Data provided: Formula Sheet, incorporating Critical Values Tables

MAS61002

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2020–2021**

Medical Statistics

This is an open book exam.

*Answer **all** questions.*

You can work on the exam during the 24 hour period starting at 10am (BST), and you must submit your work within 2.5 hours of accessing the exam paper or by the end of the 24 hour period (whichever is earlier). ***Late submission will not be considered without extenuating circumstances.***

Calculations should be performed by hand. A university-approved calculator may be used. The use of any other calculational device, software or service is not permitted. To gain full marks, you will need to show your working.

By uploading your solutions you declare that your submission consists entirely of your own work, that any use of sources or tools other than material provided for this module is cited and acknowledged, and that no unfair means have been used.

Total marks available: 60.

1 Premature babies often suffer from 'patent ductus arteriosus', or PDA, a condition in which the the ductus arteriosus (DA) blood vessel fails to close. Babies with PDA have problems breathing and surgical or pharmacological treatments may be necessary.

In a study aimed at assessing the effectiveness of ibuprofen in closing the DA, the following trial was conducted. Two hundred and fifty premature babies with PDA were recruited and split at random into two groups of size 125. In the Treated Group T the subjects were treated with ibuprofen and in Control Group C no treatment was applied. The endpoint of interest was DA72, i.e. whether or not the vessel was still open at time 72 hours after treatment.

(i) (a) Comment on the particular ethical issues involved in conducting trials on babies. ***(3 marks)***

(b) Comment on the fact that no treatment is applied in Group C. ***(2 marks)***

(c) The babies were assigned to groups using block randomization. Explain, with reasons, why this randomization might have been used rather than simple randomization or minimization. ***(2 marks)***

(ii) The researchers noted that, without treatment, the DA closes within 72 hours in around 60% of cases. They want to be able to detect an improvement, from 60% to 80%, in the percentage of cases in which the DA closes within 72 hours following treatment. If they wanted to be 90% certain of detecting such an improvement, was the proposed sample size (250) a sensible number to recruit? ***(10 marks)***

(iii) In fact, only 225 babies completed the trial, 120 in Group T and 105 in Group C. The data on DA72 for the 225 babies is shown below.

	DA72		
	Closed	Still Open	
Ibuprofen	100	20	120
Control	61	44	105

(a) What can you conclude about the effectiveness of ibuprofen in closing the DA? ***(7 marks)***

(b) The researchers, in fact, decided to present their findings using the relative risk that DA72 is 'Still Open' under ibuprofen and control. Explain the key benefit of using the relative risk to summarize their results, even though it is more often used for observational studies. Provide the relative risk together with an approximate 95% confidence interval. ***(6 marks)***

2 A clinical trial was conducted on 24 patients with multiple myeloma, a cancer of plasma cells. Treatment entails high-dose chemotherapy and anti-cancer medication involving one of two distinct drugs (Bortezomib, and Lenalidomide–dexamethasone). Patients were accordingly randomized into two groups of treatment and followed up. The first group of 12 patients was given Bortezomib (Drug A) while the second group of 12 patients was given Lenalidomide-dexamethasone (Drug B). The outcome of interest was mortality. Some patients were lost to follow-up; these censored observations and denoted by asterisks(*). The data below show each patient's time until death in months.

	Drug A	Drug B
Patient	Time (months)	Time (months)
1	7.60	3.00
2	0.60*	22.70
3	1.20*	14.00
4	10.00	1.10
5	9.50	19.90
6	9.10*	2.50*
7	5.80	10.00
8	7.90*	6.10*
9	5.10*	0.80*
10	11.10	4.70*
11	8.40	5.10
12	3.90	11.10
Total	80.2	101.0

Table 1: Time to death

(i) Table 2 below shows the number at risk, number of deaths and expected number of deaths, at separate time points per drug indexed by subscripts A and B, correspondingly. Note that, i denotes the order of death events such that $t_{(i)}$ indicates the time of death, $r_i = r_{Ai} + r_{Bi}, d_i = d_{Ai} + d_{Bi}$. The total number of expected deaths in each group is indicated by E_A and E_B and the total observed number of deaths by O_A and O_B.

(a) Complete the missing values in the Table 2, using the data provided in Table 1.

(16 marks)

(b) Compute the log-rank statistic to compare the two survival distributions at the 5% level. ***(3 marks)***

i	$t_{(i)}$	Number at risk r_{Ai}	r_{Bi}	r_i	Number of deaths d_{Ai}	d_{Bi}	d_i	Expected number of deaths e_{Ai}	e_{Bi}
1	1.1				0	1	1		
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
Total					$O_A =$	$O_B =$		$E_A =$	$E_B =$

Table 2: Log-rank statistic implementation

3 A clinical trial aims to compare the efficacy of Bendamustin and Ribomustin in patients with cell lung cancer. Subjects are followed for up to 1 year for a potential relapse. Patients still in remission after the end of study are considered right-censored. The data are stored in `MARC` and coding for the different variables is shown below:

Coding:

Sex:	0 = female ; 1 = male
Treatment:	0 = Bendamustin (Treatment A); 1= Ribomustin (Treatment B)
Age:	age of patient centred on 45 years (i.e. 55yrs old is 10)
Time:	time until relapse (months)
Status:	0 = censored ; 1 = relapse

 Question 3 continued on next page

3 (continued)

(i) An analysis was implemented in R producing the following output:

```
Model 1
survreg(formula = MARC.sv ~ Sex + Age + Treatment, dist = "weibull")
               Value Std. Error      z       p
(Intercept)  0.83494    0.11982   6.97 3.2e-12
Sex          0.11917    0.13667   0.87    0.38
Age         -0.06231    0.00423 -14.75   2e-16
Treatment   -0.77086    0.13644  -5.65 1.6e-08
Log(scale)   0.03517    0.05090   0.69    0.49

Scale= 1.04

Weibull distribution
Loglik(model)= -309.5   Loglik(intercept only)= -391.4
Chisq= 163.78 on 3 degrees of freedom, p= 2.8e-35
Number of Newton-Raphson Iterations: 5
n= 300
```

(a) Write down the survival function of the estimated model (Model 1) ***(1 mark)***

(b) A researcher wishes to estimate an alternative model assuming that relapse time is exponentially distributed. How would you expect this to affect the estimated parameters? ***(2 marks)***

(ii) Further analysis was undertaken, giving the following results:

```
Model 2
coxph(formula = MARC.sv ~ Sex + Age + Treatment)

  n= 300, number of events= 231

               coef exp(coef)  se(coef)      z  Pr(>|z|)
Sex       -0.082669  0.920656  0.132791 -0.623    0.534
Age        0.056941  1.058594  0.005206 10.937   2e-16
Treatment  0.705232  2.024316  0.138101  5.107  3.28e-07

Likelihood ratio test= 138.6  on 3 df,          p=<2e-16
Wald test            = 129    on 3 df,          p=<2e-16
Score (logrank) test = 138.5  on 3 df,          p=<2e-16
```

(a) Based on the regression analysis (Model 2), would you say that Ribomustin (treatment B) seems to be more effective than Bendamustin (treatment A)? ***(2 marks)***

 Question 3 continued on next page

3 (continued)

(b) Explain why the sign of the estimated Treatment coefficient in Model 2, differs compared to Model 1. ***(3 marks)***

(c) Using Model 2, estimate the hazard ratio comparing two males, 43 and 65 years old respectively, both under treatment B. How would this hazard ratio estimate change if both individuals were instead females, aged 43 and 65 years old respectively, and both receiving treatment A? Justify your answer. ***(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 $r = min(n_1, n_2)$

$$t_r = |\frac{\bar{X_1} - \bar{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}|$$

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

$$t_r = |\frac{\bar{X_1} - \bar{X_2}}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}(\frac{1}{n_1} + \frac{1}{n_2})}}|$$

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

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

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

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

so e.g.

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

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%
`qnorm`	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, ν`)` 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, ν) 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