Stats 756 — Assignment 1

October 30, 2022

---

1. Describe what are adaptive clinical trials and explain their need and merits/demerits.

2. In a study comparing a treatment against a control, one would ideally like to terminate the experiment as soon as possible to reach a decision about the efficacy of the treatment; however, there are some advantages and disadvantages in doing so. Discuss them in detail.

3. Suppose among the work force of 2,864 individuals in a large company, a certain illness is present with a prevalence rate of about 0.06. A screening test for the illness is performed on all the workers, and it yielded the following results:

	Illness	*state*
Test result	Yes	No
Positive	154	584
Negative	18	2,108
Total	172	2,692

   From the above results, determine the following measures:
   Exact prevalence, Sensitivity, False Negative Rate, Specificity, False Positive Rate, Probability of Positive Test, Probability of Negative Test, Positive Predictive Value, Negative Predictive Value, Accuracy, Likelihood Ratio, Regret Given Positive Test, Posterior Probability of Illness given Positive Test, and Posterior Probability of Illness given Negative Test.

4. Given the cumulative distribution function

$$F(t) = 1 - \left\{1 - \lambda \left(\frac{t}{\sigma}\right)^{1/\alpha}\right\}^{1/\lambda}, \quad t \geq 0,$$

   with parameters $\alpha > 0$, $\sigma > 0$ and $\lambda \leq 0$, show that the hazard function is

$$h(t) = \frac{(t/\sigma)^{(1/\alpha)-1}}{\alpha\sigma\{1 - \lambda(t/\sigma)^{1/\alpha}\}}.$$

   Discuss its behavior.

5. The density of the logistic distribution is

$$f(y) = \frac{1}{\tau} \frac{e^{(y-\nu)/\tau}}{\left\{1 + e^{(y-\nu)/\tau}\right\}^2}, \quad -\infty < y < \infty,$$

where ν is the location parameter and $\tau > 0$ is the scale parameter. Let $Y = ln(T)$ follow this distribution. Then, prove that T follows the *log-logistic distribution* with density function

$$g(t) = \frac{\alpha t^{\alpha-1}\beta^{\alpha}}{\{1 + (t\beta)^{\alpha}\}^2}, \quad t \geq 0,$$

where $\alpha = \frac{1}{\tau}$ and $\beta = e^{-\nu}$.

6. Find the survival function and hazard function of the log-logistic distribution given above, and show that the hazard function has a maximum if $\alpha > 1$ and is monotonically decreasing if $\alpha < 1$.

7. The *odds* of an individual to fail before time t is given by $\frac{1-S(t)}{S(t)}$, where $S(t)$ is the survival function.

   (a) Evaluate the odds when T follows the above presented log-logistic distribution;

   (b) Show that the ratio of the odds between two individuals with different values of β is independent of t;

   (c) Suppose survival in days after kidney transplant can be modelled by a log-logistic distribution with $\alpha = 1.5$ and $\beta = 0.05$. Then, find the median survival time, the time at which the hazard function reaches a maximum, and the probability of surviving until that time.

8. The following data, taken from Bryson and Siddiqui (1969) who attribute them to a study by Siddiqui and Gehan, present the survival times (in days from diagnosis) of 43 patients suffering from chronic granulocytic leukemia:

7	47	58	74	177	232	273	285	317
429	440	445	455	468	495	497	532	571
579	581	650	702	715	779	881	900	930
968	1077	1109	1314	1334	1367	1534	1712	1784
1877	1886	2045	2056	2260	2429	2509		

   (a) Fit $Weibull(\alpha, \theta)$ distribution for these lifetime data;

   (b) Would an $Exponential(\theta)$ distribution be adequate to model these lifetime data? Explain;

   (c) Estimate the mean and median survival times in days after diagnosis.

9. Weeks to death of 51 adults with recurrent gliomas (**A**=astrocytoma, **G**=glioblastoma), as given by Bland and Altman (2004) who attribute them to a study by Rostomily et al., are as follows:

				A					
6	13	21	30	31*	37	38	47*	49	50
63	79	80*	82*	82*	86	98	149*	202	219
				G					
10	10	12	13	14	15	16	17	18	20
24	24	25	28	30	33	34*	35	37	40
40	40*	46	48	70*	76	81	82	91	112
181									

Here, * denotes right censored lifetimes.

(a) Fit $Weibull(\alpha, \theta)$ distributions for these two data sets on survival times for **A** and **G**, plot the corresponding survival functions, and compare them and comment;

(b) Fit $Gamma(\beta, \delta)$ distributions for these two data sets on survival times for **A** and **G**, plot the corresponding survival functions, and compare them and comment;

(c) Fit $Lognormal(\mu, \sigma)$ distributions for these two data sets on survival times for **A** and **G**, plot the corresponding survival functions, and compare them and comment;

(d) Use Akaike Information Criterion values based on the obtained fits of the three models and comment on which model you would select as working model.

10. The followiing data, taken from Klein and Moeschberger (1997) who attribute them to a study by Avalos et al., present the time (in days) from treatment until death of the appearance of complications in 27 patients suffering from Hodgkin's lymphoma (group **HL**) or other types of lymphoma (group **NHL**):

				HL				
30	36	41	52	62	108	132	180*	307*
406*	446*	484*	748*	1290*	1345*			
				NHL				
42	53	57	63	81	140	176	210*	252
476*	524	1037*						

Here, * denotes right censored lifetimes.

(a) Fit $Weibull(\alpha, \theta)$ distributions for these two data sets on survival times for **A** and **G**, plot the corresponding survival functions, and compare them and comment;

(b) Fit $Gamma(\beta, \delta)$ distributions for these two data sets on survival times for **A** and **G**, plot the corresponding survival functions, and compare them and comment;

(c) Fit $Lognormal(\mu, \sigma)$ distributions for these two data sets on survival times for **A** and **G**, plot the corresponding survival functions, and compare them and comment;

(d) Use Akaike Information Criterion values based on the obtained fits of the three models and comment on which model you would select as working model.

Good luck!