Assignment 1 Answers

Problem 1

(a) $E(X_3 - 2X_1 + 1) = E(X_3) - 2E(X_1) + 1 = -2$

$Var(X_3 - 2X_1 + 1) = Var(X_3) + 4Var(X_1) - 4Cov(X_3, X_1) = 9$

(b) $E(3X_1 + 2X_2 + X_3 + 1) = E(3X_1) + 2E(X_2) + E(X_3) + 1 = 8$

$Var(3X_1 + 2X_2 + X_3 + 1) = 9Var(3X_1) + 4Var(X_2) + Var(X_3) + 12Cov(X_1, X_2) + 6Cov(X_1, X_3) + 4Cov(X_2, X_3) = 50$

(c) $Cov(2X_1, X_3 + 1) = 2Cov(X_1, X_3) + 2Cov(X_1, 1) = 0$

$Cov(X_1 - X_2, X_1 - X_3) = Cov(X_1, X_1) + Cov(X_1, -X_3) + Cov(-X_2, X_1) + Cov(-X_2, -X_3) = 0$

Problem 2

(a) $E(X) = 14, Var(x) = 54, X \sim N(19, 54)$

(b) $X \sim \chi^2_2$

(c) distribution unknown (it's not a χ^2_2 as X could be negative).

(d) $X = \frac{(Y_5-5)/\sqrt{5}}{\sqrt{Z_1/5}}, X \sim t_5$

(e) $X \sim \chi^2_7$

(f) $E(Y_1 + Y_2) = 2, Var(Y_1 + Y_2) = 6, X = \frac{\chi^2_1/1}{\chi^2_3/3} \sim F(1, 3)$

3. Question 2.6.

Y : Selling price / 1000

x_1 : taxes / 1000

a. $\hat{Y} = 13.32 + 3.32x_1$ (See R output)

b. H_0: The regression is not significant (or $\beta_1 = 0$)

H_1: the regression is significant (or $\beta_1 \neq 0$).

ANOVA Table:

Source	df	SS	MS	F	P-value
Regression	1	636.16	636.16	72.56	2.051×10^{-8}
Error	22	192.89	8.77		
Total	23				

P-value < 0.05, the regression is significant.

c. $R^2 = 76.7\%$. Thus 76.7% of the total variability in selling price is explained by this model.

d. A 95% CI on β_1 is $\hat{\beta}_1 \pm t_{0.025, 22}\, s.e.(\hat{\beta}_1)$

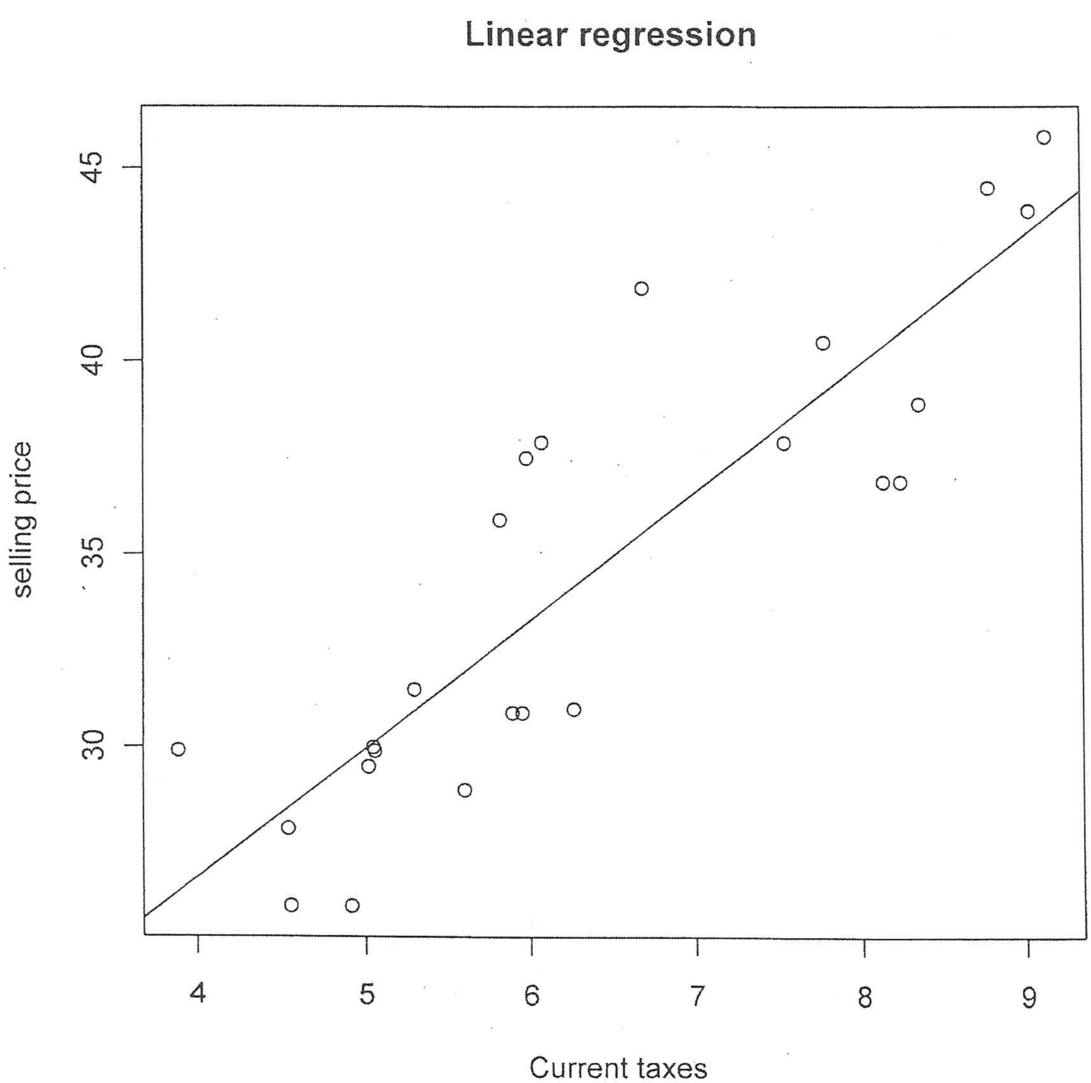
i.e $3.32 \pm 2.074\,(0.3903) = (2.51, 4.13)$.

For 1 unit increase in x_1, Y increases by β_1. If the current tax goes up by \$1000, the selling price increases by \$3320.

e. A 95% CI on the mean selling price with $x_1 = 7.5$ is $(36.72, 39.79) \times 1000$.

f. See the attached plot.

```
#Question 2.6:  R codes and output

#Read the data file in R
Data2=read.table(file="TableB4.prn", header=T)
y=Data2[,1]
x1=Data2[,2]
data3=data.frame(cbind(x1,y))
print(data3)  #Check for the data set

#Scatter plot of y versus x1
#Check for linear relationship
plot(x1,y, xlab="Current taxes", ylab="selling price",
main="Linear regression")

#Fit a simple linear regression model
l1=lm(y~x1,data=data3)
abline(l1, col="blue")   #Add the LS line to the scatter plot

#LS estimates, R-squared, standard errors of the estimates,
#F-test, t-tests
summary(l1)

#Construct ANOVA table for the linear regression model including the
#F-test
anova(l1)

#Prediction for x1=7.5  ($7500/1000)
predict.lm(l1,newdata=data.frame(x1=7.50),interval="confidence")

#Output:
#---------------------------------------------
> summary(l1)
Call:
lm(formula = y ~ x1, data = data3)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8343 -2.3157 -0.3669  1.9787  6.3168

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.3202     2.5717   5.179 3.42e-05 ***
x1            3.3244     0.3903   8.518 2.05e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.961 on 22 degrees of freedom
Multiple R-squared: 0.7673,     Adjusted R-squared: 0.7568
F-statistic: 72.56 on 1 and 22 DF,  p-value: 2.051e-08

> anova(l1)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x1         1 636.16  636.16  72.556 2.051e-08 ***
Residuals 22 192.89    8.77
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> predict.lm(l1,newdata=data.frame(x1=7.50),interval="confidence")
       fit      lwr      upr
1 38.25296 36.71776 39.78816

#---------------------------------------------------------------
```


Linear regression
selling price
45
40
35
30
4
5
6
7
8
9
Current taxes