 + 1.645\right)^2 = 98.4$

Hence 99 patients would be needed in each arm to have a power greater than 90%.

(iv) Calculate the sample size required to test $H_0 : \tau \leq -5$ vs $H_1 : \tau > -5$ for a 5% level test, again assuming that $\tau = 0$ under the alternative hypothesis.

Solution

These are hypotheses for a non-inferiority trial. From the notes the formula for sample size is

$n = \frac{2\sigma^2}{\tau_N^2}\left(z_\alpha + z_\beta\right)^2$ where (1-α) is the coverage of the one-sided confidence interval and $(1-\beta)$ is the power.

From tables $z_\beta = z_{0.1} = 1.282$ and $\tau_N = 5$. Other values are the same as part (iii)

Substitution give $n = \frac{2\sigma^2}{\tau_N^2}\left(z_\alpha + z_{\beta/2}\right)^2 = \frac{2 \times 10.66^2}{25}\left(1.645 + 1.282\right)^2 = 77.9$

Hence, if a non-inferiority analysis is to be used, 78 patients would be needed in each arm to have a power greater than 90%.

2. A randomised controlled trial is carried out to compare a standard vaccine with a new vaccine that is easier to deliver. The vaccines are allocated to two equal size groups of 200 children each. One week after inoculation the immune response is measured in each child. An immune response is found in 86%(172/200) children with the standard vaccine and of 84%(168/200) children in the new vaccine. The researchers have decided that the vaccines would be considered to have equivalent efficacy if the proportion of children with an immune response in the two treatments differs by less than 5%. Assuming that a $(1-2\alpha)$ confidence interval of proportions corresponds to a test of proportions with a significance level α, test whether the immune response of the new vaccine can be considered to be equivalent to the existing vaccine using a 5% significance level.

Solution

From the notes on the *Analysis of Binary Outcome Measures* a $(1-\alpha)$ confidence interval is given by

$p_T - p_C \pm z_{\alpha/2} SE[p_T - p_C]$ where $SE[p_T - p_C] = \sqrt{\frac{p_T(1-p_T)}{n_T} + \frac{p_C(1-p_C)}{n_C}}$.

It is suggested that a $(1-2\alpha)$ confidence interval of proportions be used to test with significance level α.

For a 5% level one uses a 90% confidence interval. Hence $z_{\alpha/2} = z_{0.05} = 1.645$ in the formula above.

For the new vaccine, $n_T = n_C = 200$, $p_T = 168/200$ and $p_C = 172/200$.

Therefore

$$SE[p_T - p_C] = \sqrt{\frac{p_T(1-p_T)}{n_T} + \frac{p_C(1-p_C)}{n_C}} = \sqrt{\frac{168\times 32}{200^3} + \frac{172\times 28}{200^3}} = 0.0356.$$

Substitution gives the 90% confidence interval for the difference for $p_T - p_C$ to be $-\frac{4}{200} \pm 1.65 \times 0.356$.

Hence, the 90% confidence interval is –7.9% to 3.9%. From the question the range of equivalence is $(-5\%, 5\%)$. Since the confidence interval does not lie within the range of equivalence there is insufficient evidence to conclude the two treatments are equivalent at a 5% significance level.

3. A randomised controlled non-inferiority trial compared cognitive behavioural therapy delivered by telephone with that delivered face-to-face for patients with obsessive compulsive disorder. The trial was testing whether telephone treatment was non-inferior to stnadrd face-to-face treatment. The outcome measure was the Yale-Brown Obsessive Compulsive Scale (YBOCS) was measured at 3 months after the completion of treatment. The limit for non-inferiority was 3 units on the YBOCS scale. Using the information in the print-out below test whether telephone treatment could be considered to be non-inferior to face-to-face treatment using

   (i) a 2.5% significance level.

   (ii) a 5% significance level.

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Telephone|      34    12.64706    1.280515    7.466622    10.04183    15.25229
Face-to-F|      29    12.93103    1.428019    7.690119    10.00587     15.8562
---------+--------------------------------------------------------------------
combined |      63    12.77778    .9461823    7.510089    10.88639    14.66917
---------+--------------------------------------------------------------------
    diff |           -.2839757    1.913504               -4.110264    3.542313
------------------------------------------------------------------------------
    diff = mean(Telephone) - mean(Face-to-Face)                   t =  -0.1484
Ho: diff = 0                                     degrees of freedom =       61

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.4413         Pr(|T| > |t|) = 0.8825          Pr(T > t) = 0.5587
```

Solution

In the question a limit of non-inferiority of 3 units has been suggested for the YBOC scale. The question states that the benefit of treatment is a reduction in the YBOCS. Hence, one requires the $(1-\alpha)$ one-sided confidence interval for the difference between *telephone* treatment and *face-to-face* treatment to be less than 3 for a significance level α. Formally, if τ is the treatment effect of telephone compared to face-to-face the hypotheses are $H_0 : \tau \geq \tau_N$ vs $H_1 : \tau < -\tau_N$. The upper limit of the usual (1-2α) two-sided significant level can therefore be considered.

(i) The print-out gives a two-sided 95% confidence interval for telephone as compared to face-to-face as (-4.110264 to 3.542313). Since the upper limit of the difference is above 3 units one cannot reject the null hypothesis at a 2.5% level.

(ii) From the notes the one-sided upper confidence interval for a α% level test is given by $\overline{Y}_T - \overline{Y}_C + t_\alpha(\nu) SE\left[\overline{Y}_T - \overline{Y}_C\right]$. From the output $\overline{Y}_T - \overline{Y}_C = -.2839757$ and $SE\left[\overline{Y}_T - \overline{Y}_C\right] = 1.913504$.

Degrees of freedom (ν) = 61. Tables don't give $\nu = 61$, but do give 60 so we can use this value as a good approximation (it will give a trivially wider confidence interval). $t_{0.05}(60) = 1.67$.

The one-sided upper confidence interval is

$\overline{Y}_T - \overline{Y}_C + t_\alpha(\nu) SE\left[\overline{Y}_T - \overline{Y}_C\right] = -.2839757 + 1.67 \times 1.913504 = 2.911$. Since this is less than 3, the null hypothesis of inferiority of telephone treatment can be rejected.

4. Consider a continuous and normally distributed outcome measure *Y* for which lower scores correspond to a better outcome. It is suggested in the notes that the null hypothesis $H_0 : \tau \geq \tau_N$ in a non-inferiority trial is rejected if the upper $(1-\alpha)$ single-sided confidence interval is in the region $(-\infty, +\tau_N)$.

   (i) Assuming that the variance σ^2 is known, derive an expression for probability of rejection of H_0.

Solution

The upper confidence interval is used defined by $\overline{Y}_T - \overline{Y}_C + t_\alpha(\nu) SE\left[\overline{Y}_T - \overline{Y}_C\right]$

H_0 is rejected if $\overline{Y}_T - \overline{Y}_C + t_\alpha(\nu) SE\left[\overline{Y}_T - \overline{Y}_C\right] < \tau_N$.

Assuming that the variance σ^2 is known enables a normal approximation to the t-distribution, so we can replace $t_\alpha(\nu)$ by z_α. Now $SE\left[\overline{Y}_T - \overline{Y}_C\right] = \sigma\lambda$ where $\lambda = \sqrt{\frac{1}{n_T} + \frac{1}{n_C}}$ so that $\overline{Y}_T - \overline{Y}_C$ has a distribution $N\left[\tau, \sigma^2\lambda^2\right]$

$$\begin{aligned}\Pr\left[\text{Reject } H_0 | \ \tau\right] &= \Pr\left[\overline{Y}_T - \overline{Y}_C + z_\alpha\sigma\lambda < -\tau_N\right] \\ &= \Pr\left[\overline{Y}_T - \overline{Y}_C < \tau_N - z_\alpha\sigma\lambda\right]\end{aligned}$$

Since $\overline{Y}_T - \overline{Y}_C$ has the distribution $N\left[\tau, \sigma^2\lambda^2\right]$, it follows that $\Pr\left[\text{Reject } H_0 | \ \tau\right] = \Phi\left(\frac{\tau_N - z_\alpha\sigma\lambda - \tau}{\sigma\lambda}\right)$.

   (ii) Show that the Type I error rate has a maximum when $\tau = \tau_N$.

Solution

We can find the maximum of this by differentiation with respect to τ. The derivative is

$$\frac{d}{d\tau}\Pr\left[\text{Reject } H_0 | \ \tau\right] = \frac{d}{d\tau}\ \Phi\left(\frac{\tau_N - z_\alpha\sigma\lambda - \tau}{\sigma\lambda}\right) = -\frac{1}{\sigma\lambda}\phi\left(\frac{-\tau_N + z_\alpha\sigma\lambda - \tau}{\sigma\lambda}\right),$$

where ϕ is the standard normal density. Since ϕ is greater than zero for finite values, it follows that $\Pr\left[\text{Reject } H_0 | \ \tau\right]$ is monotone decreasing for τ. Hence, the type I error rate has a maximum when τ has a minimum under the null, which is $\tau = \tau_N$.

   (iii) Show that the Type I error is $\leq \alpha$.

Solution

At the maximum $\tau = \tau_N$

$$\Pr\left[\text{Reject } H_0 | \ \tau\right] = \Phi\left(\frac{\tau_N - z_\alpha\sigma\lambda - \tau_N}{\sigma\lambda}\right) = \Phi\left(-z_\alpha\right) = \alpha$$

$$\Pr\left[\text{Reject } H_0 | \ -\tau_N\right] = 1 - \Phi\left(\frac{-\tau_N + z_\alpha\sigma\lambda - \left(-\tau_N\right)}{\sigma\lambda}\right) = 1 - \Phi\left(\frac{z_\alpha\sigma\lambda}{\sigma\lambda}\right) = 1 - \Phi\left(z_\alpha\right) = \alpha .$$

5. Consider a continuous and normally distributed outcome measure *Y* for which lower scores correspond to a better outcome. Suppose the null hypothesis $H_0 : \tau \geq \tau_N$ in a non-inferiority trial is rejected if the $(1-\alpha)$ single-sided confidence interval is in the region $(-\infty, +\tau_N)$.

(i) Suppose that $\tau = \tau_A$ under the alternative hypothesis. Write down an expression for the power of the trial.

Solution

From the previous question $\Pr\left[\text{Reject } H_0 |\ \tau\right] = \Phi\left(\frac{\tau_N - z_\alpha \sigma\lambda - \tau}{\sigma\lambda}\right)$

If $\tau = \tau_A$ under the alternative hypothesis, $(1\text{-}\beta) = \Pr\left[\text{Reject } H_0 | \tau_A\right] = \Phi\left(\frac{\tau_N - z_\alpha \sigma\lambda - \tau_A}{\sigma\lambda}\right)$

(ii) Show that the sample size per group required in a trial with two equal size treatment groups to test the hypothesis $H_0 : \tau \geq \tau_N$ vs $H_1 : \tau < \tau_N$ is $n = \frac{\sigma^2}{(\tau_N - \tau_A)^2}(z_\beta + z_\alpha)^2$,when $\tau = \tau_A$ under the alternative hypotheses.

Solution

$$(1\text{-}\beta) = \Phi\left(\frac{\tau_N - z_\alpha \sigma\lambda - \tau_A}{\sigma\lambda}\right)$$

Taking inverses

$$z_\beta = \frac{\tau_N - z_\alpha \sigma\lambda - \tau_A}{\sigma\lambda} = -z_\alpha + \frac{\tau_N - \tau_A}{\sigma\lambda}$$

Assuming two equal size group, say n, $\lambda = \sqrt{\frac{2}{n}}$.

Hence

$$z_\beta + z_\alpha = \frac{\tau_N - \tau_A}{\sigma}\sqrt{\frac{n}{2}}.$$

Rearranging gives $n = \frac{2\sigma^2}{(\tau_N - \tau_A)^2}(z_\beta + z_\alpha)^2$ as required.

(iii) Sketch a plot of the sample size per group (*n*) against $\tau_A \in (-\infty, \tau_N)$ marking the intercept with $\tau_A = 0$.

Solution Sketch plot of n against τ_A