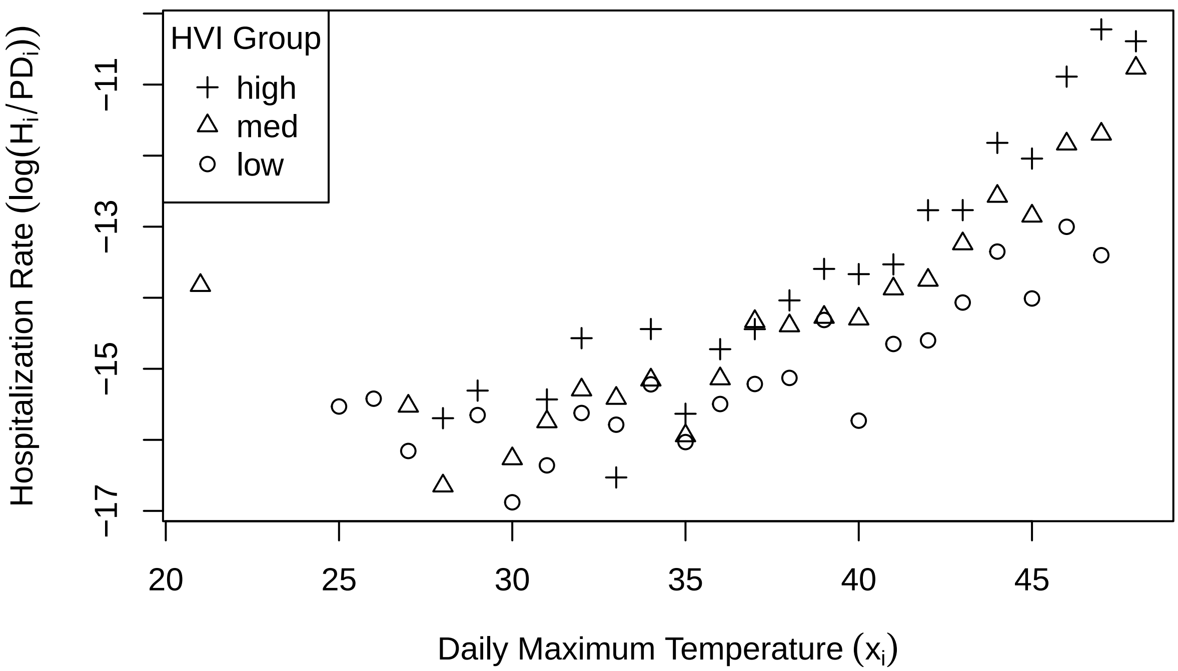
33.5/34

INF 511 Assignment 3

Jen Diehl (6179236), Sam Watson (6174574), Natasha Wesely (6180693)

2022-10-03

```
hw3.df<- readRDS(file="hw3.df.RDS")
plot(log(H/PD) ~ x, data=hw3.df, pch=as.numeric(HVI),
      xlab=expression("Daily Maximum Temperature" ~ (x[i])),
      ylab=expression("Hospitalization Rate" ~ (log(H[i]/PD[i]))))
legend("topleft", legend=c("high", "med", "low"), pch=c(3,2,1),title="HVI Group")
```



Question 1

```
lmod <- lm(log(H/PD)~I(x-30)+I((x-30)^2)+HVI ,data = hw3.df)
summary(lmod)
```

```
##
## Call:
## lm(formula = log(H/PD) ~ I(x - 30) + I((x - 30)^2) + HVI, data = hw3.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.53075 -0.17166  0.03969  0.33353  0.82653
##
## Coefficients:
##                 Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   -16.242074   0.127876 -127.015  < 2e-16 ***
## I(x - 30)      -0.009785   0.021081   -0.464 0.644243
## I((x - 30)^2)   0.014748   0.001419   10.391 5.98e-15 ***
## HVIM            0.615151   0.157495    3.906 0.000244 ***
## HVIH            1.140516   0.162304    7.027 2.43e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5208 on 59 degrees of freedom
## Multiple R-squared:  0.8996, Adjusted R-squared:  0.8928
## F-statistic: 132.1 on 4 and 59 DF,  p-value: < 2.2e-16
```

Question 2

```
summary(lmod)$r.squared
```

```
## [1] 0.8995752
```

The proportion of variability of log hospitalization rate is 0.8995752 explained by the association of temperature and HVI.

Question 3

The intercept from this model output is around -16. If we were to draw a line for the data on the plot above the intercept is also around -16. In the model we have shifted the y-axis to the right where 30 now equals 0 (or where the 20 is on the above plot). Therefore, it is appropriate to infer about the intercept from the model.

Question 4

i

H_0: y = $\beta_0 + \beta_1$ * (Temperature - 30) + β_2 * (Temperature - 30)^2

H_a: y = $\beta_0 + \beta_1$ * (Temperature - 30) + β_2 * (Temperature - 30)^2 + β_3 * HVIM + β_4 * HVIH

```
lmod2 <- lm(log(H/PD)~I(x-30)+I((x-30)^2) ,data = hw3.df)
summary(lmod2)
```

```
##
## Call:
## lm(formula = log(H/PD) ~ I(x - 30) + I((x - 30)^2), data = hw3.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.88669 -0.42529  0.07863  0.42095  1.29114
##
## Coefficients:
##                  Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   -15.736960   0.129438 -121.579  < 2e-16 ***
## I(x - 30)       0.003001   0.027916    0.107    0.915
## I((x - 30)^2)   0.014435   0.001886    7.655 1.72e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6952 on 61 degrees of freedom
## Multiple R-squared:  0.815,  Adjusted R-squared:  0.8089
## F-statistic: 134.4 on 2 and 61 DF,  p-value: < 2.2e-16
```

```
modanova <- anova(lmod, lmod2)
anova(lmod, lmod2)
```

anova(lmod2, lmod)

-0.5

```
## Analysis of Variance Table
##
## Model 1: log(H/PD) ~ I(x - 30) + I((x - 30)^2) + HVI
## Model 2: log(H/PD) ~ I(x - 30) + I((x - 30)^2)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     59 16.003
## 2     61 29.482 -2   -13.479 24.847 1.487e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ii

The F statistic is 24.847.

iii

The p-value is 1.487e-08.

iv

We can reject our null hypothesis that the reduced model is sufficient because our p-value is less than 0.05.

Question 5

i

There are 5 parameters in the full model.

ii

There are 3 parameters in the reduced model.

iii

For the full model DF is 59 and the reduced model DF is 61.

iv

```
pval = 1 - pf(modanova$F[2],
              df1 = modanova$Res.Df[1],
              df2 = modanova$Res.Df[2])
print(pval)
```

```
## [1] 0
```

The p-value is very small number and R is rounding to zero.

Question 6

```
b4hat = summary(lmod)$coefficients[5,1]
seB4hat = summary(lmod)$coefficients[5,2]
tcrit = qt(0.95, df = 59)

(CIlower = b4hat - (tcrit*seB4hat))
```

```
## [1] 0.8692906
```

We are 95% confident the mean effect of high HVI (relative to low HVI) is at least 0.8692906 (or greater).

Question 7

H_0: $\beta_4 = 0$

H_a: $\beta_4 > 0$

```
b4hat = summary(lmod)$coefficients[5,1]
seB4hat = summary(lmod)$coefficients[5,2]

testStat = (b4hat - 0) / seB4hat
(pval = 1 - pt(testStat, df = 59))
```

```
## [1] 1.217393e-09
```

The test statistic is far into the right tail of the t distribution (test stat t-value = 7.027036), and p-value is close to zero (1.217393e-09). Therefore, we can reject the null hypothesis (that there is no effect of high HVI). The effect of high HVI is significantly greater than zero.