Q1. Researchers are interested in the age at which liver cancer is commonly diagnosed in Taiwan. The data, data_HW4_1.csv, collected for this purpose includes the age at first diagnosis, follow-up time, whether cancer is present (0/1: no/yes), whether the individual is a carrier of hepatitis B (0/1: no/yes), gender (0/1: female/male), alcohol consumption (0/1: no/yes), elevated mild liver index (0/1: no/yes), and elevated moderate to high liver index (0/1: no/yes).

1. If we are interested in studying the factors that influence the incidence rate of age at which individuals develop liver cancer using the Cox model, what specific model would you recommend?
2. Can we consider the Age as a covariate in the model? (explain)
3. Utilizing the syntax fit=coxph(Surv(your model)); plot(survfit(fit)) to examine at which age individuals exhibit statistical significance in developing liver cancer.
4. What are the implications when statistics show significance—whether they indicate harmful or protective effects? (explain)
5. Overall, do you think the Cox model is suitable for use in this context?

Provide a conclusion by the key points from 1 to 5.

Q2. Researchers are investigating the correlation between the incidence rate of liver cancer and gender (0/1: female/male). The data collected for this purpose, found in data_HW4_2.csv, includes information such as follow-up time and the presence of cancer (0/1: no/yes).

6. Do you believe the Cox model is appropriate for application in this context?
7. What is the impact of gender? (provide an explanation)
8. Is the effect statistically significant?
9. What is the difference in 5-year survival rates between genders?

Summarize the key points from questions 6 to 9 to draw a conclusion.

Q3. If researchers are more focused on studying the incidence rate of liver cancer rather than mortality, it's crucial to bear in mind the close connection between mortality and the incidence of liver cancer. Please use the DATA_HW3.txt file to analyze the incidence rate of liver cancer. DATA_HW3.txt includes information on the time to liver cancer (T_1), time to death (T_2), and cancer status (1/0: no/yes), as well as the states of HBV (1/0: carrier/healthy) respectively order in the 1st to 4th columns.

10. Inference the effect from the states of HBV by comparing their survival function of T_1 (and given the comment!)
11. You will need the R-code in the supplement.
12. In the code, we stratified the data into groups with and without HBV. Explain why stratification was chosen over using the Cox model.
13. If we insist on using the Cox model, what assumption do we need to make?

Hint: The estimated θ, $\hat{\theta} = 42.158$ and 15.09 under state of HBV=0 and 1, respectively.

14. In the code, why we need to using the PLA function before we calculate the survival function of T_1.

Supplement

```
library("Copula.surv")

typeS=names(table(data$S))
Z=data$S
Z[data$S==typeS[1]]=0
Z[data$S==typeS[2]]=1
x.obs_0=data$X1[Z==0]
y.obs_0=data$X2[Z==0]
dy_0=data$D[Z==0]
dx_0=ifelse(x.obs_0==y.obs_0,0,1)
x.obs_1=data$X1[Z==1]
y.obs_1=data$X2[Z==1]
dy_1=data$D[Z==1]
dx_1=ifelse(x.obs_1==y.obs_1,0,1)
z.obs=ifelse(data$X1<=data$X2,data$X1,data$X2)
dz=data$D
dz1=which(data$X1<data$X2)
dz[dz1]=1
z.obs_0=z.obs[Z==0]
z.obs_1=z.obs[Z==1]
dz_0=dz[Z==0]
dz_1=dz[Z==1]
#########################################################################

theta_0=U2.Clayton(x.obs_0,y.obs_0,dx_0,dy_0)
theta_1=U2.Clayton(x.obs_1,y.obs_1,dx_1,dy_1)



#########################################################################
Sz0=survfit(Surv(z.obs_0,dz_0) ~ 1, conf.type = "log-log")
Sz1=survfit(Surv(z.obs_1,dz_1) ~ 1, conf.type = "log-log")
Sy0=survfit(Surv(y.obs_0,dy_0) ~ 1, conf.type = "log-log")
Sy1=survfit(Surv(y.obs_1,dy_1) ~ 1, conf.type = "log-log")
```

```
pLA=function(yy,tt,LL){
pla=function(yi){
loc=sum(yi>=tt)
if(loc==0)ans=1
if(loc>0)ans=LL[loc]
ans
}
apply(matrix(yy),1,pla)
}
###################################

Sx0.y=(Sz0.y^(-theta_0[1])-Sy0.y^(-theta_0[1])+1 )^(1/(-theta_0[1]))
Sx1.y=(Sz1.y^(-theta_1[1])-Sy1.y^(-theta_1[1])+1 )^(1/(-theta_1[1]))
```