Stat 482

Homework #4 Solutions

1. input = number of copiers, response = service time

(a) confidence band at $x = 3$: $[40.28, 48.77]$

(b) As we increase the scope of the estimation problem, we increase the probability of data incompatible with a model. Thus, we need to increase the range of compatibility

(c) see plot of confidence band and confidence intervals

2. input = age, response = muscle mass

(a) $E(MSR) = \sigma^2 + SS_x \cdot \beta_1^2$, $E(MSE) = \sigma^2$

(b) If $\beta_1 \approx 0$, then $E(MSR) \approx E(MSE)$.

Small F^* indicates data compatible with the null model.

Large F^* indicates data supporting the alternative model.

(c) $F^* = 174.06$, $p = .000$ Since the data is not compatible with the no effect model, we accept the model which includes age as a predictor of muscle mass.

(d) $r^2 = .75$ We estimate that 75% of the variation in muscle mass is explained by age.

HW #4-1 Computing

Data from Exercise 1.20

A sample of service calls is selected to investigate the relationship between the number of copiers to be serviced (x) and the service time (y)

:

```
hw4_1.data = read.table(
'http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KuntnerData/Chapter%20%201%20Data%20Sets/CH01PR20.txt')
colnames(hw4_1.data)=c("time","number")

attach(hw4_1.data)

reg.mod = lm(time~number)

x.all = c(1:10)
b0 = reg.mod$coefficients[1]
b1 = reg.mod$coefficients[2]
y.hat = b0 + b1*x.all


e = reg.mod$residuals
n = length(e)
sse = sum(e^2)
dfe = n-2
mse = sse / dfe

x.sample = number
x.bar = mean(x.sample)
x.star = x.sample - x.bar
ssx = sum(x.star^2)

y.lower = y.hat - sqrt(2*qf(.95,2,dfe))*sqrt(mse*(1/n+(x.all-x.bar)^2/ssx))
y.upper = y.hat + sqrt(2*qf(.95,2,dfe))*sqrt(mse*(1/n+(x.all-x.bar)^2/ssx))

y.all = matrix(c(y.hat,y.lower,y.upper),ncol=3)

colnames(y.all) = c("mean.y","lower.limit","upper.limit")
```

```
cbind(x.all,y.all)
```

```
##       x.all    mean.y lower.limit upper.limit
##  [1,]     1  14.45509    8.396337    20.51385
##  [2,]     2  29.49034   24.403822    34.57686
##  [3,]     3  44.52559   40.278528    48.77265
##  [4,]     4  59.56084   55.927234    63.19444
##  [5,]     5  74.59608   71.224244    77.96792
##  [6,]     6  89.63133   86.090677    93.17199
##  [7,]     7 104.66658  100.579543   108.75362
##  [8,]     8 119.70183  114.815910   124.58775
##  [9,]     9 134.73708  128.902591   140.57156
## [10,]    10 149.77232  142.901302   156.64335
```

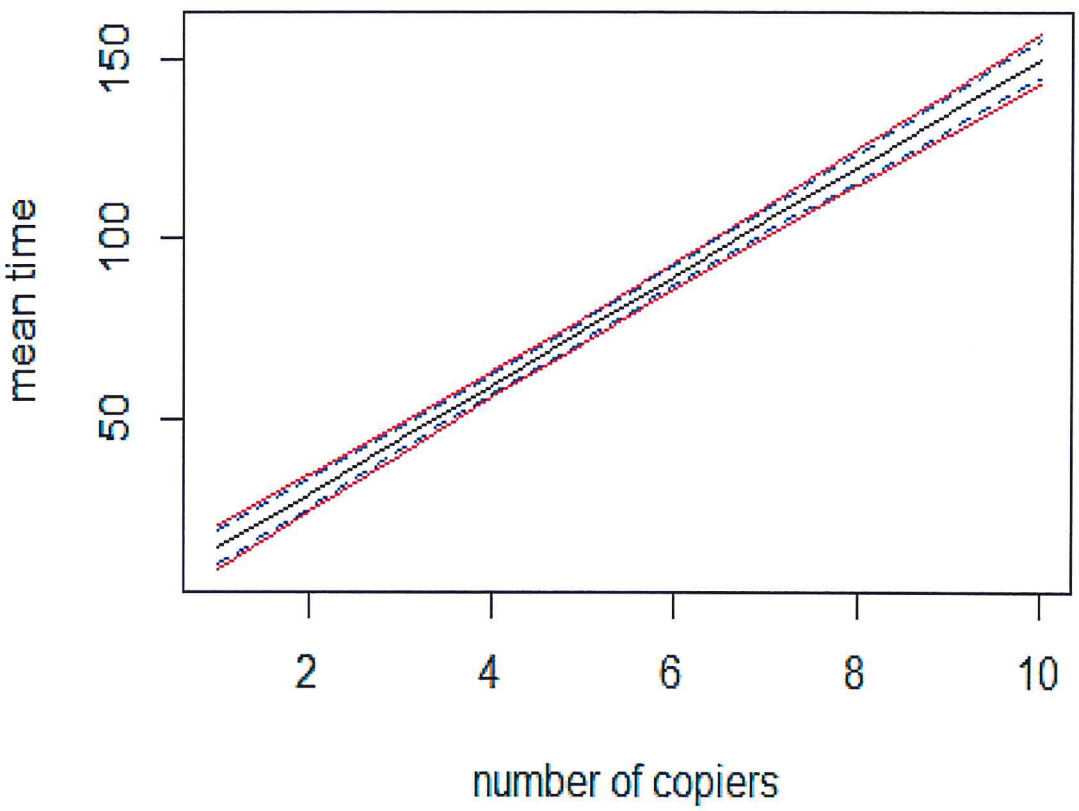
```
matplot(x.all,y.all,type="l",lty=1,col=c("black","red","red"),
        xlab = "number of copiers",ylab = "mean time")
```



```
ci = predict(reg.mod,data.frame(number=x.all),interval = "confidence")
ci = cbind(x.all,ci)
ci
```

```
##    x.all       fit       lwr       upr
## 1      1  14.45509   9.63614  19.27404
## 2      2  29.49034  25.44468  33.53600
## 3      3  44.52559  41.14760  47.90357
## 4      4  59.56084  56.67078  62.45089
## 5      5  74.59608  71.91422  77.27794
## 6      6  89.63133  86.81520  92.44746
## 7      7 104.66658 101.41587 107.91729
## 8      8 119.70183 115.81571 123.58794
## 9      9 134.73708 130.09650 139.37765
## 10    10 149.77232 144.30732 155.23733
```

```
y.lwr = ci[,3]
y.upp = ci[,4]
points(x.all,y.lwr,type="l",lty=2,col="blue")
points(x.all,y.upp,type="l",lty=2,col="blue")
```


mean time
50
100
150
2
4
6
8
10
number of copiers

HW #4-2 Computing

Data from Exercise 1.27

A sample of women is selected to investigate the relationship between age (x) and muscle mass (y)

:

```
hw4_2.data = read.table(
'http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Chapter%20%201%20Data%20Sets/CH01PR27.txt'
  )
colnames(hw4_2.data)=c("muscle.mass","age")

hw4_2.mod = lm(muscle.mass ~ age, data=hw4_2.data)

anova(hw4_2.mod)
```

```
## Analysis of Variance Table
##
## Response: muscle.mass
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## age        1 11627.5 11627.5  174.06 < 2.2e-16 ***
## Residuals 58  3874.4    66.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(hw4_2.mod)
```

```
##
## Call:
## lm(formula = muscle.mass ~ age, data = hw4_2.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.1368  -6.1968  -0.5969   6.7607  23.4731
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 156.3466     5.5123   28.36   <2e-16 ***
## age          -1.1900     0.0902  -13.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.173 on 58 degrees of freedom
## Multiple R-squared:  0.7501, Adjusted R-squared:  0.7458
## F-statistic: 174.1 on 1 and 58 DF,  p-value: < 2.2e-16
```