Introduction to Econometrics

Assignment One

M. Randy Reid

2025-05-22

Question 1

Using the following 5 observations on X and Y, estimate the intercept and regression coefficient (estimate α and β in yi = α + βxi + ei) using R:

```
x <- c(1, 2, 3, 4, 5)
y <- c(2, 2, 3, 3, 6)
```

Report your output - annotate and explain. Provide a summary regression output from R:

You can also embed plots, for example:

```
#fit a linear model for y on x, where y is the dependent variable and x is the independent variable
#we're examining how changes in x impact y
model <- lm(y ~ x)

#summarize the fit
print(model)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##         0.5          0.9
```

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##   1   2   3   4   5
##  0.6 -0.3 -0.2 -1.1  1.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.500      0.995   0.503   0.6499
## x              0.900      0.300   3.000   0.0577 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9487 on 3 degrees of freedom
## Multiple R-squared:   0.75,  Adjusted R-squared:  0.6667
## F-statistic:     9 on 1 and 3 DF,  p-value: 0.05767
```

```
#Intepretation of Results:
#The slope is estimated to be .90 which reflects the change in y with a single unit change in X
#Our intercept is estimated to be 0.50. As such, Y is 0.50 when x is 0.
#R-Squared is 0.75 meaning that three fortha of the variation in Y is attributable to (or explained by) X.
#Our model shows a positive linear relationship between the variables.
```

Question 2

Obtain predicted values and the residuals, and print these along with the values of Y and X.

```
#recalculate residuals from previous model
model_residuals <- resid(model)

#variables and residuals in single table or date frame
q2_table <- data.frame(x=x,y=y,model_residuals=model_residuals)

print(q2_table)
```

```
##   x y model_residuals
## 1 1 2             0.6
## 2 2 2            -0.3
## 3 3 3            -0.2
## 4 4 3            -1.1
## 5 5 6             1.0
```

Question 3

If another observation (obs. number 6) will have X=3, what is your best prediction for Y? Can you calculate the residual for this prediction?

```
#we know that our model is y(hat) = 0.5 + 0.9x and can therefore predict a new value of y given x = 3
#by substitution
y_estimate <- 0.5 + 0.9*3
print(y_estimate)
```

```
## [1] 3.2
```

Question 4

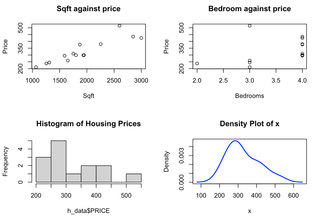
We're using downloaded data (excel format) so I need install the necessary packages and establish a working directory

```
#We're using downloaded data so I need to establish a working directory
setwd('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files')
#import data
h_data <- read.csv('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files/house.csv')
#print for visual inspection
#print(h_data)

#calculate the means and SD for each variable and organize in a data frame
h_averages <- sapply(h_data, mean)
h_sigma <- sapply(h_data, sd)

h_mean_sigma <- data.frame(house_variables = names(h_data), averages = h_averages, SD = h_sigma)
print(h_mean_sigma)
```

```
##        house_variables    averages          SD
## PRICE            PRICE  327.564286  86.9954694
## SQFT              SQFT 1919.714286 580.4868552
## BATHS            BATHS    2.250000   0.6044705
## BEDRMS          BEDRMS    3.571429   0.6462062
```

##Question 5 Provide scatterplots for sqft x price and bedrms x price (price on the vertical axis) on two separate plots, also plot a histogram of house prices, and a density plot. Report all 4 plots on one plot window - in R, you can use the par(mfrow=c(2,2))

```
setwd('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files')
h_data <- read.csv('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files/house.csv')

par(mfrow=c(2,2))
#plot 1
plot(h_data$SQFT,h_data$PRICE, main = "Sqft against price", ylab = "Price", xlab = "Sqft")
#plot 2
plot(h_data$BEDRMS,h_data$PRICE, main = "Bedroom against price", ylab = "Price", xlab = "Bedrooms")
#plot 3
hist(h_data$PRICE, main = "Histogram of Housing Prices")
#plot 4
dh_price <- density(h_data$PRICE)
plot(dh_price, main = "Density Plot of x", xlab = "x", ylab = "Density", col = "blue", lwd = 2)
```



##Question 6 Use OLS to estimate the equation: pricei = α + β · sqf ti + ei. Report the summary statistics for the regression results. What are the estimated values of α and β?

```
setwd('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files')
h_data <- read.csv('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files/house.csv')

#OLS
h_model <- lm(h_data$PRICE ~ h_data$SQFT)

#summary stats
summary(h_model)
```

```
##
## Call:
## lm(formula = h_data$PRICE ~ h_data$SQFT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.517 -28.285  -0.688   6.795  95.526
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 68.20074   37.43707   1.822   0.0935 .
## h_data$SQFT  0.13511    0.01872   7.216 1.66e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.19 on 12 degrees of freedom
## Multiple R-squared:  0.8127, Adjusted R-squared:  0.7971
## F-statistic: 52.07 on 1 and 12 DF,  p-value: 1.663e-05
```

```
#pulling coefficients
estimate_intercept <- coef(h_model)[1]
estimate_beta <- coef(h_model)[2]

#print estimates
print(estimate_beta)
```

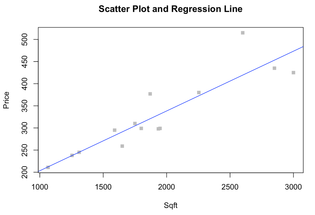
```
## h_data$SQFT
##   0.1351053
```

```
print(estimate_intercept)
```

```
## (Intercept)
##    68.20074
```

##Question 7 Using the results from the above regression, plot price and sqft (i.e. scatter plot with the regression line) on one graph, and plot the residuals on another graph

```
setwd('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files')
#import data
h_data <- read.csv('/Users/mrr/Desktop/Economics/Intro.Econometrics/Data Files/house.csv')

#OLS
h_model <- lm(h_data$PRICE ~ h_data$SQFT)

#syntax for scatter plot
plot(h_data$PRICE ~ h_data$SQFT, main = "Scatter Plot and Regression Line", xlab = "Sqft", ylab = "Price", pch=1
5, col="grey", )

#regressionline
abline(h_model,col="blue", lwd=1)
```




```
#residuals
h_residuals <- resid(h_model)
plot(h_residuals)
```