

1. You are given the following data:

Time	Group 1		Group 2		Group 3	
t_i	d_i	c_i	d_i	c_i	d_i	c_i
1	1	1	1	1	0	0
2	1	1	2	1	1	1
3	1	0	0	2	3	0
4	0	1	1	0	2	1

where d_i and c_i are the numbers of death and right-censored observations at time t_i from respective groups.

A Cox regression model is used for modelling the data. Let e^{β_1} and e^{β_2} be the hazard ratio from groups 2 and 3 compared with group 1. Write down the score functions of β_1 and β_2 based on this sample using the Breslow's approximation for handling tie observations.

2. In a class of 20 students, they are classified into 3 groups according to their mid-term scores: 8 of them in group "Excellent", 8 of them in group "Good" and 4 of them in "Fair". Professor B gave out 5 Grade A, 2 from group "Excellent", 2 from group "Good" and 1 from group "Fair". Is there any evidence that Professor B gave out Grade A randomly?
3. A study was conducted to investigate the waiting times for subsidised residential care services for the elderly and a random sample of 10 observations was presented as follows.

Observation	Age to join the queue	Age left the queue	Received the service?
1	73.2	75.3	Yes
2	68.5	70.1	No
3	72.6	76.4	No
4	65.1	70.2	Yes
5	79.0	83.5	No
6	64.9	68.3	Yes
7	75.2	78.0	Yes
8	74.7	76.2	Yes
9	68.3	71.2	No
10	75.4	76.2	No

The government claimed that the waiting times are exponentially distributed with mean 2.5 years. Test if the data support the claim.

4. The data file "data3.csv" contains the following data.

- Time : The number of days between registration and earlier of death or study analysis time in July, 1986.
Ind : 0 if Time is time to death, 1 if time to end of study.
Z1 : Treatment Code, 1= D-penicillamine, 2=placebo.
Z2 : Age in days.
Z3 : Sex, 0=male, 1=female.
Z4 : Presence of hepatomegaly, 0 = no, 1 = yes.
Z5 : Platelet count; coded value is number of platelets per-cubic-milliliter of blood divided by 1000.
Z6 : Prothrombin time, in seconds.
Z7 : Histologic stage of disease, graded 1, 2 3, or 4.

Use your seat number as the seed to generate a subsample of size equal to two third of the total observations. Use this subsample to answer the following questions.

- (a) Suppose we use Cox regression model to model the lifetimes using covariates $Z1 - Z7$ as main effects only (i.e., no interaction effects). Write down hazard rate at time t for individual with covariate vector \mathbf{Z} . How many model parameters we have in the model?
- (b) What is the log likelihood test to test the hypothesis that the variables $Z2, Z5$ and $Z6$ have no effect on the lifetime distribution? State the degree of freedom of the test and p value as well.
- (c) What is the score test to test the hypothesis that the variable $Z3, Z4$ and $Z7$ have no effect on the lifetime distribution? State the degree of freedom of the test and p value as well.
- (d) What is the Wald test to test the hypothesis that the variable $Z1, Z2$ and $Z4$ have no effect on the lifetime distribution? State the degree of freedom of the test and p value as well.
- (e) Construct a 95% confidence interval for the relative risk for the individual's age increased by 1 year given other conditions are the same. Assuming there are 365 days per year.
- (f) Construct a 95% confidence interval for the relative risk for the individual with $Z7 = 2$ compared with the individual with $Z7 = 4$, given other conditions are the same, based on model in (a).