MT30772 Exam Solution 2006-7

A1

(i) In the context of clinical trials briefly explain what is meant by the term bias.

(ii) Describe two possible sources of bias in clinical trials.

[5 Marks]

(i) In the context of a clinical trial bias is a factor that deviates the estimate of the treatment effect systematically away from the true estimate.
[1 mark]

(ii)_A brief description of any two of the following (i) selection (ii) allocation (iii) performance (iv) follow-up (v) outcome assessment or (vi) analyses biases

[4 Marks]

A2

(i) How many patients have been allocated to each treatment? [marks 1]

13 Acupuncture 12 Homeopathic patients

The characteristics of the next three patients to entering the trial are:

26^{th} (Male, Migraine)
27^{th} (Female, Migraine)
Determine the treatment allocation of each patient.

Table with updated totals

		Male		Female		Migraine		Tension		total		Allocated treatment
		(A)	(H)	(A)	(H)	(A)	(H)	(A)	(H)	(A)	(H)	
26	Male, Migraine	8	5	5	7	7	6	6	6			
27	Female, Migraine	8	**6**	5	7	7	**6**	6	6	8+7 =15	5+6 =11	(H)
28		8	7	**6**	7	**8**	7	6	7	5+7 =12	7+6 =13	(A)

The next two patients are allocated to Homeopathic , Acupuncture.

[Marks 4]

[total 5 marks]

A3 A randomised controlled equivalence trial is being planned to compare a new treatment (T) and with a control treatment (C)**.** Suppose *Y* is the continuous and normally distributed outcome measure and suppose we

consider the two treatments to be equivalent provided the 1-2α% confidence interval for $\overline{Y}_T - \overline{Y}_C$ is in the range [-δ,δ]. The power to demonstrate treatment equivalence is given by the expression $2\Phi\left(\frac{\delta\sqrt{n}}{\sigma\sqrt{2}} - z_\alpha\right) - 1$ where σ is the within group standard deviation, n is the sample size of each group, and Φ is the cumulative density function of a standardised normal distribution.

Suppose a range of equivalence [-2,2] is to be used and the within group standard deviation has been estimated to be 4 and α equal to *5%*.

(i) Estimate the power of the study 50 subjects in each treatment.

Power $= 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma\sqrt{2}} - z_\alpha\right) - 1$ *with* δ*=2* σ*=4 and* α*=0.05. From tables of the normal distribution* z_α*=1.645.*

Power $= = 2\Phi\left(\frac{2\times\sqrt{50}}{4\sqrt{2}} - 1.645\right) - 1 = 2\Phi\left(\frac{2\times\sqrt{50}}{4\sqrt{2}} - 1.645\right) - 1 = 2\Phi\left(0.855\right) - 1$

From tables $\Phi\left(0.855\right)$ *=0.803 Hence power =2x0.803-1=0.606*

[5 marks]

(ii) Determine the sample size required to obtain 95% power.

0.95= $2\Phi\left(\frac{\delta\sqrt{n}}{\sigma\sqrt{2}} - z_\alpha\right) - 1$

Substitution with δ*=2* σ*=4 and* α*=0.05 gives*

$$\Phi\left(\frac{2\sqrt{n}}{4\sqrt{2}} - 1.645\right) = \frac{1.95}{2} = 0.975$$

Therefore $\left(\frac{2\sqrt{n}}{4\sqrt{2}} - 1.645\right) = \Phi^{-1}\left(0.975\right)$

From tables $\Phi^{-1}\left(0.975\right) = 1.96$ *Therefore* $\sqrt{n} = 2\sqrt{2}\left(1.96 + 1.645\right)$ *giving n= 103.9.*

Therefore the minimum sample size is 104 in each group.
[5 marks]
[total 10]

A4

(i) Derive a test statistic to compare the treatment effect in patients with pain in both knees with the treatment effect for patients with pain in just one knee.

Let δ_A be the treatment effect patients with both knees affected and let δ_B the treatment effect patients with just one knees effected. To investigate a subgroup effect we are test the null hypothesis that

$H_0:\delta_A=\delta_B$ vs $H_1:\delta_A\neq\delta_B$

Let sample means for each subgroup and treatment combination be $\overline{Y}_{CA}, \overline{Y}_{TA}, \overline{Y}_{CB}$ and $\overline{Y}_{TB}$ with sample standard deviations S_{CA}, S_{TA}, S_{CB} , and S_{TB} and with sample sizes n_{CA}, n_{TA}, n_{CB} and n_{TB}.
$\hat{\delta}_A = \overline{Y}_{TA} - \overline{Y}_{CA}$ is an estimate of the treatment effect in sub-group A $\hat{\delta}_B = \overline{Y}_{TB} - \overline{Y}_{CB}$ is an estimate of the treatment effect in sub-group B

Since patients in each of the four subgroups are independent, it follows that $\overline{Y}_{CA}, \overline{Y}_{TA}, \overline{Y}_{CB}$ and $\overline{Y}_{TB}$ are independent. Hence

$$Cov\left[\overline{Y}_{ij}, \overline{Y}_{i'j'}\right] = 0 \quad \text{for } i = T,C \text{ and } j = A,B.$$

Hence

$$Var\left(\left(\overline{Y}_{TA} - \overline{Y}_{CA}\right) - \left(\overline{Y}_{TB} - \overline{Y}_{CB}\right)\right) = Var\left[\overline{Y}_{TA}\right] + Var\left[\overline{Y}_{CA}\right] + Var\left[\overline{Y}_{TB}\right] + Var\left[\overline{Y}_{CB}\right].$$

Since $\hat{Var}\left[\overline{Y}_{ij}\right] = \frac{S_{ij}^2}{n_{ij}}$ for $i = T,C$ and $j = A,B$,

it follows that

$$\hat{SE}\left(\left(\overline{Y}_{TA} - \overline{Y}_{CA}\right) - \left(\overline{Y}_{TB} - \overline{Y}_{CB}\right)\right) = \sqrt{\left(\frac{S_{TA}^2}{n_{TA}} + \frac{S_{CA}^2}{n_{CA}}\right) + \left(\frac{S_{TB}^2}{n_{TB}} + \frac{S_{CB}^2}{n_{CB}}\right)}.$$

It follows from the central limit theorem that $T_{AB} = \dfrac{\left(\overline{Y}_{TA} - \overline{Y}_{CA}\right) - \left(\overline{Y}_{TB} - \overline{Y}_{CB}\right)}{\sqrt{\left(\frac{S_{TA}^2}{n_{TA}} + \frac{S_{CA}^2}{n_{CA}}\right) + \left(\frac{S_{TB}^2}{n_{TB}} + \frac{S_{CB}^2}{n_{CB}}\right)}}$ will have a standardized normal distribution under the null hypothesis asymptotically.

Hence T_{AB} may be used to test $H_0:\delta_A=\delta_B$ vs $H_1:\delta_A\neq\delta_B$.
[5 marks]

(i)
$\overline{Y}_{CA}, \overline{Y}_{TA}, \overline{Y}_{CB}$ $\overline{Y}_{TB}$ S_{CA}, S_{TA}, S_{CB} , S_{TB} ,n_{CA}, n_{TA}, n_{CB} and n_{TB} from the above formula can be extracted from the computer output. Substitution in the formula give

$$T_{AB} = \frac{(-9) - (2)}{\sqrt{\left(\frac{15^2}{32} + \frac{18^2}{30}\right) + \left(\frac{16^2}{12} + \frac{17^2}{15}\right)}} = \frac{11}{7.644} = 1.439$$

Reference to tables of the normal distribution give a p=0.15 for a two-tailed test.
[4 marks]

(ii) From the computer output there is no evidence of an overall benefit. Nevertheless the researcher has concluded that the new treatment (T) is more effective than the standard treatment (C) in patients with pain in both knees. This is because a test of the treatment effect is significant at a 5% level for the both knees analysis and not for just one knee. This analysis and the conclusion that is drawn are flawed for two reasons. Firstly failure to reject the null hypothesis in just one knee does not imply that there is no effect. Secondly separate statistical tests do not compare the magnitude of the treatment effect for each subgroup. The test of a difference in treatment effect for both knees as compare to just one knees does not support the investigators conclusion of a treatment effect in one sub-group but not the other. Therefore I disagree with the researcher. [5 marks]

[14 marks]

A5.

(i) 1. C must be a cause of D and

2. C must be correlated with E in the study sample. When there are just exposed and unexposed subjects, this means that the distribution of C is different in these two groups.

2 marks

(ii) The *stratification* method. In this approach the data are divided into a number of strata defined by the value of the confounder variable. The estimate of the effect of exposure on disease is estimated within each strata. (Usually a weighted average of the separate estimates is then calculated).

Method based on *regression* modelling. In this approach we set up a model for disease risk as a function of both exposure, E, and the confounder variable C, eg a logistic regression model. If the model is correct, then the β coefficient for the exposure variable is an estimate of its effect assuming that the confounder variable stays constant.

4 marks

A6

(i) Since A and B are indept,

$$\Pr(A=a, B=b\} = \Pr\{A=a\}\Pr(B=b\} = \frac{\exp(-\beta\lambda T_A)(\beta\lambda T_A)^a}{a!}\frac{\exp(-\lambda T_B)(\lambda T_B)^b}{b!} \quad (1).$$

The distribution of A+B is Poisson with mean $\beta\lambda T_A + \lambda T_B$, ie

$$\Pr\{A+B=a+b\} = \frac{\exp(-\beta\lambda T_A - \lambda T_B)(\beta\lambda T_A + \lambda T_B)^{a+b}}{(a+b)!} \quad (2).$$

Also $\Pr\{A=a \mid A=B=a+b\} = \frac{\Pr\{A=a\}\Pr\{B=b\}}{\Pr\{A+B=a+b\}}$. Substituting from (1) and (2), this becomes:

$$\frac{\exp(-\beta\lambda T_A - \lambda T_B)(\beta\lambda T_A)^a(\lambda T_B)^b(a+b)!}{a!b!\exp(-\beta\lambda T_A - \lambda T_B)(\beta\lambda T_A + \lambda T_B)^{a+b}} = \frac{(a+b)!}{a!b!}\left(\frac{\beta\lambda T_A}{\beta\lambda T_A + \lambda T_B}\right)^a\left(\frac{\lambda T_B}{\beta\lambda T_A + \lambda T_B}\right)^b$$

Cancelling out λ, we see that this last expression describes a Binomial distn with parameters a+b and $\frac{\beta T_A}{\beta T_A + T_B}$.

[5 marks]

(ii) Under H_0, the binomial probability becomes $\frac{1*0.5T_B}{1*0.5T_B + T_B} = \frac{1}{3}$, so distn is Binomial with parameters 8 and 1/3.

[2 marks]

(iiii) To test H_0, we need to evaluate $\Pr\{A \geq 5 | A+B=8, H_0\} =$

$$\sum_{a=6}^{a=8} \binom{8}{a} \left(\frac{1}{3}\right)^a \left(\frac{2}{3}\right)^{8-b} = \binom{8}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^2 + \binom{8}{7}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^1 + \binom{8}{8}\left(\frac{1}{3}\right)^8$$

$= (28*2^2 + 8*2+1)/3^8 = 129/6561 = 0.01966$. Since $0.01966 < 0.05$, reject H_0.

[3 marks]

SECTION B

B7

For a parallel group trial comparing a control treatment (C) with a new intervention (T) suppose *y* is a continuous, normally distributed outcome variable and *x* is the value of the same variable recorded at baseline prior to randomisation. Suppose that δ is the treatment effect such that:

$y = \mu_y + \varepsilon_y$ and $x = \mu_x + \varepsilon_x$ for treatment C

$y = \mu_y + \delta + \varepsilon_y$ and $x = \mu_x + \varepsilon_x$ for treatment T.

with $E\left[\varepsilon_x\right] = E\left[\varepsilon_y\right] = 0$, $Var\left[\varepsilon_y\right] = \sigma_y^2$, $Var\left[\varepsilon_x\right] = \sigma_x^2$, and $Cov\left[\varepsilon_x, \varepsilon_y\right] = \sigma_{xy}$

Let $d = y - x$ with $\bar{d}_C$ and $\bar{d}_T$ the sample means of treatment C and treatment T respectively.

Define $\hat{\delta}(\theta) = (\bar{y}_T - \theta\bar{x}_T) - (\bar{y}_C - \theta\bar{x}_C)$.

(i) Show that

$E\left[\hat{\delta}(\theta)\right] = \delta$

Define $\hat{\delta}(\theta) = (\bar{y}_T - \theta\bar{x}_T) - (\bar{y}_C - \theta\bar{x}_C)$.

Values of θ *equal to 0 , 1 and* β *correspond to the treatment effect in an unadjusted, change and adjusted model analyses.*

Now $E\left[\hat{\delta}(\theta)\right] = E\left[\bar{y}_T - \bar{y}_C\right] - \theta E\left[\bar{x}_T - \bar{x}_C\right]$

Since $E\left[\bar{y}_T - \bar{y}_C\right] = \tau + \beta E\left[\bar{x}_T - \bar{x}_C\right]$, *it follows that*

$$E\left[\hat{\delta}(\theta)\right] = \delta + (\beta - \vartheta) E\left[\bar{x}_T - \bar{x}_C\right]$$

Randomisation means that $E\left[\bar{X}_T\right] = E\left[\bar{X}_C\right]$.

Therefore $E\left[\hat{\delta}(\theta)\right] = \delta$.

[4 marks]

Show that $Var\left[\hat{\delta}(\theta)\right] = \lambda^2\left(\sigma_y^2 + \theta^2\sigma_x^2 - 2\theta\sigma_{xy}\right)$ where $\lambda = \sqrt{\frac{1}{n_T} + \frac{1}{n_C}}$

$Var\left[\hat{\delta}(\theta)\right] = Var\left[\bar{y}_T - \bar{y}_C - \theta\left(\bar{x}_T - \bar{x}_C\right)\right]$

$= Var\left[\bar{y}_T - \bar{y}_C\right] + Var\left[\theta\left(\bar{x}_T - \bar{x}_C\right)\right] + 2Cov\left[\bar{y}_T - \bar{y}_C, \theta\left(\bar{x}_T - \bar{x}_C\right)\right].$

Now $Var\left[\bar{y}_T - \bar{y}_C\right] = Var\left[\bar{y}_T\right] + Var\left[\bar{y}_C\right]$ *since groups are independent.*

Since $Var\left[\bar{y}_T\right] = Var\left[y_T\right]/n_T$ *and* $Var\left[\bar{y}_C\right] = Var\left[y_C\right]/n_C$ *, it follows that* $Var\left[\bar{y}_T - \bar{y}_C\right] = \lambda^2\sigma_y^2$ *(1)*

where $\lambda = \sqrt{\frac{1}{n_T} + \frac{1}{n_C}}$.

Similarly $Var\left[\bar{x}_T - \bar{x}_C\right] = \lambda^2\sigma_x^2$ *.(2)*

Now $Cov\left[\bar{y}_T - \bar{y}_C, \bar{x}_T - \bar{x}_C\right] = Cov\left[\bar{y}_T, \bar{x}_T\right] + Cov\left[\bar{y}_C, \bar{x}_C\right].$

Since $Cov\left[\bar{y}_T, \bar{x}_T\right] = \frac{1}{n_T}\sigma_{xy}$ *and* $Cov\left[\bar{y}_C, \bar{x}_C\right] = \frac{1}{n_C}\sigma_{xy}$,

it follows also that

Now $Cov\left[\bar{y}_T - \bar{y}_C, \bar{x}_T - \bar{x}_C\right] = \lambda^2\sigma_{xy}$ *.(3)*

Combining (1) (2) and (3) gives

$Var\left[\hat{\delta}\left(\theta\right)\right] = \lambda^2\left(\sigma_y^2 + \theta^2\sigma_x^2 - 2\theta\sigma_{xy}\right)$

[7 marks]

Show that this has a minimum when θ =β

Differentiation with respect to θ *gives*

$$\frac{\partial}{\partial\theta}Var\left[\hat{\delta}\left(\theta\right)\right] = \lambda^2\left(2\theta\sigma_x^2 - 2\sigma_{xy}\right).$$

This equals zero when $\theta = \sigma_{xy}/\sigma_x^2$.

The second derivative $\frac{\partial^2}{\partial\theta^2}Var\left[\hat{\delta}\left(\theta\right)\right] = 2\lambda^2\sigma_x^2$.

As this is positive , it follows that $Var\left[\hat{\delta}\left(\theta\right)\right]$ *has a minimum when* $\theta = \sigma_{xy}/\sigma_x^2$.

[5 marks]

Comment on the implications of these results for the choice between an unadjusted comparison of the two treatments based on y , and analysis using the change score y-x and an adjusted analysis using a linear model of y with treatment group and x as covariates.

Values of θ *equal to 0 , 1 and* β *correspond to the treatment effect in an unadjusted, change and linear adjusted model analyses. All three estimates are unbiased, but an estimate of the treatment effect based on a linear model smaller variance compared to an unadjusted analysis or a change analysis. Reducing the variance of the treatment effect estimate increases the power of the analysis. As a*

consequence if a baseline variable is thought to be correlated with outcome, an analysis adjusting for baseline is recommended, and where the baseline value of the outcome is recorded a linear model analysis is superior to an analysis based on change.

[5 marks]

Suppose μ_y^C μ_y^T μ_x^C and μ_x^T are the mean for each treatment as baseline and outcome. One type of statistical analysis seen within the medical literature is test the hypothesis H_0 $\mu_y^C = \mu_x^C$ and H_0 $\mu_y^T = \mu_x^T$ using paired t-test. Change within treatment is interpreted as a treatment effect and rejection of the hypothesis for one treatment but not the other is interpreted as a difference between the two treatment. Give two reasons why this type of analysis is flawed.

This type of analysis is flawed because

- *Tests of within group change are often statistically significant, but change within a treatment group may not be due to treatment. It may occurs because the condition naturally resolves.A treatment efffect is define as a differeence between treatments not change within subjects. [Not a treatment affect]*
- *Failure to reject the null hypothesis, for a treatment does not imply no change. The change within groups can be the same but unequal variances or groups sizes affect the probability of rejecting the null hypothesis and this may differ between groups. [Type II Error]*
- *The two p-values relate to two separate hypotheses test and so do not directly tested the difference between groups. [Inappropriate hypothesis testing]*

[Two reasons for 4 marks]

B8.

For an AB/BA crossover trial a model for a continuous outcome y_{ij} of the i^{th} patient in the j^{th} period can be written as

$y_{i1} = \mu + \delta + \xi_i + \varepsilon_{i1}$ for a patient in sequence AB in period 1,
$y_{i2} = \mu + \phi + \xi_i + \varepsilon_{i2}$ for a patient in sequence AB in period 2,
$y_{i1} = \mu + \xi_i + \varepsilon_{i1}$ for a patient in sequence BA in period 1,
$y_{i2} = \mu + \delta + \phi + \xi_i + \varepsilon_{i2}$ for a patient in sequence BA in period 2.

where μ is the mean for the sequence BA in period 1 , δ is the treatment effect of A relative to B, γ is the carryover effect, ξ_i is a random variable representing patient *i* with mean zero and variance σ_B^2, and ε_{ij} is the error term for patient i in period *j* assumed to be normally distributed with mean zero and variance σ_ε^2.
Defining $d_i = y_{i2} - y_{i1}$ let $\bar{d}_{AB}$, μ_{AB}^D, $\bar{d}_{BA}$ and μ_{BA}^D be the sample and population means of these for sequences AB and BA respectively.

(i) Show that a test of the null hypothesis $H_0 : \mu_{AB}^d = \mu_{BA}^d$ is the same as a test of no treatment effect,

$H_0 : \delta = 0$

$\mu_{AB}^d = E\left[y_{i2} - y_{i1}\right] = E\left[\left(\mu + \phi + \varepsilon_{i2}\right) - \left(\mu + \delta + \varepsilon_{i1}\right)\right] = \phi - \delta$ since $E\left[\varepsilon_{i2}\right] = E\left[\varepsilon_{i1}\right] = 0$
Similarly
$\mu_{BA}^d = E\left[y_{i2} - y_{i1}\right] = E\left[\left(\mu + \phi + \delta + \varepsilon_{i2}\right) - \left(\mu + \varepsilon_{i1}\right)\right] = \phi + \delta$
Hence $\mu_{BA}^d - \mu_{AB}^d = 2\delta$
$H_0 : \mu_{AB}^d = \mu_{BA}^d$ iff $H_0 : \delta = 0$.
[3 Marks]

(ii) Two anti-cholesterol lowering drugs were compared in a randomised controlled crossover trial. Ten patients were randomly allocated to sequence drug A then drug B and nine patients were randomly allocated to sequence drug B then drug A. The table below summarizes the sample mean and standard deviation for each sequence and period. interval of the treatment effect. Test the hypothesis $H_0 : \delta = 0$ and compute the 95% confidence interval and comment on the result.

	Period 1		*Period 2*		*Period 2 – Period 1*		
Sequence	*mean*	*s.d.*	*mean*	*s.d.*	*mean*	*s.d*	*n*
AB	6.42	0.81	6.01	0.72	-0.41	0.52	10
BA	6.23	0.63	6.11	0.71	-0.12	0.43	8

The hypothesis H_0: δ=0 vs. H_1: $\delta \neq 0$ can be tested using a two-sample t-test of the means of the differences $H_0 : \mu_{AB}^d = \mu_{BA}^d$. *The test statistic T_d is defined as*

$$T_d = \frac{\bar{d}_{BA} - \bar{d}_{AB}}{\hat{SE}\left[\bar{d}_{BA} - \bar{d}_{AB}\right]} \text{ where } \hat{SE}\left[\bar{d}_{BA} - \bar{d}_{AB}\right] = s_d\sqrt{\frac{1}{n_{AB}} + \frac{1}{n_{BA}}} \text{ and } s_d = \sqrt{\frac{(n_{AB}-1)s_{d_{AB}}^2 + (n_{BA}-1)s_{d_{BA}}^2}{n_{AB} + n_{BA} - 2}}.$$

Substidution gives $s_d = \sqrt{\frac{9\times 0.52^2 + 7\times 0.43^2}{16}} = 0.483$ *Hence* $\hat{SE}\left[\bar{d}_{BA} - \bar{d}_{AB}\right] = 0.483\sqrt{\frac{1}{10}+\frac{1}{8}} = 0.229$

Hence $T_d = \frac{\text{-.41-(-0.12)}}{0.229} = -1.266$ *Under the null hypothesis this will have a t distribution with 16 d.f.*

From tables p>0.05. Hence the null hypothesis is not reject.

A (1-α)-size confidence interval for the treatment effect δ is defined by

$$\frac{1}{2}\left(\bar{d}_{BA} - \bar{d}_{AB}\right) \pm \frac{1}{2} t_{\alpha/2}\left(n_1+n_2-2\right)\hat{SE}\left[\bar{d}_{BA} - \bar{d}_{AB}\right].$$

From tables $t_{\alpha/2}\left(n_1+n_2-2\right) = t_{0.05}\left(16\right) = 2.12$

Therefore the 95% confidence interval

$\frac{1}{2}\left(\bar{d}_{BA} - \bar{d}_{AB}\right) \pm \frac{1}{2} t_{\alpha/2}\left(n_1+n_2-2\right)\hat{SE}\left[\bar{d}_{BA} - \bar{d}_{AB}\right]$ *which is* $\frac{1}{2}\left(0.29\right) \pm \frac{1}{2} 2.12\times .229$ *giving*

[-0.388 , 0.098] [-0.39 , 0.10]

Based on the hypothesis test there is no evidence that there is any difference between the two treatment. The 95% confidence interval is [-0.388 , 0.098]

[7 marks]

(iii) Define $c_i = y_{i1} - y_{i2}$ for sequence AB and $c_i = y_{i2} - y_{i1}$ for sequence BA. Let μ^c_{AB} μ^c_{BA}, $\bar{c}_{AB}$ and $\bar{c}_{BA}$ be the population and sample means of these for sequences AB and BA respectively. Show that a test of the null hypothesis $H_0: \mu^c_{AB} = \mu^c_{BA}$ is the same as a test of the period effect, $H_0 : \phi = 0$.

$$\mu^c_{AB} = E\left[y_{i1} - y_{i2}\right] = E\left[\left(\mu+\delta+\varepsilon_{i2}\right) - \left(\mu+\phi+\varepsilon_{i1}\right)\right] = \delta - \phi$$

$$\mu^c_{BA} = E\left[y_{i2} - y_{i1}\right] = E\left[\left(\mu+\phi+\delta+\varepsilon_{i2}\right) - \left(\mu+\varepsilon_{i1}\right)\right] = \phi + \delta$$

Therefore $\mu^C_{BA} - \mu^C_{AB} = -2\phi$.

Hence the test $H_0: \mu^c_{AB} = \mu^c_{BA}$ *is equivalent to a test of the period effect* $H_0 : \phi = 0$.

[3 marks]

(iv) For data in the table above test the null hypothesis $H_0: \phi = 0$.

The hypothesis H_0: ϕ=0 vs. H_1: $\phi\neq 0$ can be tested using a two-sample t-test of the means of the differences $H_0 : \mu^c_{AB} = \mu^c_{BA}$. *An appropriate test statistic T_d is defined as*

$$T_C = \frac{\bar{c}_{BA} - \bar{c}_{AB}}{\hat{SE}\left[\bar{c}_{BA} - \bar{c}_{AB}\right]} \text{ where } \hat{SE}\left[\bar{c}_{BA} - \bar{c}_{AB}\right] = s_C\sqrt{\frac{1}{n_{AB}}+\frac{1}{n_{BA}}} \text{ and } s_C = \sqrt{\frac{\left(n_{AB}-1\right)s^2_{C_{AB}} + \left(n_{BA}-1\right)s^2_{C_{BA}}}{n_{AB}+n_{BA}-2}}$$

Where s_C is the pooled standard deviation and $s_{C_{AB}}, s_{C_{BA}}$ are the standard deviations for each sequence.

Now $\mu^c_{AB} = -\mu^d_{AB}$ *=0.41 and* $\mu^c_{BA} = \mu^d_{BA}$ *=-0.12*

$s_{C_{AB}} = s_{D_{AB}} =, s_{C_{BA}} = s_{D_{AB}}$ *Therefore* $\hat{SE}\left[\bar{c}_{BA} - \bar{c}_{AB}\right] = \hat{SE}\left[\bar{d}_{BA} - \bar{d}_{AB}\right] = 0.229$

Therefore $T_C = \dfrac{0.41 + 0.12}{0.229} = 2.31$

From tables $t_{\alpha/2}\left(n_1+n_2-2\right) = t_{0.05}\left(16\right) = 2.12$ *Therefore one can reject the null hypothesis that* H_0*:* $\phi=0$

[mark 5]

(v) Briefly comment on the result of the trial

From the test of the hypothesis $\boldsymbol{H_0: \delta = 0}$ ***t****here is no evidence of a treatment effect. In contrast there is evidence of a period effect. From inspection of the data in the table one can see that cholesterol levels reduce for both sequences.*

[2 marks]

(vi) It is sometimes suggested that the treatment effect δ can be estimated the overall sample mean of the differences c_i say $\bar{c} = \dfrac{\sum_{i=1}^{N} c_i}{N}$, where *N* is the total number of subjects in the trial. Show $\bar{c}$ may be biased estimate of δ. In what circumstances is $\bar{c}$ an unbiased estimate of δ ?

[5 marks]

Considering $\bar{c}$ $E[\bar{c}\,] = \Sigma E[c_i]/N$

$E\left[c_i\right] = -\phi + \delta$ *for sequence AB and* $E\left[c_i\right] = \phi + \delta$ *for sequence BA. Therefore*

$E[\bar{c}\,] = (N_{AB}(-\phi+\delta) + N_{BA}(\phi+\delta))/N = \delta + \phi\,(N_{BA} - N_{AB})/N.$

Hence $\bar{c}$ *is a biased estimator of* δ *if numbers allocated to each sequence* $N_{\boldsymbol{BA}}$ *&* N_{AB} *are unequal and there is a period effect. If the numbers in each sequence are equal or there is no period effect the estimator will be unbiased.*

[5 marks]

B9

(i) $\frac{d}{d\pi}\left(\ln\frac{\nu}{1-\nu}\right)=\frac{(1-\nu)}{\nu}.\frac{1}{(1-\nu)^2}$ *and* $Var(\nu)=\frac{\nu(1-\nu)}{n}$.

Then using approximation,

$$Var\left(\ln\frac{\pi}{(1-\pi)}\right)\approx\left[\frac{d}{d\pi}\left(\ln\frac{\pi}{1-\pi}\right)\right]^2 Var(\pi)=\frac{1}{n\pi(1-\pi)}=\frac{1}{n\pi}+\frac{1}{n(1-\pi)}.$$

Hence $Var\left(\ln\left[\frac{\pi_A(1-\pi_B)}{(1-\pi_A)\pi_B}\right]\right)=Var\left(\ln\frac{\pi_A}{1-\pi_A}\right)+Var\left(\ln\frac{\pi_B}{1-\pi_B}\right)=\frac{1}{n_A\pi_A}+\frac{1}{n_A(1-\pi_A)}+\frac{1}{n_B\pi_B}+\frac{1}{n_B(1-\pi_B)}$

7 **marks**

(ii) Assuming that $\ln(\hat{\gamma})\sim Normal(\ln\gamma, Var(\ln\hat{\gamma}))$.

A 95% CI for ln(γ) can be found using $\ln(\hat{\gamma})\pm 1.96*\sqrt{Var(\hat{\gamma})}$.

From the data $\hat{\pi}_A=10/100=0.1,\ n_A\hat{\pi}_A=10, n_A(1-\hat{\pi}_A)=90$ and

$$\hat{\pi}_B=12/250=0.048,\ n_B\hat{\pi}_B=12, n_B(1-\hat{\pi}_B)=238.$$

The point estimate $\hat{\gamma}$ of γ is obtained from the cross-product ratio $\frac{10.238}{90*12}=2.204$ and ln (2.204)=0.7901.

$V[\ln(\hat{\gamma})]$ is estimated by substituting estimates of π_A etc into the formula in (i) giving

$$\sqrt{Var(\hat{\gamma})}=\sqrt{\frac{1}{10}+\frac{1}{90}+\frac{1}{12}+\frac{1}{238}}=0.4457.$$

Approx 95% CI for lnγ is therefore 0.7901 ± 1.96*0.4457, ie (-0.0835, 1.6637)

and 95% CI for γ is exp(-0.0835, 1.6637) = (0.92, 5.28). 7 **marks**

(iii) $\Pr(E\cap D\}=\Pr\{E\}*\Pr\{D|E\}=q\pi_A$, etc. Hence :

	D	**not D**
E	$q\pi_A$	$q(1-\pi_A)$
not E	$(1-q)\pi_B$	$(1-q)(1-\pi_B)$

$$\nu_1=\Pr\{E\mid D\}=\frac{\Pr\{E\cap D\}}{\Pr\{E\}}=\frac{q\pi_A}{q\pi_A+(1-q)\pi_B}.$$

Similarly, $\nu_0=\Pr\{E\mid notD\}=\frac{q(1-\pi_A)}{q(1-\pi_A)+(1-q)(1-\pi_B)}.$

From these:

$$\frac{\nu_1}{1-\nu_1} = \frac{q\pi_A}{(1-q)\pi_B} \text{ and } \frac{\nu_0}{1-\nu_0} = \frac{q(1-\pi_A)}{(1-q)(1-\pi_B)},$$

hence $$\frac{\nu_1/(1-\nu_1)}{\nu_0/(1-\nu_0)} = \frac{\pi_A/\pi_B}{(1-\pi_A)/(1-\pi_B)} = \frac{\pi_A(1-\pi_B)}{(1-\pi_A)\pi_B} = \gamma$$

7 marks

(iv) We could choose indept random samples of people with the disease and people without disease: from the former one could estimate ν_1 and from the latter ν_0 and hence γ could be estimated using the formula in (iii).

4 marks