Midterm Survival

Tonghao Zhang

2016/10/20

Problem 1

In a small clinical trail, the event times for 8 subjects are 10, 30, 20+, 30+, 20, 50+, 60+, 30, where the sign "+" means right-censored subjects. **Solve this problem manually without running R or SAS.**

1.1

We assume that the censoring here is random and non-informative. Please explain the "non-informativeness" for a random censoring scheme.

Solution

Non-informative means the selection of which patiens to be censored has provide no information on the their survival probability. For example, you can not censor a patient when he is about to die. To plan a random censoring scheme, we can determine the censoring time C_i and how many patients to be censored each time. Randomly select patients still at risk at each C_i (perhaps number the patients and use a random number generator).

1.2

Estimate the survival function S(t) using both the Kaplan-Meier estimator and Nelson- Aalen estimator, and plot the estimates.

Solution

Kaplan-Meier: S(10)=0.875, S(20)=0.75, S(30)=0.45

Nelson-Aalen: H(10)=0.125, H(20)=0.26786, H(30)=0.66786

—

—

—

—

—

—

—

—

—

1.3

Foucs on Kaplan-Meier estimator. Estimate the 95% confidence interval for $\hat{S}(25)$.

Solution

For t = 25, $\sigma_t^2(25) = \sum_{t_i \le 25} \frac{d_i}{Y_i(Y_i - d_i)} = 0.04167$. So the confidence interval is [0.622128, 0.87787].

1.4

Focus on Kaplan-Meier estimator. Perform the test that the 8 subjects are from a population with hazard rate function h0(t) = 1 when $0 < t \le 30$ and h0(t) = 2 when $t > 30$.

Solution

Use one-sample log-rank test, $W(t_i) = Y_i$. $O(\tau) = 4$, $E(\tau) = 300$, here $\tau = 60$. $Z(\tau) = O(\tau) - E(\tau)$. $V[Z(\tau)] = \int_0^\tau W^2(s) \frac{h_0(s)}{Y(s)} = 300$. Test statistic $\frac{Z(\tau)^2}{V[Z(\tau)]} = 292.05$, following a χ_1^2 distribution. Reject the null hypothesis.

1.5

Focus on Kaplan-Meier estimator. Assuming 2 more subjects are left censored at time 15 and 40, write out the first 2 iterations for estimating the survival function with both the 2 left censored subjects, 4 exact death time, and the 4 right censored subjects. Note that you just need to write out the first iterations without worrying about convergence.

Solution

—

—

—

—

—

—

—

—

—

Problem 2

Read Section 1.11. The data is available in R `library(KMsurv) data(tongue)` and `help(tongue)`.

2.1

Test the hypothesis that the survival rates of patients with cancer of the tongue are the same for patients with aneuploid and diploid tumors using the log-rank test.

```
library(survival)
library(KMsurv)
data("tongue");attach(tongue)
survdiff(Surv(time,delta)~type,rho = 0)
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ type, rho = 0)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## type=1 52       31     36.6     0.843      2.79
## type=2 28       22     16.4     1.873      2.79
##
##  Chisq= 2.8  on 1 degrees of freedom, p= 0.0949
```

Solution

H_0: hazard function is the same for two groups. H_A hazard function not the same.

The log-rank test-statistic value is 2.8, following a χ^2_1. p-value is 0.0949, there is not enough evidenvce to reject H_0 on a confidence level of 0.95.

2.2

If primary interest is in detecting differences in survival rates between the two types of cancers which occur soon after the diagnosis of the cancer, repeat Part 2.1 using a more appropriate test statistic.

```
group1<-tongue[type==1,];group2<-tongue[type==2,]
fit1<-summary(survfit(Surv(group1$time,group1$delta)~1))
fitnull<-summary(survfit(Surv(tongue$time,tongue$delta)~1))
tD<-max(max(group1$time*group1$delta),max(group2$time*group2$delta))
D<-length(fitnull$time[fitnull$time<=tD])
di1<-rep(0,D);yi1<-rep(0,D)
for(i in 1:D){
  di1[i]<-nrow(group1[group1$time*group1$delta==fitnull$time[i],])
  yi1[i]<-nrow(group1[group1$time>=fitnull$time[i],])
}
numerator<-0;denominator<-0
for(i in 1:D){
  numerator<-numerator+fitnull$n.risk[i]*(di1[i]-yi1[i]*fitnull$n.event[i]/fitnull$n.risk[i])
  denominator<-denominator+fitnull$n.risk[i]^2*
    (yi1[i]/fitnull$n.risk[i])*(1-yi1[i]/fitnull$n.risk[i])*
      ((fitnull$n.risk[i]-fitnull$n.event[i])/(fitnull$n.risk[i]-1))*fitnull$n.event[i]
}
denominator<-sqrt(denominator)
Z<-numerator/denominator
Z*Z
```

```
## [1] 3.305493
```

```
1-pchisq(Z*Z,df=1)
```

```
## [1] 0.06904864
```

```
detach(tongue)
```

Solution

H_0: hazard function is the same for two groups. H_A hazard function not the same.

If the primary interest is in early time, using the Gehan test which assign heavier weight to early time is appropriate. Z = 3.3054925 the p-value is 0.0690486. Conclude H_0 on a confidence level of 95%.

Problem 3

Read Section 1.3 about the bone marrow transplant data, which is available in R `library(KMsurv) data(bmt)` and `help(bmt)`. Several event times are described for patients receiving a bone marrow transplant for leukemia. Consider the time to development of acute graft-versus-host disease (AGVHD). As a prophylactic treatment, patients at two of the hospitals were given a treatment combining methotrexate (MTX) with cyclosporine and possibly methylprednisilone. Patients at the other hospitals were not given methotrexate but rather a combination of cyclosporine and methylprednisilone. Of primary interest in studying AGVHD is a test of the effectiveness of the MTX regime to prevent AGVHD. Use Breslows method for handling ties.

3.1

Using an appropriate Cox model, test the hypothesis of no difference in the rate of development of AGVHD between MTX and no MTX patients. Find a point estimate and a 95% confidence interval for the relative risk of AGVHD for patients on the MTX protocol as compared to those not given MTX.

```
data("bmt");attach(bmt)
##z10 MTX indicator
##ta Time To Acute Graft-Versus-Host Disease
##da Acute GVHD Indicator 1-Developed Acute GVHD 0-Never Developed Acute GVHD)
surv3<-Surv(ta,da)
coxmod1<-coxph(surv3~z10,method = "breslow")
coxmod1re<-coxph(surv3~0,method = "breslow")
LR<--2*(coxmod1re$loglik-coxmod1$loglik[2])
p<-1-pchisq(LR,1)
```

Solution

Construct the Cox model with one variable z10-indicator for MTX usage, the baseline is patients without MTX treatment. Likelihood ratio test statistic is 0.4342122, following a χ_1^2 distribution. p-value is 0.5099293. Conclude H_0: no effect in MTX on a confidence level of 95%. The relative risk for MTX patients compared to baseline is 0.741584, and the 95% confidence interval for relative risk is [0.297777, 1.846841].

3.2

Patients were also grouped into risk categories based on their status at the time of transplantation. These categories were as follows: acute lymphoblastic leukemia (ALL) with 38 patients and acute myeloctic leukemia (AML). The latter category was further subdivided into low-risk-first remission (54 patients) and high-risk-second remission or untreated first relapse or second or greater relapse or never in remission (45 patients). Test the hypothesis of interest (no effect of MTX on development of AGVHD) adjusting for the three disease categories.

```
##group Disease Group 1-ALL, 2-AML Low Risk, 3-AML High Risk
cox2full<-coxph(surv3~z10+as.factor(group),method = "breslow")
cox2fullSum<-summary(cox2full)
cox2re<-coxph(surv3~as.factor(group),method = "breslow")
cox2reSum<-summary(cox2re)

LR<--2*(cox2re$loglik[2]-cox2full$loglik[2])
p<-1-pchisq(LR,3)
```

Solution

In order to test the null hypothesis: no effect in MTX treatment, I performed a local test using log-rank. The full model is constructed with 3 variables - indicator of MTX, AML-low risk, AML-high risk, the baseline is patients in ALL group without MTX treatment.

Test statistic is 0.588486, following a χ^2_3 distribution. The according p-value is 0.8990631. Conclude the null hypothesis on a confidence level of 95%.

3.3

Test for the possibility of an interaction effect on AGVHD between the disease categories and the use MTX.

```
G2<-1*(group==2);G3<-1*(group==3)
I2<-G2*z10;I3<-G3*z10
cox3full<-coxph(surv3~z10+G2+G3+I2+I3,method = "breslow")
cox3re<-coxph(surv3~z10+G2+G3,method = "breslow")

LR3<--2*(cox3re$loglik[2]-cox3full$loglik[2])
p<-1-pchisq(LR3,5)
```

Solution

In order to test the null hypothesis: no interaction between the usage of MTX and disease catergory, I performed a local test using log-rank. The full model is constructed with 5 variables - indicator of MTX, AML-low risk, AML-high risk and two interaction terms, the baseline is patients in ALL group without MTX treatment.

Test statistic is 1.0616613, following a χ^2_5 distribution. The according p-value is 0.9574406. Conclude the null hypothesis on a confidence level of 95%.

3.4

Using the factors of age, sex, CMV status, FAB class, waiting time to transplant, and disease category as defined in Example 8.5, find the best model to test the primary hypothesis of no MTX effect on the occurrence of AGVHD. Test the primary hypothesis and find an estimate of the relative risk of occurrence of AGVHD for an MTX patient as compared to a non-MTX patient.

Solution

```
##G2 AML-low risk, G3 AML high risk
G2<-1*(group==2);G3<-1*(group==3)
##A1 patient age -28, A2 donor age -28, A3 = A1*A2
A1<-z1-28;A2<-z2-28;A3<-A1*A2
##S1 sex of patient, 1 for male;S2 sex of donor, 1 for male; S3 = S1*S2
S1<-z3;S2<-z4;S3<-S1*S2
##C1 patient CMV status, 1 positive;C2 donor CMV; C3 = C1*C2
C1<-z5;C2<-z6;C3<-C1*C2
##F1 FAB class
F1<-z8

fit.group<-coxph(surv3~z10+G2+G3)
fit.age<-coxph(surv3~z10+A1+A2+A3)
fit.sex<-coxph(surv3~z10+S1+S2+S3)
fit.CMV<-coxph(surv3~z10+C1+C2+C3)
fit.FAB<-coxph(surv3~z10+F1)
```

```
AIC.group<--2*fit.group$loglik[2]+2*3;AIC.group
```

```
## [1] 252.7605
```

```
AIC.age<--2*fit.age$loglik[2]+2*4;AIC.age
```

```
## [1] 247.2634
```

```
AIC.sex<--2*fit.sex$loglik[2]+2*4;AIC.sex
```

```
## [1] 255.6673
```

```
AIC.CMV<--2*fit.CMV$loglik[2]+2*4;AIC.CMV
```

```
## [1] 252.567
```

```
AIC.FAB<--2*fit.FAB$loglik[2]+2*2;AIC.FAB
```

```
## [1] 252.1753
```

From the AIC values, we include age into the model.

```
fit.age.group<-coxph(surv3~z10+A1+A2+A3+G2+G3)
fit.age.sex<-coxph(surv3~z10+A1+A2+A3+S1+S2+S3)
fit.age.CMV<-coxph(surv3~z10+A1+A2+A3+C1+C2+C3)
fit.age.FAB<-coxph(surv3~z10+A1+A2+A3+F1)

AIC.age.group<--2*fit.age.group$loglik[2]+2*6;AIC.age.group
```

```
## [1] 247.2352
```

```
AIC.age.sex<--2*fit.age.sex$loglik[2]+2*7;AIC.age.sex
```

```
## [1] 252.6194
```

```
AIC.age.CMV<--2*fit.age.CMV$loglik[2]+2*7;AIC.age.CMV
```

```
## [1] 250.7727
```

```
AIC.age.FAB<--2*fit.age.FAB$loglik[2]+2*5;AIC.age.FAB
```

```
## [1] 249.1293
```

In this step, all AIC value start to increase. I validate my conclusion by further inference on the significance of these variable using local likelihood ratio test.

```
LRgroup<--2*(fit.age$loglik[2]-fit.age.group$loglik[2]);p<-1-pchisq(LRgroup,6)
LRgroup;p
```

```
## [1] 4.028207
```

```
## [1] 0.672859
```

```
LRsex<--2*(fit.age$loglik[2]-fit.age.sex$loglik[2]);p<-1-pchisq(LRsex,7)
LRsex;p
```

```
## [1] 0.6440764
```

```
## [1] 0.9987295
```

```
LRCMV<--2*(fit.age$loglik[2]-fit.age.CMV$loglik[2]);p<-1-pchisq(LRCMV,7)
LRCMV;p
```

```
## [1] 2.490688
```

```
## [1] 0.9277966
```

```
LRFAB<--2*(fit.age$loglik[2]-fit.age.FAB$loglik[2]);p<-1-pchisq(LRFAB,5)
LRFAB;p
```

```
## [1] 0.1341204
```

```
## [1] 0.9996659
```

All variables are not significant. So the model should be Cox proportional hazard model with patients without MTX treatment and donor patient age all equal to 28 being baseline, factors are MTX indicator, donor age, patient age and intercation between donor age and patient age.

```
fit.reduce<-coxph(surv3~A1+A2+A3)
LR<--2*(fit.reduce$loglik[2]-fit.age$loglik[2]);p<-1-pchisq(LR,4)
fit.age
```

```
## Call:
## coxph(formula = surv3 ~ z10 + A1 + A2 + A3)
##
##         coef exp(coef) se(coef)     z    p
## z10 -0.51176   0.59944  0.47396 -1.08 0.28
## A1   0.04725   1.04838  0.03184  1.48 0.14
## A2   0.03431   1.03491  0.03257  1.05 0.29
## A3  -0.00299   0.99701  0.00235 -1.27 0.20
##
## Likelihood ratio test=9.43  on 4 df, p=0.0511
## n= 137, number of events= 26
```

```
LR;p
```

```
## [1] 1.263423
```

```
## [1] 0.86755
```

H_0: MTX has no effect, H_A: MTX has effect.

The likelihood ratio test statistic calculated is 1.2634233, following χ^2_4 distribution, the p-value is 0.86755. There is not enough evidence to reject H_0. Conclusion: MTX has no effection on a confidence level of 95%.

The relative risk is 0.59944.