Lab 5

Ziwen Zhang

2022-12-08

# Question 1

Consider data tongue from the package KMsurv. Do a simple graphical check of the proportional hazards assumptions and state your conclusions.

library(KMsurv)  
library(survival)

## Warning: package 'survival' was built under R version 4.2.2

library(survminer)

## Warning: package 'survminer' was built under R version 4.2.2

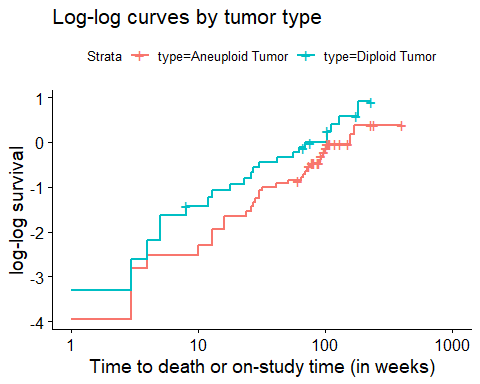
## Loading required package: ggplot2

## Loading required package: ggpubr

##   
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':  
##   
## myeloma

data(tongue)  
  
tongue <- tongue %>%  
 mutate(type = ifelse(type == 1, "Aneuploid Tumor", "Diploid Tumor"))  
  
tongue\_fit <- survfit(Surv(time, delta) ~ type, data = tongue)  
  
ggsurvplot(tongue\_fit, fun = "cloglog", data = tongue) +   
 labs(title = "Log-log curves by tumor type",  
 y = "log-log survival",   
 x = "Time to death or on-study time (in weeks)")



The proportional hazards assumption is violated since the log-log survival curves are not parallel.

# Question 2

For this problem we use data kidtran (death times of 863 kidney transplantpatients) from KMsurv package. Patients can be classified by race and sex into four groups.

• Test the hypothesis that there is no difference in survival between the four groups. • for each gender group, conduct individual tests of the hypothesis of no racial differences in survival rates.

data("kidtran")  
  
log\_rank <- survdiff(Surv(time, delta) ~ race + gender, data = kidtran)  
log\_rank

## Call:  
## survdiff(formula = Surv(time, delta) ~ race + gender, data = kidtran)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## race=1, gender=1 432 73 69.25 0.2025 0.4013  
## race=1, gender=2 280 39 47.39 1.4860 2.2531  
## race=2, gender=1 92 14 14.52 0.0184 0.0205  
## race=2, gender=2 59 14 8.84 3.0173 3.2245  
##   
## Chisq= 4.7 on 3 degrees of freedom, p= 0.2

H0: There is no difference in survival between the four groups. H1: There is at least one difference in survival among the four groups.

Since the p-value is 0.2 > 0.05, there is no evidence to reject the null hypothesis and thus there is no difference in survival among the four groups.

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ✔ purrr 0.3.5

## Warning: package 'purrr' was built under R version 4.2.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

# for male (gender=1)  
male <- kidtran %>%  
 filter(gender == 1)  
  
log\_rank\_male <- survdiff(Surv(time, delta) ~ race, data = male)  
log\_rank\_male

## Call:  
## survdiff(formula = Surv(time, delta) ~ race, data = male)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## race=1 432 73 71.9 0.0168 0.097  
## race=2 92 14 15.1 0.0801 0.097  
##   
## Chisq= 0.1 on 1 degrees of freedom, p= 0.8

# for female (gender=2)  
female <- kidtran %>%  
 filter(gender == 2)  
  
log\_rank\_female <- survdiff(Surv(time, delta) ~ race, data = female)  
log\_rank\_female

## Call:  
## survdiff(formula = Surv(time, delta) ~ race, data = female)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## race=1 280 39 44.79 0.748 4.85  
## race=2 59 14 8.21 4.076 4.85  
##   
## Chisq= 4.8 on 1 degrees of freedom, p= 0.03

For male group:

H0: There is no difference in survival between the two race groups. H1: There is a difference in survival between the two race groups.

Since the p-value is 0.8 > 0.05, there is no evidence to reject the null hypothesis and thus there no difference in survival between the two race groups.

For female group:

H0: There is no difference in survival between the two race groups. H1: There is a difference in survival between the two race groups.

Since the p-value is 0.03 < 0.05, there is an evidence to reject the null hypothesis and thus there is a difference in survival between the two race groups.

# Question 3

For the female subset from the previous problem (kidtran) estimate the difference in RMST for τ = 2500 using package survRM2.

library(survRM2)

## Warning: package 'survRM2' was built under R version 4.2.2

female <- female %>%  
 mutate(race = ifelse(race == 1, 0, 1))  
  
RMSTdiff <- rmst2(time = female$time, status = female$delta, arm = female$race, tau = 2500, covariates = NULL, alpha = 0.05)  
RMSTdiff

##   
## The truncation time: tau = 2500 was specified.   
##   
## Restricted Mean Survival Time (RMST) by arm   
## Est. se lower .95 upper .95  
## RMST (arm=1) 1993.597 124.863 1748.869 2238.325  
## RMST (arm=0) 2209.651 45.547 2120.381 2298.921  
##   
##   
## Restricted Mean Time Lost (RMTL) by arm   
## Est. se lower .95 upper .95  
## RMTL (arm=1) 506.403 124.863 261.675 751.131  
## RMTL (arm=0) 290.349 45.547 201.079 379.619  
##   
##   
## Between-group contrast   
## Est. lower .95 upper .95 p  
## RMST (arm=1)-(arm=0) -216.054 -476.555 44.447 0.104  
## RMST (arm=1)/(arm=0) 0.902 0.793 1.027 0.119  
## RMTL (arm=1)/(arm=0) 1.744 0.984 3.093 0.057

The difference in RMST is 216.054

# Question 4

The cumulative hazard function for the Weibull distribution is given by

Derive a simple graphical diagnostic to check whether the Weibull distribution is adequate for data.

Please note that this question is not about specific data set. Your solution should be applicable for any data when you want to use Weibull distribution

Solution: Take the logarithm to the left and right of the equation:

Therefore, we can calculate

and plot

versus

. If the data can fit in this linear model, the Weibull distribution is adequate for data.