# STAT537: Statistics for Research I: HW#11

Due on Nov. 10, 2016

*Dr. Schmidhammer TR 11:10am – 12:25pm*

**Wenqiang Feng**

# Contents

# Problem 1

Ott Exercise 16.5b

*Solution.* (a) **Provide a scatterplot of the data with regression lines which would allow a visual assessment of whether there is a significant relationship between risk index of CVD and the number of cigarettes smoked:** The scatterplot of the data with regression lines can be found in Figure 1 which indicates that there is a significant relationship between risk index of CVD and the number of cigarettes smoked.

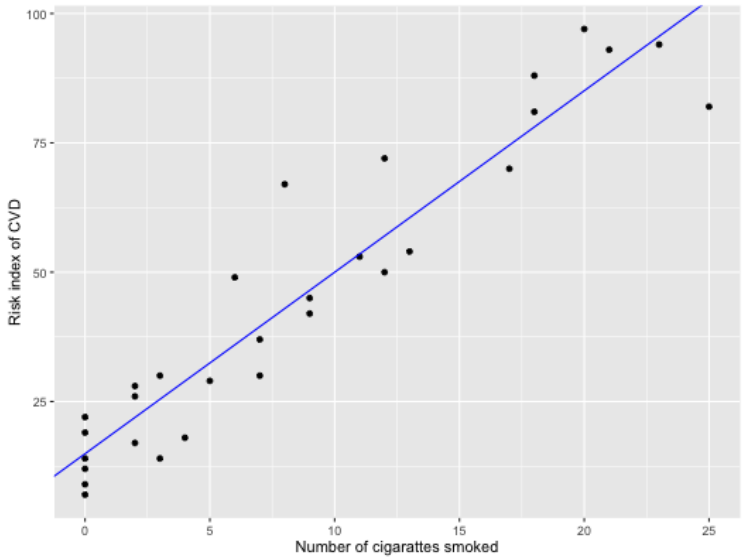


Figure 1: Scatterplot for these data, with number of cigarattes smoked on the horizontal axis, and Risk index of CVD on the vertical axis.

□

# Problem 2

Ott Exercise 16.6 ab

*Solution.* (a) **Test the hypothesis that the relationships between risk index and number of cigarettes have equal slopes for the three treatments at the $\alpha = 0.05$ level.** The fitted results indicate that the p-value of the interaction term is $0.06186 > 0.05$. Hence we do not have enough information to reject $H_0$ : the covariates have equal slopes. Hence, we may conclude that slopes for the three treatments at the $\alpha = 0.05$ level are not equal to each other. Moreover, the fitted results for each treatment confirms our conclusion. The fitted results are as follows:

```
Analysis of Variance Table

Response: RISK
                Df  Sum Sq Mean Sq   F value    Pr(>F)
NOCIG            1 21410.9 21410.9 1055.3276 < 2.2e-16 ***
TREATMENT        2  1820.1   910.1   44.8561 7.814e-09 ***
NOCIG:TREATMENT  2   127.1    63.5    3.1319   0.06186 .
```

```
Residuals       24    486.9    20.3
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

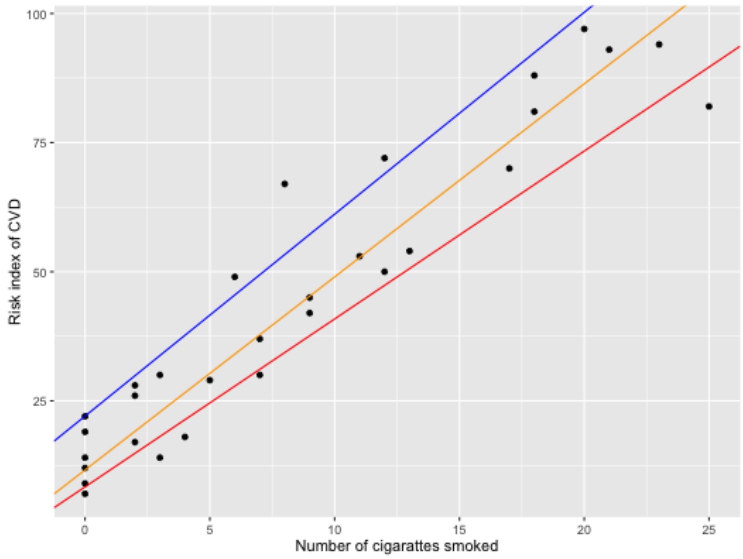


Figure 2: Scatterplot for these data, with number of cigarattes smoked on the horizontal axis, and Risk index of CVD on the vertical axis.

(b) **Does there appear to be a difference in the mean risk index for the three treatments?** The following fitted results indicate that the p-value for TREATMENT is $1.674e-08 < 0.05$ which shows there is a difference in the mean risk index for the three treatments.

```
Analysis of Variance Table

Response: RISK
          Df  Sum Sq Mean Sq F value    Pr(>F)
NOCIG      1 21410.9 21410.9 906.642 < 2.2e-16 ***
TREATMENT  2  1820.1   910.1  38.536 1.674e-08 ***
Residuals 26   614.0    23.6
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

□

# Appendix

## R code for HW#11

Listing 1: Source code for problem 1

```
# reference: http://www.stat.columbia.edu/~martin/W2024/R3.pdf
```

```r
rm(list = ls())
# set the path or enverionment
#linux
#setwd("/home/feng/Dropbox/UTK_Course/Stat537/Excel/CH16")
#Mac
setwd("/Users/wenqiangfeng/Dropbox/UTK_Course/Stat537/Excel/CH16")

# 16.5 (b)
library(readxl)
rawdata = read_excel("ex16-5.xls",sheet = 1)
attach(rawdata)
library(car)
#plot(NOCIG,RISK, main="Scatterplot",
#     xlab="Number of cigarettes smoked daily, ", ylab="Risk index for CVD ", pch=19)
#abline(lm(NOCIG~RISK,data = rawdata),col="blue",untf=T)
library(ggplot2)
lm = lm(RISK~NOCIG, data=rawdata)
ggplot(rawdata, aes(x=NOCIG, y=RISK)) +
  geom_point() +
  geom_abline(intercept = lm$coef[1], slope = lm$coef[2], col="blue") +
  xlab("Number of cigarattes smoked") +
  ylab("Risk index of CVD")

# 16.6 (a)
fit_lm = lm(RISK~NOCIG+TREATMENT+NOCIG:TREATMENT, data=rawdata)
anova(fit_lm)

treatment_C = rawdata[TREATMENT=='C',]
treatment_1 = rawdata[TREATMENT=='I',]
treatment_2 = rawdata[TREATMENT=='II',]

fit_C = lm(RISK~NOCIG, data=treatment_C)
fit_1 = lm(RISK~NOCIG, data=treatment_1)
fit_2 = lm(RISK~NOCIG, data=treatment_2)

ggplot(rawdata, aes(x=NOCIG, y=RISK)) +
  geom_point() +
  geom_abline(intercept = fit_C$coef[1], slope = fit_C$coef[2],
                               col="blue",show.legend = TRUE) +
  geom_abline(intercept = fit_1$coef[1], slope = fit_1$coef[2],
                              col="red",show.legend = TRUE) +
  geom_abline(intercept = fit_2$coef[1], slope = fit_2$coef[2],
                              col="orange",show.legend = TRUE) +
  xlab("Number of cigarattes smoked") +
  ylab("Risk index of CVD")

# (b)
fit = lm(RISK~NOCIG+TREATMENT, data=rawdata)
anova(fit)
```