Statistical tables are attached

Two Hours

UNIVERSITY OF MANCHESTER

May 2007 Final Draft

Medical Statistics MT3772

Electronic calculators may be used provided that they cannot store text

---

Answer **ALL** six questions in **SECTION A** (50 Marks)

Answer **TWO** of the three questions in **SECTION B** (25 marks each)

The total number of marks on the paper is 100

SECTION A

Answer **ALL** six questions

SECTION A

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]

A2.

In a trial comparing an Acupuncture treatment (A) with a Homeopathic treatment (H) for patient suffering from chronic headaches , patients are allocated to treatment using *deterministic minimization* controlling for sex and type of head ache (migraine, tension). The numbers of patients with each characteristic for each treatment are given in the table below after twenty-five patients have entered the trial.

Patient Characteristic	Male		Female		Migraine		Tension	
Treatment	(A)	(H)	(A)	(H)	(A)	(H)	(A)	(H)
Number of Patients	8	5	5	7	7	6	6	6

(i) How many patients have been allocated to each treatment?

(ii) The characteristics of the next two patients to entering the trial are:

26^{th} (Male, Migraine)

27^{th} (Female, Migraine)

Determine the treatment allocation of each patient.

[5 marks]

A3 A randomised controlled equivalence trial is being planned to compare a new treatment (T) 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)100 % confidence interval for $\bar{y}_T - \bar{y}_C$ 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 treatment group standard deviation, *n* is the sample size of both groups, and Φ is the cumulative distribution function of a standardised normal distribution.

Suppose [-2,2] is to be used as the range of equivalence and the within treatment group standard deviation has been estimated to be 4 .

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

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

[10 Marks]

A4. A researcher has carried out a randomized clinical trial to compare a new treatments (T) for chronic knee pain with a standard treatment (C). The two treatments are randomly allocated to 89 patients including 62 with pain in both knees and 27 with pain in just one knee. After three months follow-up patient's pain is measured using a 100mm visual analogue scale with higher scores representing greater pain. The researcher has carried out the following statistical analysis using a statistical software package, first analyzing all patients and then patients with pain in just one knee and both knee separately.

All subjects

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean      Std. Dev.         [95% Conf. Interval]
---------+--------------------------------------------------------------------
       T |      44          29       16               24.13556    33.86444
       C |      45          34       17               28.89263    39.10737
---------+--------------------------------------------------------------------
    diff |                  -5                        -11.95869    1.958691
------------------------------------------------------------------------------
diff = mean(T) - mean(C)  T-Test of difference = 0 (vs not =):
T-Value = -1.43  P-Value = 0.157  DF = 87
```

One knee

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean      Std. Dev.         [95% Conf. Interval]
---------+--------------------------------------------------------------------
       T |      12          36       16               25.83409    46.16591
       C |      15          34       17               24.58571    43.41429
---------+--------------------------------------------------------------------
    diff |                   2                        -11.21511    15.21511
------------------------------------------------------------------------------
diff = mean(T) - mean(C)    T-Test of difference = 0 (vs not =):
T-Value = 0.312  P-Value = 0.758 DF = 25
```

Both knees

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean      Std. Dev.         [95% Conf. Interval]
---------+--------------------------------------------------------------------
       T |      32          26     15                20.59192    31.40808
       C |      30          35     18                28.27869    41.72131
---------+--------------------------------------------------------------------
    diff |                  -9                        -17.39686   -.6031408
------------------------------------------------------------------------------
diff = mean(T) - mean(C)     T-Test of difference = 0 (vs not =):
T-Value = -2.1440    P-Value = 0.036  DF = 60
```

Using a 5% significance level, the researcher concludes from the computer output that the new treatment (T) is more effective than the standard treatment (C) in patients with pain in both knees but no better in patients with pain in just one knee.

(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.

(ii) Apply this statistical test to the data to compare the treatment effect in patients with pain in both knees with the treatment effect for patients with pain in just one knee.

(iii) Comment on the researcher's analysis and conclusions. Do you agree with them?

[14 Marks]

A5

(i) A study measures the association between an exposure E and a disease D by comparing disease rates in exposed and unexposed subjects. It is claimed that this association is biased because of a confounding factor, C. State two conditions which must be satisfied by C for C to be a confounder.

(ii) Briefly describe the principles underlying two statistical approaches for removing bias due to a confounder.

[6 marks]

A6

Samples of exposed and unexposed subjects are chosen independently and followed over time to monitor new cases of disease. Let A and B be random variables describing the number of new cases of disease seen in exposed and unexposed respectively, T_A and T_B be the corresponding total person- time of observation in each group and let $\beta\lambda$ and λ be the incidence density rates in exposed and unexposed populations.

(i) Assume that A and B have independent Poisson distributions with means $\beta\lambda T_A$ and λT_B respectively. Show that the *conditional* distribution of A, given A+B=a+b, is Binomial with parameters (a+b) and $\frac{\beta T_A}{\beta T_A + T_B}$, using the fact that the sum of two Poisson variables also has a Poisson distribution.

(ii) Suppose that $T_A = 0.5T_B$. What is the condition distribution of A given A+B=8 under H_0:β=1?

(iii) Suppose that a= 6, b=2. Test H_0:β=1 versus H_1:β>1 using a test of size 0.05.

[10 marks]

SECTION B

B7

For a parallel group trial comparing a control treatment (C) with a new treatment (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

$E\left[\hat{\delta}(\theta)\right] = \delta$.**[4 marks]**

(ii) 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 $\lambda = \sqrt{\frac{1}{n_T} + \frac{1}{n_C}}$, n_T is the numbers of patient allocated to the new treatment and n_C is the number allocated to the control treatment. **[7 marks]**

(iv) Show that this has a minimum when $\theta = \sigma_{xy} / \sigma_x^2$ .. **[5 marks]**

(iv) 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. **[5 marks]**

(v) 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 to test the hypothesis H_0 $\mu_y^C = \mu_x^C$ and H_0 $\mu_y^T = \mu_x^T$ using paired t-test. 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. **[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, 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, $H_0 : \delta = 0$ **[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 eight 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. **[7 marks]**

	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

(iii) Define $c_i = y_{i1} - y_{i2}$ for sequence AB and $c_i = y_{i2} - y_{i1}$ for sequence BA. Let μ_{AB}^c μ_{BA}^c, $\bar{c}_{AB}$ 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 $H_0 : \mu_{AB}^c = \mu_{BA}^c$ is the same as a test of the period effect, $H_0 : \phi = 0$. **[3 marks]**

(iv) For data in the table above test the null hypothesis $H_0 : \phi = 0$. **[5 marks]**

(v) Briefly comment on the result of the trial **[2 marks]**

(vi) It is sometimes suggested that the treatment effect δ can be estimated by 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 that $\bar{c}$ may be a biased estimate of δ. **[5 marks]**

B9 (i) To investigate whether an exposure, E, is a cause of a disease D, random samples of exposed and unexposed men, of size n_A and n_B respectively, are chosen from a population. The cumulative incidence of disease in each group over a 10 year period is measured. Assume that the numbers of subjects in each group who develop the disease have independent Binomial distributions with probability parameters π_A and π_B respectively and consider the odds ratio, $\gamma = \dfrac{\pi_A(1-\pi_B)}{(1-\pi_A)\pi_B}$.

Using the approximate relationship $Var[y] \approx \left(\dfrac{dy}{dx}\right)^2_{x=E[x]} Var[x]$, show that

$$Var[\ln \hat{\gamma}] \approx \frac{1}{n_A\pi_A} + \frac{1}{n_A(1-\pi_A)} + \frac{1}{n_B\pi_B} + \frac{1}{n_B(1-\pi_B)}$$

(ii) The following data are obtained:

	Disease	*No Disease*	*Sub-total*
Exposed	10	90	100
Unexposed	12	238	150

Use this formula to calculate an approximate 95% Confidence Interval for γ. You may assume that $\ln(\hat{\gamma})$ is approximately Normally distributed with mean ln(γ).

(iii) Consider a different population with the same disease probabilities, π_A and π_B. A proportion, q, of this population is exposed to E. Give expressions for the proportions in the population in the four cells formed by cross-classifying E and D. Use these to show that $\nu_1 = \Pr\{E \mid D\} = \dfrac{q\pi_1}{q\pi_1 + (1-q)\pi_0}$ and derive a similar expression for

$\nu_0 = \Pr\{E \mid no\, D\}$. Hence show that $\dfrac{\nu_1(1-\nu_0)}{(1-\nu_0)\nu_0} = \gamma$.

(iv) On the basis of the result in (iii) suggest an alternative study design to that in (i) for estimating γ.