



The
University
Of
Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS

**Autumn Semester
2015–16**

Medical Statistics

2 hours

*Candidates may bring to the examination a calculator that conforms to University regulations. All questions will be marked, but credit will be given for only the best **THREE** answers. All questions carry equal marks. Total marks 60.*

**Please leave this exam paper on your desk
Do not remove it from the hall**

Registration number from U-Card (9 digits)
to be completed by student

--	--	--	--	--	--	--	--	--

Blank

- 1 Withdrawals can cause problems in clinical trials.
- (i) Using a ‘Worst Case Scenario’ (WCS) approach to attribution of unknown outcomes where appropriate, provide both

- (a) per protocol *(4 marks)*
and
- (b) intention to treat *(4 marks)*

analyses of the following (artificial) data from a trial comparing Drug with Placebo?

Assessment	Drug	Placebo	TOTAL
Success	20	20	40
Failure	4	16	20
Total Assessed	24	36	60
Withdrawn	16	4	20
Total Randomized	40	40	80

- (c) Explain why the WCS approach is not easy when the primary endpoint is quantitative. *(2 marks)*
- (d) Which analysis do you feel is more appropriate and why? *(2 marks)*
- (e) What is your overall conclusion about the value of the Drug? *(2 marks)*
- (ii) If the trial coordinator adopted conventional values for many of the parameters in his sample size calculations, namely
 20% allowance for dropouts
 5% probability of a type I error
 90% power
 and the success rate on Placebo was only anticipated to be 50%, what Clinically Relevant Difference was the trial sized to detect? *(6 marks)*

- 2 A randomized, double-blind comparative trial of drugs A and B is to be conducted. A statistician proposes that the small (16 patient) study might be conducted according to the following schematic diagram

Drug A	Drug B
before after	before after
$X \longleftrightarrow X$	$X \longleftrightarrow X$
$X \longleftrightarrow X$	$X \longleftrightarrow X$
\vdots	\vdots
$X \longleftrightarrow X$	$X \longleftrightarrow X$

where ‘X’ indicates that the quantitative response is measured and measurements on the same patient are linked. Thus measurements are taken both before and after application of the drug.

2 (continued)

The following summary statistics for the original responses (A before, A after, B before, B after), and some derived variables, are available:

	n	mean	s.d.
A before	8	96.1	17.53
A after	8	92.4	17.14
B before	8	89.5	17.63
B after	8	91.1	17.26
A after - A before	8	-3.8	2.19
B after - B before	8	1.6	4.34
B before - A before	8	-6.6	5.45
B after - A after	8	-1.3	5.01

- (i) Suggest how you might test whether or not the drugs had the same effect using the summary statistics available. *(4 marks)*
- (ii) State the distributional assumptions underpinning your suggestion in part (i). *(2 marks)*
- (iii) Assuming the assumptions are valid, use appropriate values from the summary statistics given to perform the test you suggested in (i). *(3 marks)*
- (iv) Explain why such a study might be preferable to the simpler trial (of the same size) represented in the following schematic diagram. *(2 marks)*

Drug A	Drug B
X	X
X	X
\vdots	\vdots
X	X

- (v) The 16 patients are required to be allocated randomly to the two groups. Explain the benefits of a blocked allocation over a simple random allocation. *(2 marks)*
- (vi) Suppose the effects of the drugs are thought to depend on the sex (Male/Female) and smoking status (Non-smoker/Smoker) of the patients.
 - (a) Explain how this would affect the analysis you would advise for the trial. *(2 marks)*
 - (b) Explain how this would affect the randomization you would advise. *(1 mark)*

2 (continued)

- (c) If the sex and smoking status of the patients are as given below, use the following random digits
 3 0 4 5 8 4 9 2 0 7 6 2 3 5 8 4 1 5 3 2
 to allocate the 16 patients to the two drugs. Explain your procedure clearly. (4 marks)

Patient number	Sex	Smoking status
1	Male	Smoker
2	Male	Smoker
3	Female	Smoker
4	Male	Non-smoker
5	Male	Smoker
6	Male	Non-smoker
7	Male	Smoker
8	Female	Non-smoker
9	Male	Non-smoker
10	Female	Non-smoker
11	Male	Smoker
12	Female	Non-smoker
13	Male	Non-smoker
14	Male	Non-smoker
15	Male	Smoker
16	Male	Smoker

- 3 24 patients with multiple myeloma (a type of cancer arising from plasma cells) were allocated to one of two forms of treatment and followed up. 11 patients were given drug *A* and 13 patients drug *B*. The primary outcome of interest was mortality. The table below shows the data collected. Here the status variable records whether the patient was observed to die (status = 1) or whether they were lost to follow-up beforehand (status = 0).

	Drug A		Drug B	
	Time (mths)	Status	Time (mths)	Status
	18.70	1	8.90	0
	1.40	1	14.90	1
	0.40	1	14.80	0
	9.20	1	16.50	0
	12.40	1	14.10	1
	23.70	1	13.80	1
	12.10	1	12.80	1
	3.10	0	11.10	0
	26.40	0	19.50	1
	33.50	0	17.30	0
	6.50	1	9.30	1
			19.90	1
			10.50	1
Total	147.4	8	183.4	8

3 (continued)

Some R analysis is shown below:

```
> t1 <- c(18.7,1.4,0.4,9.2,12.4,23.7,12.1,3.1,26.4,33.5,6.5)
> t2 <- c(8.9,14.9,14.8,16.5,14.1,13.8,12.8,11.1,19.5,17.3,9.3,19.9,10.5)
> time <- c(t1, t2)
>
> c1 <- c(1,1,1,1,1,1,1,1,0,0,0,1)
> c2 <- c(0,1,0,0,1,1,1,0,1,0,1,1,1)
> status <- c(c1, c2)
>
> type <- rep(c("A", "B"), c(11, 13))
>
> Comp.sv <- Surv(time, status, type = "right")
```

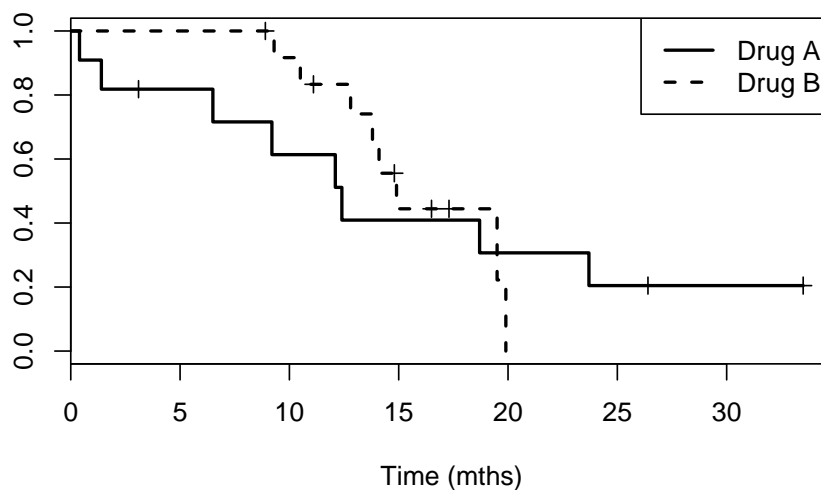
3 (continued)

```
> summary(survfit(Comp.sv ~ type))
Call: survfit(formula = Comp.sv ~ type)
```

type=A							
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI
0.4	11	1	0.909	0.0867	0.7541	1.000	
1.4	10	1	0.818	0.1163	0.6192	1.000	
6.5	8	1	0.716	0.1397	0.4884	1.000	
9.2	7	1	0.614	0.1526	0.3769	0.999	
12.1	6	1	0.511	0.1578	0.2793	0.936	
12.4	5	1	0.409	0.1559	0.1939	0.863	
18.7	4	1	0.307	0.1467	0.1202	0.783	
23.7	3	1	0.205	0.1286	0.0597	0.701	

type=B							
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI
9.3	12	1	0.917	0.0798	0.7729	1.000	
10.5	11	1	0.833	0.1076	0.6470	1.000	
12.8	9	1	0.741	0.1295	0.5259	1.000	
13.8	8	1	0.648	0.1426	0.4211	0.998	
14.1	7	1	0.556	0.1493	0.3281	0.941	
14.9	5	1	0.444	0.1554	0.2240	0.882	
19.5	2	1	0.222	0.1753	0.0474	1.000	
19.9	1	1	0.000	NaN	NA	NA	

```
> plot(survfit(Comp.sv ~ type), lty = c(1,2), lwd = 2, xlab = "Time (mths)")
```



- (i) Without making any model assumptions, estimate the median mortality times for the two drugs. *(3 marks)*

3 (continued)

(ii) It is suggested that the survival times are Exponentially distributed with rates λ_A and λ_B respectively. Under this assumptions:

(a) Estimate λ_A and λ_B and hence the mean mortality times with approximate 95% confidence intervals. *(4 marks)*

(b) Perform a likelihood ratio test to assess whether there is a difference in the mortality time distributions of the two drugs. *(4 marks)*

(iii) Use the partially complete R output (the ???s indicate output which is missing)

```
> survdiff(Comp.sv ~ as.factor(type))
```

Call:

```
survdiff(formula = Comp.sv ~ as.factor(type))
```

	N	Observed	Expected	???
as.factor(type)=A	11	8	7.55	???
as.factor(type)=B	13	8	8.45	???

to perform a non-parametric test assessing whether there is a difference in mortality between the two drugs. *(3 marks)*

(iv) Do the assumptions of an Exponential survival distribution seem plausible? Explain your answer. *(2 marks)*

(v) Discuss the impact on your analysis if:

(a) patients were withdrawn from the study if their condition was seen to be deteriorating or they showed side effects which needed alternative treatment. *(2 marks)*

(b) patients withdrew from the study because they were feeling better and no longer needed continuing treatment. *(2 marks)*

4 A clinical trial was conducted to compare two treatments for breast cancer among women. After removal of the tumour, patients were randomly allocated to either treatment A or treatment B and followed up for 5 years for cancer recurrence. The data are stored in `breast` and coding for the different variables is shown below:

Coding:

Treat: treatment (0 = treatment A; 1 = treatment B)

Obese: indicator if patient is obese (0 = non-obese; 1 = obese)

Age: age of patient centred on 50 years (i.e. 55 year old is +5)

Time: time until cancer recurrence (years)

Status: indicator of relapse (1) or censoring (0)

4 (continued)

(i) Some R analysis was performed with the edited output shown below:

```
> Breast.sv <- Surv(Time, Status)
>
> Br.fit <- coxph(Breast.sv ~ Obese+Treat+Age)
> summary(Br.fit)
Call:
coxph(formula = Breast.sv ~ Obese + Treat + Age)

n= 200, number of events= 170

              coef exp(coef) se(coef)      z Pr(>|z|)
Obese  0.221694  1.248189  0.155460  1.426 0.153855
Treat -0.537459  0.584231  0.158781 -3.385 0.000712 ***
Age    0.023620  1.023902  0.007944  2.973 0.002945 **
---

```

- (a) Specify the form of the model used for the above analysis in terms of the baseline hazard function $h_0(t)$ and the covariates. (3 marks)
- (b) Describe in detail the effects of the covariates on the time to recurrence. (4 marks)
- (c) Using the model output, calculate the estimate of the hazard ratio comparing
- An obese 56 year old woman on treatment A
 - A non-obese 43 year old woman on treatment B
- (3 marks)

(ii) An alternative analysis is shown below

```
> Br.fit1 <- survreg(Breast.sv ~ Obese+Treat+Age, dist="exponential")
> summary(Br.fit1)

Call:
survreg(formula = Breast.sv ~ Obese + Treat + Age, dist = "exponential")

              Value Std. Error      z      p
(Intercept)  0.6274    0.14495  4.33 0.000015
Obese        -0.2094    0.15447 -1.36 0.175219
Treat         0.5825    0.15563  3.74 0.000182
Age          -0.0251    0.00769 -3.27 0.001085

Scale fixed at 1

Exponential distribution
Loglik(model)= -266.1  Loglik(intercept only)= -281
Chisq= 29.91 on 3 degrees of freedom, p= 1.4e-06
Number of Newton-Raphson Iterations: 5
n= 200

```

4 (continued)

- (a) Specify what type of analysis has been performed here and write down the fitted model for T , the time to recurrence. *(4 marks)*
- (b) Comment on the effects of the different covariates on the time to recurrence with this approach. Would you prefer to be on treatment A or B? *(3 marks)*
- (c) Estimate the expected time to recurrence for a 56 year old obese woman who is taking treatment B. *(3 marks)*

End of Question Paper

STANDARD FORMULAE FOR MEDICAL STATISTICS (INCLUDING TABLES OF CRITICAL VALUES)

1 Clinical Trials Formulae

Two Sample t-Test — Separate variances form $r = \min(n_1, n_2)$

$$t_r = \left| \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \right|$$

Two Sample t-Test — Pooled variance form $r = n_1 + n_2 - 2$

$$t_r = \left| \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \right|$$

Sample Size Calculations — Two sample test for proportions NB number in each group

$$n \simeq \frac{\theta_2(1-\theta_2) + \theta_1(1-\theta_1)}{(\theta_2 - \theta_1)^2} [\Phi^{-1}(\beta) + \Phi^{-1}(\alpha/2)]^2$$

Sample Size Calculations — Two sample test for means NB number in each group

$$n \simeq \frac{2\sigma^2}{(\mu_2 - \mu_1)^2} [\Phi^{-1}(\beta) + \Phi^{-1}(\alpha/2)]^2$$

Standard Error for Natural Logarithm of Relative Risk

$$s.e.[(\log_e(RR))] = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$$

Standard Error for Natural Logarithm of Odds Ratio

$$s.e.[(\log_e(OR))] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

2 Survival Analysis Formulae

Exponential Distributions — MLE of rate λ with censoring The mle

$$\hat{\lambda} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i} = \frac{\Delta}{\mathcal{T}} \quad \text{var}(\hat{\lambda}) \approx \frac{\hat{\lambda}^2}{\sum_{i=1}^n \delta_i}.$$

For any (differentiable, monotonic) function $g(\cdot)$,

$$\text{var}(g(\hat{\lambda})) \approx [\{g'(\lambda)\}^2 \text{var}(\lambda)]_{\lambda=\hat{\lambda}}.$$

so e.g.

$$\text{var}\left(\frac{1}{\hat{\lambda}}\right) = \text{var}(\hat{\mu}) \approx \frac{\hat{\mu}^2}{\sum_{i=1}^n \delta_i}$$

Exponential Distributions — MLE test

$$W = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\sqrt{\frac{\hat{\lambda}_1^2}{\Delta_1} + \frac{\hat{\lambda}_2^2}{\Delta_2}}} \approx N(0, 1).$$

Exponential Distributions — LRT test

$$2 \left\{ \Delta_1 \log \frac{\Delta_1}{\mathcal{T}_1} + \Delta_2 \log \frac{\Delta_2}{\mathcal{T}_2} - (\Delta_1 + \Delta_2) \log \frac{\Delta_1 + \Delta_2}{\mathcal{T}_1 + \mathcal{T}_2} \right\} \approx \chi_1^2$$

Log-rank Statistic

$$LR = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \sim \chi_1^2$$

3 Tables of Percentage Points (also known as Quantiles or Critical Values) for Three Standard Distributions

The tables contain values of quantiles q such that $P[X \leq q] = p$ for various probabilities p when X has the specified distribution (which may depend on particular degrees of freedom ν). In these tables, p has been expressed as a percentage rather than a decimal. The relevant R commands for generating the q are also shown. For the $N(0, 1)$ distribution, the tabulated function is also known as the Φ^{-1} function.

STANDARD NORMAL DISTRIBUTION PERCENTAGE POINTS

`qnorm(p)` where p is percentage, e.g. for 95%, $p = 0.95$

	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
<code>qnorm</code>	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

CHI-SQUARED PERCENTAGE POINTS

`qchisq(p, nu)` where p is percentage, e.g. for 95%, $p = 0.95$

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588

STUDENT'S t PERCENTAGE POINTS
 $qt(p, \nu)$ where p is percentage, e.g. for 95%, $p = 0.95$

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090