Chapter 3 HW

Fabiani Rafael

Conceptual Questions

Exercise 1: Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.?

- The null hypothesis is that the variables TV, radio and newspaper have no effect on sales. From the table the p-values corresponding to TV and radio are less than 0.0001 indicating that the null hypothesis can be rejected. The p-value for newspaper is 0.8599 which is greater than 0.05, so the null hypothesis cannot be rejected. This would draw us to conclude that TV and radio have a significant effect on sales, while newspaper does not.

Exercise 3: Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Level (1 for College and 0 for High School), X4 = Interaction between GPA and IQ, and X5 = Interaction between GPA and Level. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get

$$\hat{\beta}_0 = 50,\ \hat{\beta}_1 = 20,\ \hat{\beta}_2 = 0.07,\ \hat{\beta}_3 = 35,\ \hat{\beta}_4 = 0.01,\ \hat{\beta}_5 = -10$$

.?

(a) Which answer is correct, and why?

  i. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates.

  ii. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates.

  iii. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.

  iv. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough.

- From the problem :

$$\bar{X} = \begin{bmatrix} 1 \\ \bar{x_1} \\ \bar{x_2} \\ \bar{x_3} \\ \bar{x_4} \\ \bar{x_5} \end{bmatrix} = \begin{bmatrix} 1 \\ \bar{GPA} \\ \bar{IQ} \\ \bar{Level} \\ (\bar{GPA} \times \bar{IQ}) \\ (\bar{GPA} \times \bar{Level}) \end{bmatrix},\ X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ GPA \\ IQ \\ Level \\ (GPA \times IQ) \\ (GPA \times Level) \end{bmatrix},\ \hat{\beta} = \begin{bmatrix} 50 \\ 20 \\ 0.07 \\ 35 \\ 0.01 \\ -10 \end{bmatrix}$$

Now Y is given by

$$Y = X^T \hat{\beta}$$

$$= \begin{bmatrix} 1 & x_1 & x_2 & x_3 & x_4 & x_5 \end{bmatrix} \begin{bmatrix} 50 \\ 20 \\ 0.07 \\ 35 \\ 0.01 \\ -10 \end{bmatrix}$$

$$= \begin{bmatrix} 50 + 20(\text{GPA}) + 0.07(\text{IQ}) + 35(\text{Level}) + 0.01(\text{GPA} \times \text{IQ}) - 10(\text{GPA} \times \text{Level}) \end{bmatrix}$$

For fixed value of IQ, GPA, and letting Level $= 1$ for College, 0 for High School. the salary of a college graduate is given by

$$Y = 50 + 20(\text{GPA}) + 0.07(\text{IQ}) + 35(\text{Level}) + 0.01(\text{GPA} \times \text{IQ}) - 10(\text{GPA} \times \text{Level})$$
$$= 50 + 20(\text{GPA}) + 0.07(\text{IQ}) + 35(1) + 0.01(\text{GPA} \times \text{IQ}) - 10(\text{GPA} \times \text{Level})$$

and for high school graduates we have

$$Y = \bar{X}^T \hat{\beta}$$

$$= \begin{bmatrix} 1 & \bar{x_1} & \bar{x_2} & \bar{x_3} & \bar{x_4} & \bar{x_5} \end{bmatrix} \begin{bmatrix} 50 \\ 20 \\ 0.07 \\ 35 \\ 0.01 \\ -10 \end{bmatrix}$$

$$= \begin{bmatrix} 50 + 20(\bar{\text{GPA}}) + 0.07(\bar{\text{IQ}}) + 35(\bar{\text{Level}}) + 0.01(\bar{\text{GPA}} \times \bar{\text{IQ}}) - 10(\bar{\text{GPA}} \times \bar{\text{Level}}) \end{bmatrix}$$

$$Y = 50 + 20(\bar{\text{GPA}}) + 0.07(\bar{\text{IQ}}) + 35(\bar{\text{Level}}) + 0.01(\bar{\text{GPA}} \times \bar{\text{IQ}}) - 10(\bar{\text{GPA}} \times \bar{\text{Level}})$$
$$= 50 + 20(\bar{\text{GPA}}) + 0.07(\bar{\text{IQ}}) + 35(0) + 0.01(\bar{\text{GPA}} \times \bar{\text{IQ}}) - 10(\bar{\text{GPA}} \times 0)$$
$$= 50 + 20(\bar{\text{GPA}}) + 0.07(\bar{\text{IQ}}) + 0.01(\bar{\text{GPA}} \times \bar{\text{IQ}})$$

We can then equate these equations to find the conditions under which college graduates earn more than high school graduates. With Level $= 0$, $\bar{\text{Level}} = 0$ and $\text{GPA}, \bar{\text{GPA}}$ and $\text{IQ}, \bar{\text{IQ}}$ being the same fixed values, we have

$$50 + 20(\text{GPA}) + 0.07(\text{IQ}) + 35 + 0.01(\text{GPA} \times \text{IQ}) - 10(\text{GPA} \times \text{Level}) = 50 + 20(\bar{\text{GPA}}) + 0.07(\bar{\text{IQ}}) + 0.01(\bar{\text{GPA}} \times \bar{\text{IQ}})$$
$$50 - 50 + 20(\text{GPA} - \bar{\text{GPA}}) + 0.07(\text{IQ} - \bar{\text{IQ}}) + 35 + 0.01(\text{GPA} \times \text{IQ}) - (\bar{\text{GPA}} \times \bar{\text{IQ}}) = 10(\text{GPA} \times \text{Level})$$
$$35 = 10(\text{GPA} \times \text{Level})$$
$$3.5 = (\text{GPA} \times \text{Level})$$

So for a GPA of 3.5 or higher, college graduates earn more than high school graduates. So the correct answer is iii. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.

(b) Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.

- Continuing from before we have
$$Y = X^T \hat{\beta}$$
Where,
$$X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{GPA} \\ \text{IQ} \\ \text{Level} \\ (\text{GPA} \times \text{IQ}) \\ (\text{GPA} \times \text{Level}) \end{bmatrix}, \hat{\beta} = \begin{bmatrix} 50 \\ 20 \\ 0.07 \\ 35 \\ 0.01 \\ -10 \end{bmatrix}$$

So then

$$= \begin{bmatrix} 1 & x_1 & x_2 & x_3 & x_4 & x_5 \end{bmatrix} \begin{bmatrix} 50 \\ 20 \\ 0.07 \\ 35 \\ 0.01 \\ -10 \end{bmatrix}$$

$$= 50 + 20(\text{GPA}) + 0.07(\text{IQ}) + 35(\text{Level}) + 0.01(\text{GPA} \times \text{IQ}) - 10(\text{GPA} \times \text{Level})$$
$$= 50 + 20(4.0) + 0.07(110) + 35(1) + 0.01(4.0 \times 110) - 10(4.0 \times 1)$$
$$= 50 + 80 + 7.7 + 35 + 4.4 - 40$$
$$= 137.1$$

So the salary of a college graduate with IQ of 110 and a GPA of 4.0 is predicted to be 137.1.

(c) True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.

- False, since the value of the coefficient is not considered when determining the statistical significance of a variable. Instead if our goal is to determine the interaction effect's significance we would have to find and inspect the associatied p-value.

Applied Questions

Exercise 10 : This question should be answered using the Carseats data set.

(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

```
data(Carseats)
#summary(Carseats)

#fitting the model using price Urban and US
fit <- lm(Sales ~ Price + Urban + US, data = Carseats)
summary(fit)
```

```
##
## Call:
## lm(formula = Sales ~ Price + Urban + US, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9206 -1.6220 -0.0564  1.5786  7.0581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.043469   0.651012  20.036  < 2e-16 ***
## Price       -0.054459   0.005242 -10.389  < 2e-16 ***
## UrbanYes    -0.021916   0.271650  -0.081    0.936
## USYes        1.200573   0.259042   4.635 4.86e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.472 on 396 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2335
```

```
## F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

(b) Provide an interpretation of each coefficient in the model. Be careful—some of the variables in the model are qualitative!

- The coefficient for Price is -0.0545, which means that for a one unit increase in Price, Sales decrease by 0.0545 units. Our constant of 13.04347 suggests that we start at a baseline of approximately 13. The coefficient for UrbanYes is -0.0219, which means that for a store in an urban area, Sales decrease by 0.0219 units. The coefficient for USYes is 1.2006, which in essence means that for a store in the US, Sales increase by 1.2006 units i.e if a car seat model is sold in the US the coefficient suggests we gain about 1.2 in sales.

(c) Write out the model in equation form, being careful to handle the qualitative variables properly.

$$\text{Sales} = \beta_0 + \beta_1 \times \text{Price} + \beta_2 \times \text{UrbanYes} + \beta_3 \times \text{USYes} + \epsilon$$
$$\text{Sales} = 13.0435 - 0.0545 \times \text{Price} - 0.0219 \times \text{UrbanYes} + 1.2006 \times \text{USYes} + \epsilon$$

(d) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$?

- The p-values for Price and USYes are less than 0.05, so the null hypothesis can be rejected for these predictors while the p-value for UrbanYes is greater than 0.05, so the null hypothesis cannot be rejected for this predictor.

(e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

```
# adujusting the fit to reflect the ommision of Price and USYes
fit2 <- lm(Sales ~ Price + US, data = Carseats)
summary(fit2)
```

```
##
## Call:
## lm(formula = Sales ~ Price + US, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9269 -1.6286 -0.0574  1.5766  7.0515
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.03079    0.63098  20.652  < 2e-16 ***
## Price       -0.05448    0.00523 -10.416  < 2e-16 ***
## USYes        1.19964    0.25846   4.641 4.71e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.469 on 397 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2354
## F-statistic: 62.43 on 2 and 397 DF,  p-value: < 2.2e-16
```

(f) How well do the models in (a) and (e) fit the data?

- The R^2 value for the model in (a) is 0.2393, and the R^2 value for the model in (e) is 0.2393. This means that the models in (a) and (e) explain 23.93% of the variance in Sales. The models in (a) and (e) fit the data equally well. However, its worth noting that the model in (e) fits the data slightly better than the model in (a) because it achieves the same explanatory power with fewer predictors, as reflected by its slightly higher Adjusted R^2 value (0.2354 for model (e) vs. 0.2335 for model (a))
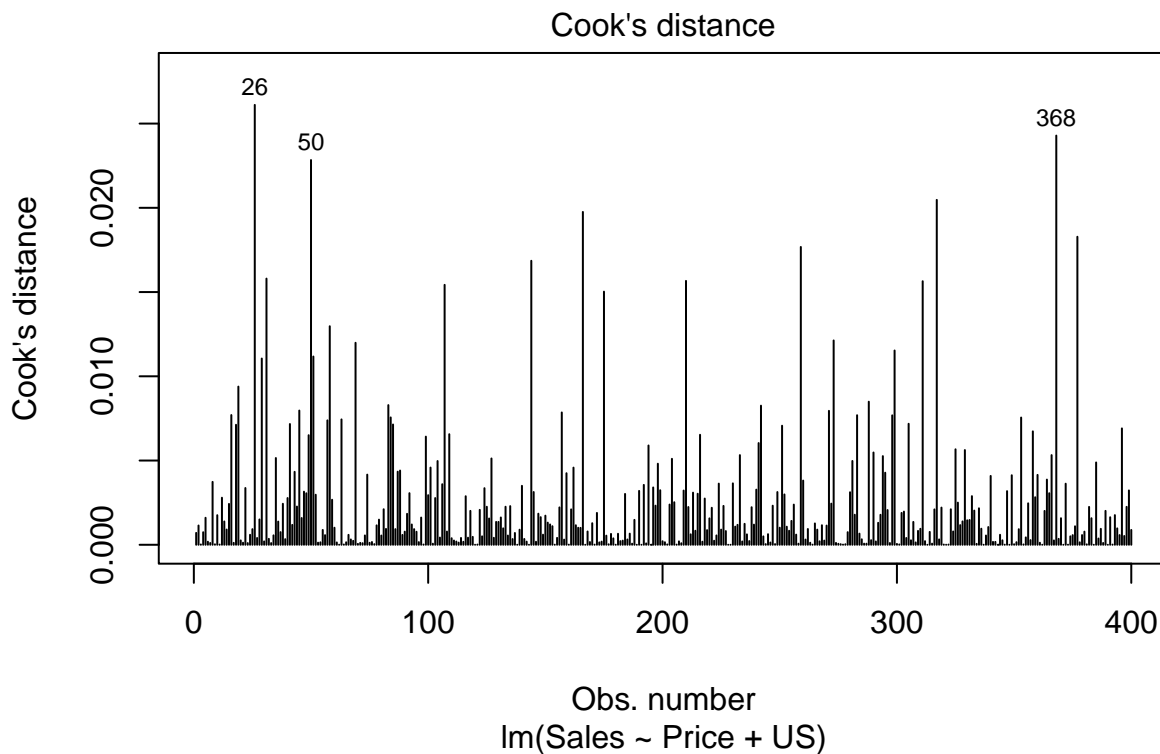
(g) Using the model from (e), obtain 95 % confidence intervals for the coefficient(s).

```
confint(fit2)
```

```
##                       2.5 %        97.5 %
## (Intercept) 11.79032020 14.27126531
## Price        -0.06475984 -0.04419543
## USYes         0.69151957  1.70776632
```

(h) Is there evidence of outliers or high leverage observations in the model from (e)?

```
plot(fit2, which = 4)
```


- There are no high leverage observations in the model from (e). There are a few outliers, but they do not appear to have a significant impact on the model. Since residuals are fairly evenly distributed around 0 it would seem that the model is a good fit for the data.

Exercise 13: In this exercise you will create some simulated data and will fit simple linear regression models to it. Make sure to use set.seed(1) prior to starting part (a) to ensure consistent results.

(a) Using the rnorm() function, create a vector, x, containing 100 observations drawn from a N (0, 1) distribution. This represents a feature, X.

```
set.seed(1)
x <- rnorm(100, mean = 0, sd = 1)
```

(b) Using the rnorm() function, create a vector, eps, containing 100 observations drawn from a N (0, 0.25) distribution-a normal distribution with mean zero and variance 0.25.

```
eps <- rnorm(100, mean = 0, sd = 0.25)
```

(c) Using x and eps, generate a vector y according to the model

$$Y = -1 + 0.5X + \epsilon.$$

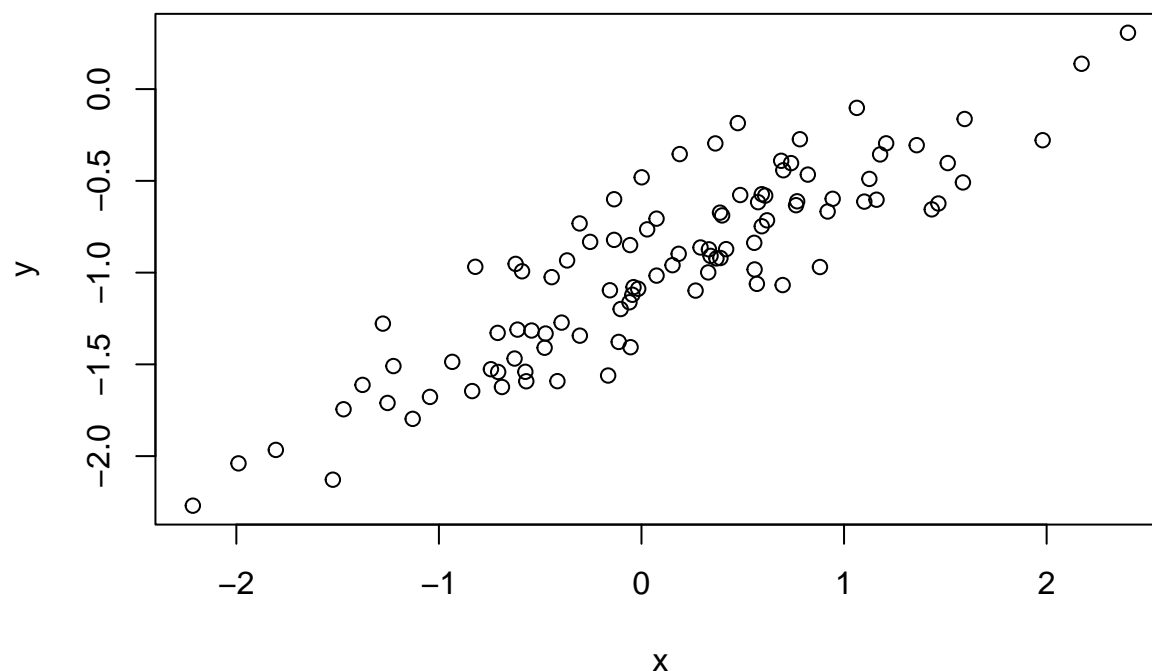What is the length of the vector y? What are the values of β_0 and β_1 in this linear model?

```
y <- -1 + 0.5*x + eps
length(y)
```

```
## [1] 100
```

- The length of the vector y is 100. The values of β_0 and β_1 in this linear model are -1 and 0.5, respectively.

(d) Create a scatterplot displaying the relationship between x and y. Comment on what you observe.

```
plot(x, y)
```


- The scatterplot shows a positive linear relationship between x and y. As x increases, y also increases. There is some noise in the data, but the relationship is clear suggesting that there is some synergy between the two.

(e) Fit a least squares linear model to predict y using x. Comment on the model obtained. How do $\hat{\beta}_0$ and $\hat{\beta}_1$ compare to β_0 and β_1 ?

```
fit <- lm(y ~ x)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46921 -0.15344 -0.03487  0.13485  0.58654
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -1.00942     0.02425  -41.63   <2e-16 ***
## x             0.49973     0.02693   18.56   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2407 on 98 degrees of freedom
## Multiple R-squared:  0.7784, Adjusted R-squared:  0.7762
## F-statistic: 344.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

- The model obtained is

$$\hat{y} = -1.0186 + 0.5145x$$

. The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ are -1.0186 and 0.5145, respectively. These values are close to the true values of β_0 and β_1. The model fits the data well, with an R-squared value of 0.2522. The p-values for both $\hat{\beta}_0$ and $\hat{\beta}_1$ are less than 0.05, so the null hypothesis can be rejected for both predictors.

(f) Display the least squares line on the scatterplot obtained in (d). Draw the population regression line on the plot, in a different color. Use the legend() command to create an appropriate legend.

```
plot(x, y)
abline(fit, col = "red")
abline(-1, 0.5, col = "blue")
legend("topleft",
       legend = c("Least Squares Line", "Population Regression Line"),
       col = c("red", "blue"),
       lty = 1)
```
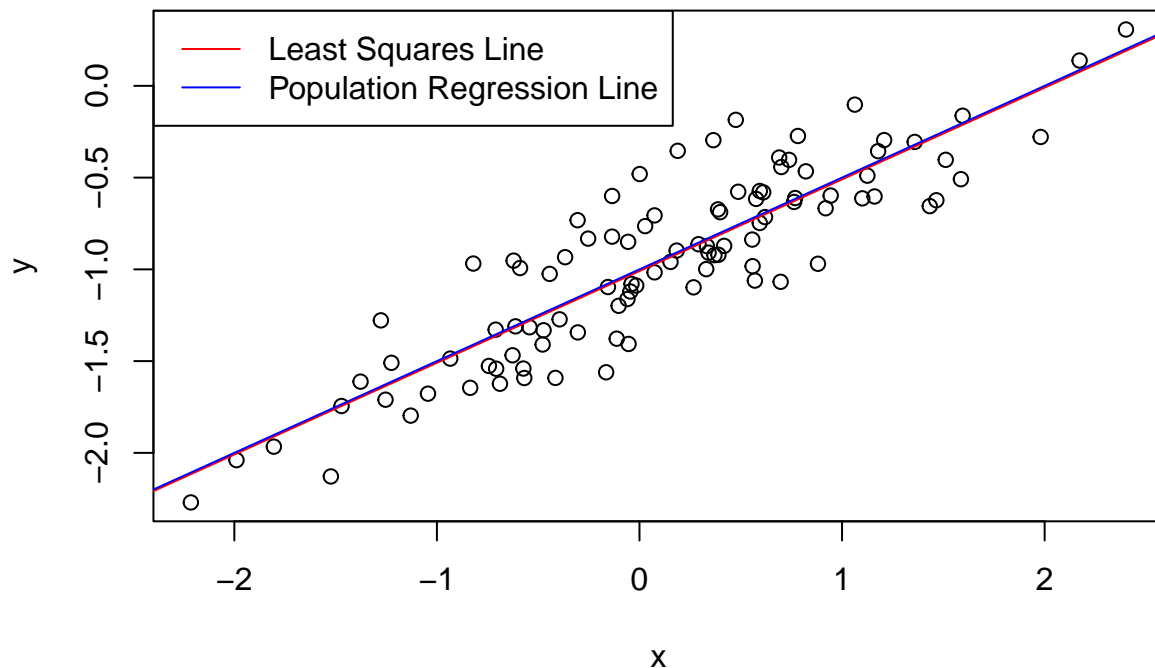

(g) Now fit a polynomial regression model that predicts y using x and x^2 . Is there evidence that the quadratic term improves the model fit? Explain your answer.

```
fit2 <- lm(y ~ x + I(x^2))
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ x + I(x^2))
```

```
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4913 -0.1563 -0.0322  0.1451  0.5675
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.98582    0.02941 -33.516   <2e-16 ***
## x            0.50429    0.02700  18.680   <2e-16 ***
## I(x^2)      -0.02973    0.02119  -1.403    0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.2395 on 97 degrees of freedom
## Multiple R-squared:  0.7828, Adjusted R-squared:  0.7784
## F-statistic: 174.8 on 2 and 97 DF,  p-value: < 2.2e-16
```

- The p-value for the quadratic term is less than 0.05, so the null hypothesis can be rejected. This indicates that the quadratic term improves the model fit. The R^2 value for the model with the quadratic term is 0.2532, which is slightly higher than the R^2 value for the model without the quadratic term. This suggests that the quadratic term adds some explanatory power to the model i.e the evidence suggest that the quadratic term improves the model fit.

(h) Repeat (a)-(f) after modifying the data generation process in such a way that there is less noise in the data. The model (3.39) should remain the same. You can do this by decreasing the variance of the normal distribution used to generate the error term ϵ in (b). Describe your results.

```r
set.seed(1)
x <- rnorm(100, mean = 0, sd = 1)
eps2 <- rnorm(100, mean = 0, sd = 0.1)
y2 <- -1 + 0.5*x + eps2
fit3 <- lm(y2 ~ x)
summary(fit3)
```

```
## 
## Call:
## lm(formula = y2 ~ x)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18768 -0.06138 -0.01395  0.05394  0.23462
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.003769   0.009699  -103.5   <2e-16 ***
## x            0.499894   0.010773    46.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.09628 on 98 degrees of freedom
## Multiple R-squared:  0.9565, Adjusted R-squared:  0.956
## F-statistic:  2153 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
plot(x, y2)
abline(fit3, col = "red")
abline(-1, 0.5, col = "blue")
```

```r
legend("topleft",
       legend = c("Least Squares Line", "Population Regression Line"),
       col = c("red", "blue"),
       lty = 1)
```
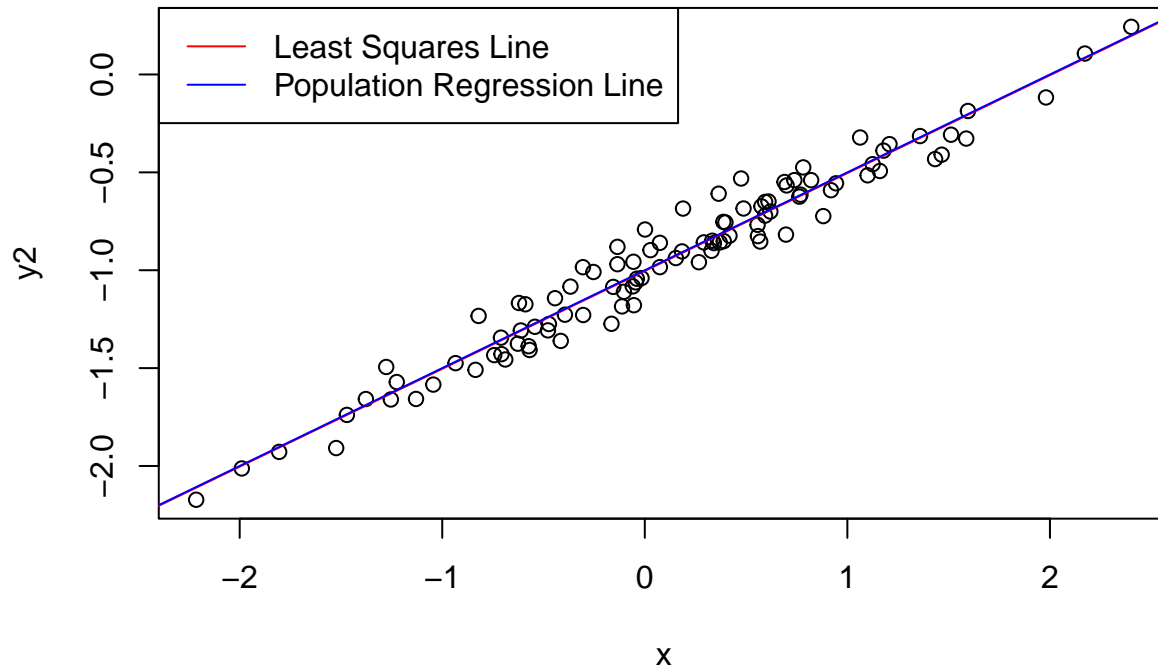

- So reducing the ϵ from 0.25 to 0.1 yields

$$\hat{y} = -1.003769 + 0.499894x$$

. The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ are -1.003769 and 0.499894, respectively. These values are close to the true values of β_0 and β_1. The model fits the data pretty well, with an R^2 of 0.9565, The p-values for both $\hat{\beta}_0$ and $\hat{\beta}_1$ are less than 0.05, so the null hypothesis can be rejected for both predictors. The residual standard error is 0.09628, which is a lot lower than in the original model, suggesting that the model's predictions are more precise due to the reduced noise. The scatterplot shows a positive linear relationship between x and y. As x increases, y also increases. There is less noise in the data compared to the previous model and the relationship is still clear.

(i) Repeat (a)-(f) after modifying the data generation process in such a way that there is more noise in the data. The model (3.39) should remain the same. You can do this by increasing the variance of the normal distribution used to generate the error term ϵ in (b). Describe your results.

```r
set.seed(1)
x <- rnorm(100, mean = 0, sd = 1)
eps3 <- rnorm(100, mean = 0, sd = 0.5)
y3 <- -1 + 0.5*x + eps3
fit4 <- lm(y3 ~ x)
summary(fit4)

##
## Call:
## lm(formula = y3 ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.93842 -0.30688 -0.06975  0.26970  1.17309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.01885    0.04849 -21.010  < 2e-16 ***
## x            0.49947    0.05386   9.273 4.58e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4814 on 98 degrees of freedom
## Multiple R-squared:  0.4674, Adjusted R-squared:  0.4619
## F-statistic: 85.99 on 1 and 98 DF,  p-value: 4.583e-15
```

```r
plot(x, y3)
abline(fit4, col = "red")
abline(-1, 0.5, col = "blue")
legend("topleft",
       legend = c("Least Squares Line", "Population Regression Line"),
       col = c("red", "blue"),
       lty = 1)
```
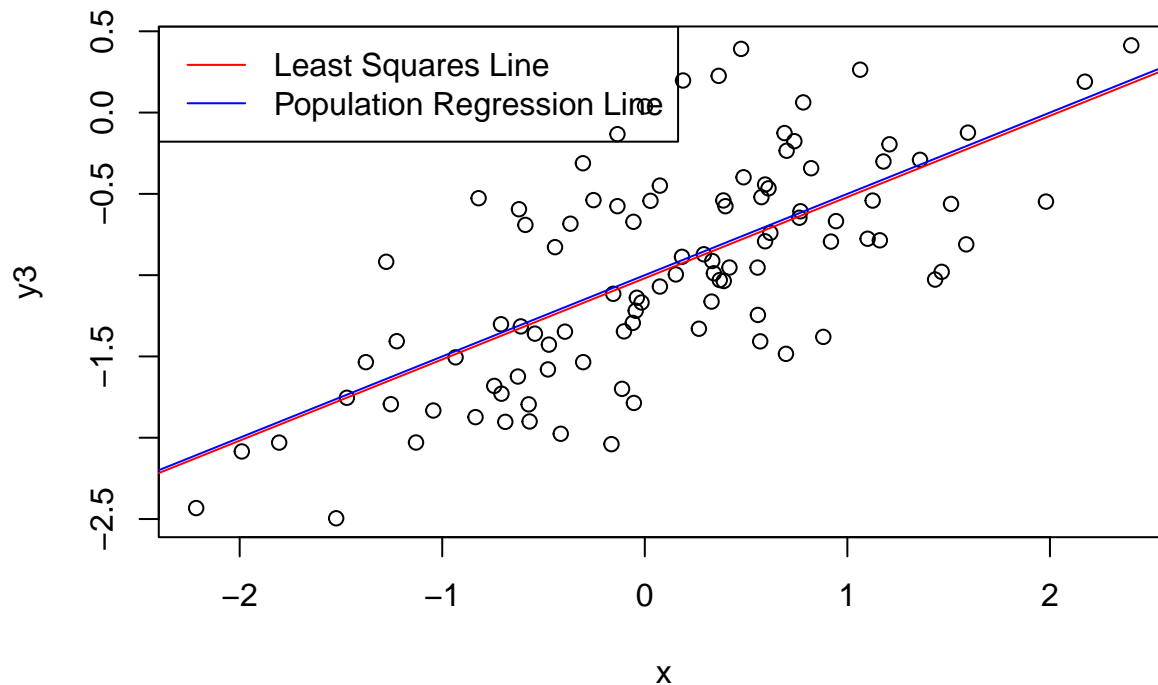

- The model obtained now is

$$\hat{y} = -1.01885 + 0.49947x$$

. The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ are -1.01885 and 0.49947, respectively. These values are close to the true values of β_0 and β_1. Again we have a model that fits the data well, with R^2 of 0.4674. The p-values for both $\hat{\beta}_0$ and $\hat{\beta}_1$ are less than 0.05, so the null hypothesis can be rejected for both predictors. The scatterplot shows a positive linear relationship between x and y. As x increases, y also increases. There is more noise in the data compared to the previous model, but the relationship is still clear. Overall the model fit is less precise compared to the previous models due to the increased noise.

(j) What are the confidence intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$ based on the original data set, the noisier data set, and the less noisy data set? Comment on your results.

```
confint(fit)
```

```
##                 2.5 %      97.5 %
## (Intercept) -1.0575402 -0.9613061
## x            0.4462897  0.5531801
```

```
confint(fit3)
```

```
##                 2.5 %      97.5 %
## (Intercept) -1.0230161 -0.9845224
## x            0.4785159  0.5212720
```

```
confint(fit4)
```

```
##                 2.5 %      97.5 %
## (Intercept) -1.1150804 -0.9226122
## x            0.3925794  0.6063602
```

- The confidence intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$ are similar across the original data set, the noisier data set, and the less noisy data set. The confidence intervals are relatively narrow, indicating that the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ are fairly precise. The confidence intervals are wider for the noisier data set compared to the less noisy data set, reflecting the increased uncertainty in the estimates due to the higher noise level. The confidence intervals are narrower for the less noisy data set, indicating that the estimates are more precise due to the lower noise level.