**1、**

## R code  
> data<-read.csv("ACTG175(speff2trial).txt")

> library(survival)

> Y<-Surv(data$days, data$cens==1)

> survdiff(Y~data$arms)

N Observed Expected (O-E)^2/E (O-E)^2/V

data$arms=0 532 181 116 37.030 47.67

data$arms=1 522 103 134 6.988 9.40

data$arms=2 524 109 132 4.158 5.58

data$arms=3 561 128 139 0.933 1.27

Chisq= 49.2 on 3 degrees of freedom, p= 1e-10

## log rank test的p-value小于0.05，所以认为四种治疗方式间有差异。

**2、**

## R code

> survdiff(Y~data$arms, rho=0.5)

N Observed Expected (O-E)^2/E (O-E)^2/V

data$arms=0 532 170.5 108 36.035 49.58

data$arms=1 522 95.5 124 6.745 9.72

data$arms=2 524 100.8 123 4.149 5.96

data$arms=3 561 119.1 130 0.899 1.31

Chisq= 51.2 on 3 degrees of freedom, p= 4e-11

## weighted log rank test的p-value小于0.05，所以认为四种治疗方式间有差异。

**3、**

## R code

> data<-read.csv("wcgsdata.csv")

> library(survival)

> Y<-Surv(data$Time169, data$Chd69==1)

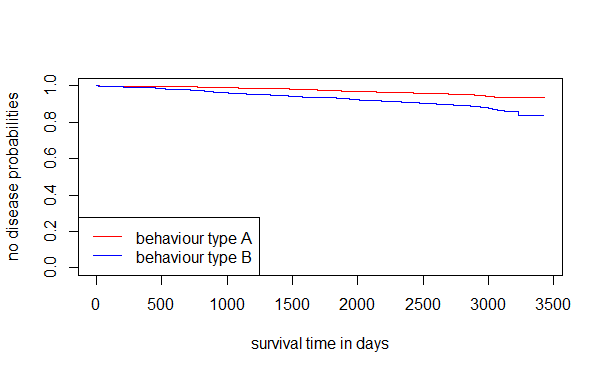
> kmfit<-survfit(Y~data$Dibpat0)

> plot(kmfit, xlab="survival time in days",

+ ylab="survival probabilities", col=c("red","blue"))

> legend("bottomleft", c("behaviour type A","behaviour type B"),

+ lty="solid", col=c("red","blue"))



## 从图上可以看出，behaviour type B危害更大，会提高心脏病发作概率。

**4、**

## R code

> survdiff(Y~data$Dibpat0)

N Observed Expected (O-E)^2/E (O-E)^2/V

data$Dibpat0=0 1565 79 131 20.5 41.8

data$Dibpat0=1 1589 178 126 21.3 41.8

Chisq= 41.8 on 1 degrees of freedom, p= 1e-10

## log rank test的p-value小于0.05，所以认为两种行为方式间有差异。

**5、**

## R code

# 换算单位

> data$Weight0<-data$Weight0 \* 0.45359

> data$Height0<-data$Height0 \* 0.0254

# 计算BMI并离散化

> data$BMI<-data$Weight0 / (data$Height0)^2

>data$BMI<-cut(data$BMI, breaks=c(-Inf, 18.5, 25.0, 30.0, Inf),

labels=c("underweight","healthyweight","overweight","obese"),

include.lowest=TRUE, right=FALSE)

# Ncigs0离散化

> data$Ncigs0[data$Ncigs0!=0]<-"smoker"

> data$Ncigs0[data$Ncigs0==0]<-"not smoker"

# Stratified Log Rank Test

> survdiff(Y ~ data$Dibpat0 + strata(data$BMI,data$Ncigs0))

N Observed Expected (O-E)^2/E (O-E)^2/V

data$Dibpat0=0 1565 79 127 18.1 36.1

data$Dibpat0=1 1589 178 130 17.6 36.1

Chisq= 36.1 on 1 degrees of freedom, p= 2e-09

## stratified log rank test的p-value小于0.05，所以认为两种行为方式间有差异。