

Homework #5 Solutions

input = time ^(in hours), response = concentration of a solution ^(in %)

① (a) $\hat{Y} = 2.575 - 0.324x$

(b) $\hat{\mu}_5 = 0.955\%$

(c) The linear regression estimate of a mean will have smaller variance than the saturated model estimate.

If the linear regression model is appropriate, then its estimate will also have small bias.

② (a) $Y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i=1, \dots, c \\ j=1, \dots, n_i \end{cases}$

(b) $SSPE = 0.1574$, $df_{PE} = n - c = 10$

$SSLF = 2.7673$, $df_{LF} = c - 2 = 3$

(c) $F_{LF}^* = 58.603$, $p = .000$

The data is not compatible with the linear regression model.
We will need a more flexible model.

(d) see plot of fitted values

HW #5 Computing

Data from Exercise 3.15

A designed experiment is conducted to study the concentration of a solution (y) over time (x) in hours.

:

```
hw5.data = read.table(
  'http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/Ku
  tnerData/Chapter%20%203%20Data%20Sets/CH03PR15.txt')
colnames(hw5.data)=c("concentration", "time")
str(hw5.data)

## 'data.frame':  15 obs. of  2 variables:
## $ concentration: num  0.07 0.09 0.08 0.16 0.17 0.21 0.49 0.58 0.53 1.22
## $ time          : num  9 9 9 7 7 7 5 5 5 3 ...

regression.mod = lm(concentration ~ time, data=hw5.data)
regression.mod

##
## Call:
## lm(formula = concentration ~ time, data = hw5.data)
##
## Coefficients:
## (Intercept)          time
##      2.575         -0.324

fact.time = as.factor(hw5.data$time)

full.mod = lm(concentration ~ fact.time - 1, data=hw5.data)
full.mod

##
## Call:
## lm(formula = concentration ~ fact.time - 1, data = hw5.data)
##
## Coefficients:
## fact.time1 fact.time3 fact.time5 fact.time7 fact.time9
##      2.8367      1.1467      0.5333      0.1800      0.0800
```

```

predict(regression.mod,data.frame(time=3))

##          1
## 1.603333

predict(full.mod,data.frame(fact.time="3"))

##          1
## 1.146667

anova(regression.mod,full.mod)

## Analysis of Variance Table
##
## Model 1: concentration ~ time
## Model 2: concentration ~ fact.time - 1
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      13 2.9247
## 2      10 0.1574  3     2.7673 58.603 1.194e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(sort(hw5.data$time), full.mod$fitted.values[order(hw5.data$time)], type
= "b",ylab = "concentration of the solution",xlab = "time in hours",ylim =
c(min(hw5.data$concentration),max(hw5.data$concentration)))

abline(regression.mod$coefficients,col="red")
points(hw5.data$time,hw5.data$concentration,pch=16)

```

