A1Q6

Undergraduate Student

(a)

```r
x <- rnorm(100, mean = 0, sd = 1)
```

(b)

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

(c)

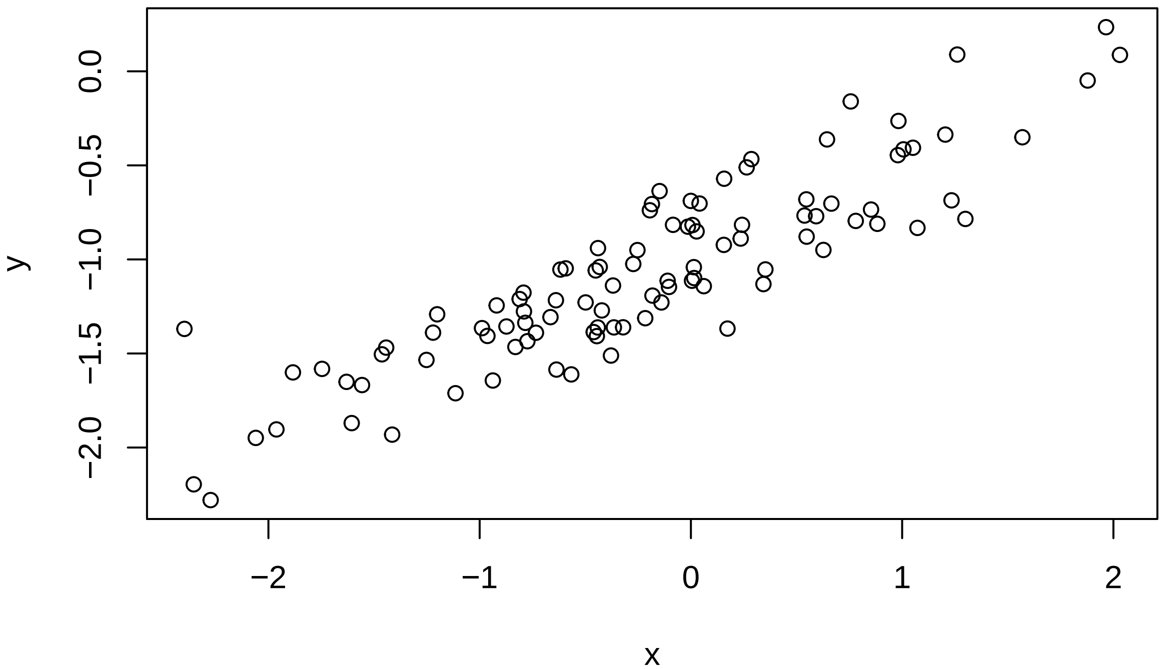
```r
y <- -1 + 0.5*x + eps
length(y)
```

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

Comment: The length of the vector y is 100. The value of β_0 is -1. The value of β_1 is 0.5.

(d)

```r
plot(x, y, xlab = "x", ylab = "y")
```



Comment: the scatter plot indicates a possible linear relationship between x and y.

(e)

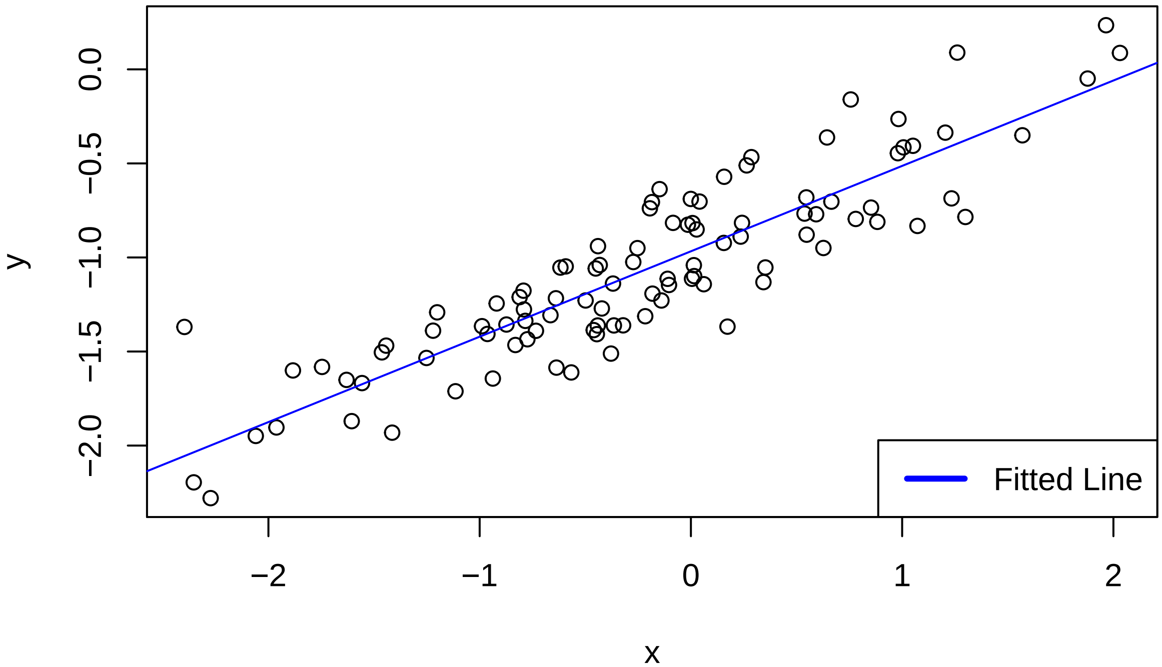
```
model1 <- lm(y~x)
summary(model1)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47915 -0.16384 -0.00315  0.14181  0.68629
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.96728    0.02299  -42.07   <2e-16 ***
## x            0.45398    0.02353   19.30   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.224 on 98 degrees of freedom
## Multiple R-squared:  0.7916, Adjusted R-squared:  0.7895
## F-statistic: 372.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

Comment: $\hat{\beta}_0$ and $\hat{\beta}_1$ are slightly different from β_0 and β_1, but they are nearly the same. We will say that the fitted model is good estimate to the data.

(f)

```r
plot(x, y, xlab = "x", ylab = "y")
abline(model1, col = "blue")
legend("bottomright",c("Fitted Line"), col = "blue", lwd=3)
```



(g)

```r
model2 <- lm(y~poly(x, 2, raw = TRUE))
summary(model2)
```

```
##
## Call:
## lm(formula = y ~ poly(x, 2, raw = TRUE))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46192 -0.17528 -0.00547  0.14659  0.59844
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.98697    0.02765 -35.699   <2e-16 ***
## poly(x, 2, raw = TRUE)1  0.46423    0.02480  18.721   <2e-16 ***
## poly(x, 2, raw = TRUE)2  0.02299    0.01805   1.274    0.206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

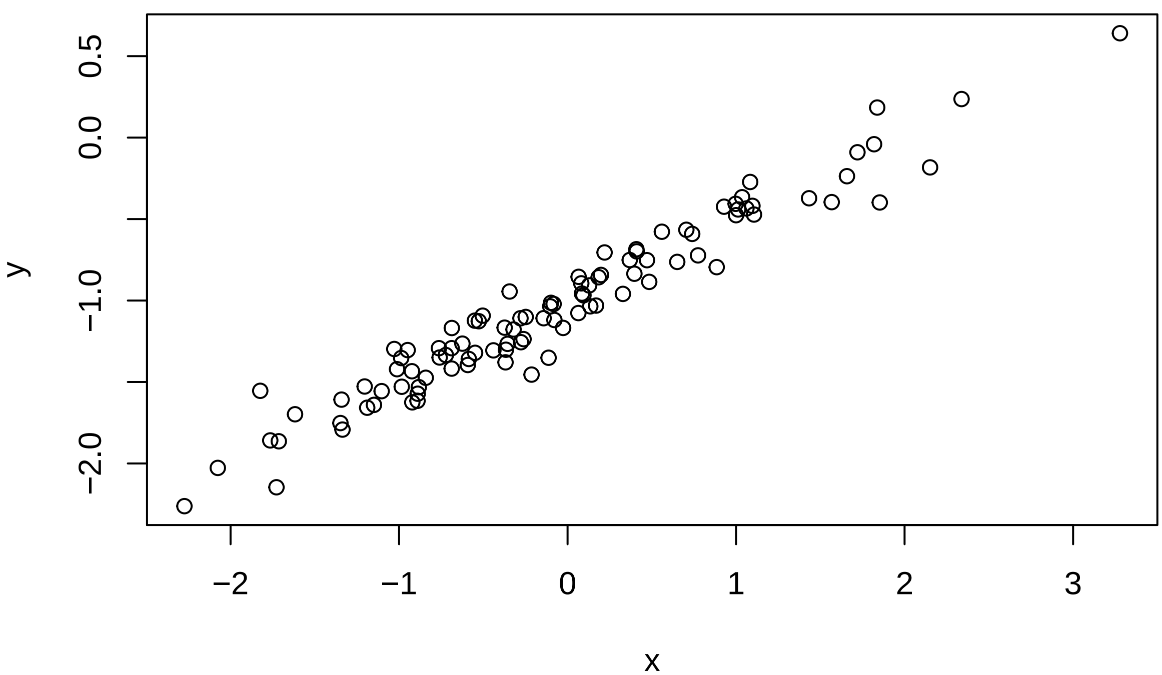
```
## Residual standard error: 0.2233 on 97 degrees of freedom
## Multiple R-squared:  0.795,  Adjusted R-squared:  0.7908
## F-statistic: 188.1 on 2 and 97 DF,  p-value: < 2.2e-16
```

Comment: From the summary, we see that the p-value of the intercept and the x term is pretty small, but the quadratic term is large. Therefore, the quadratic term is not significant and we tends to reject it. This means that there is no evidence that the model is improved.

(h)

```r
x1 <- rnorm(100, mean = 0, sd = 1)
eps1 <- rnorm(100, mean = 0, sd = 0.15)
y1 <- -1 + 0.5*x1 + eps1
```

```r
plot(x1, y1, xlab = "x", ylab = "y")
```

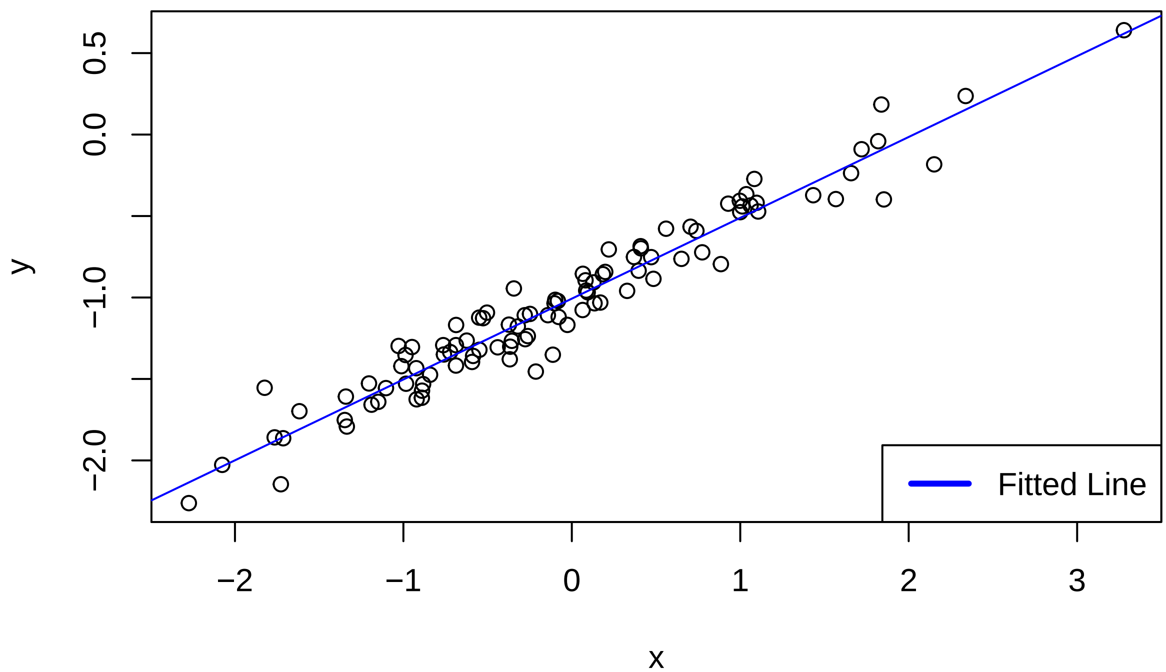



```r
model3 <- lm(y1~x1)
summary(model3)
```

```
##
## Call:
## lm(formula = y1 ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34156 -0.08323  0.02177  0.07378  0.35782
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.00700    0.01285  -78.39   <2e-16 ***
## x1           0.49615    0.01242   39.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1282 on 98 degrees of freedom
## Multiple R-squared:  0.9422, Adjusted R-squared:  0.9416
## F-statistic:  1597 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
plot(x1, y1, xlab = "x", ylab = "y")
abline(model3, col = "blue")
legend("bottomright",c("Fitted Line"), col = "blue", lwd=3)
```




```r
model4 <- lm(y1~poly(x1, 2, raw = TRUE))
summary(model4)
```

```
##
## Call:
## lm(formula = y1 ~ poly(x1, 2, raw = TRUE))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34213 -0.08359  0.02355  0.07341  0.35952
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.0063565  0.0159400 -63.134   <2e-16 ***
## poly(x1, 2, raw = TRUE)1  0.4963691  0.0128882  38.513   <2e-16 ***
## poly(x1, 2, raw = TRUE)2 -0.0005843  0.0085157  -0.069    0.945
## ---
```

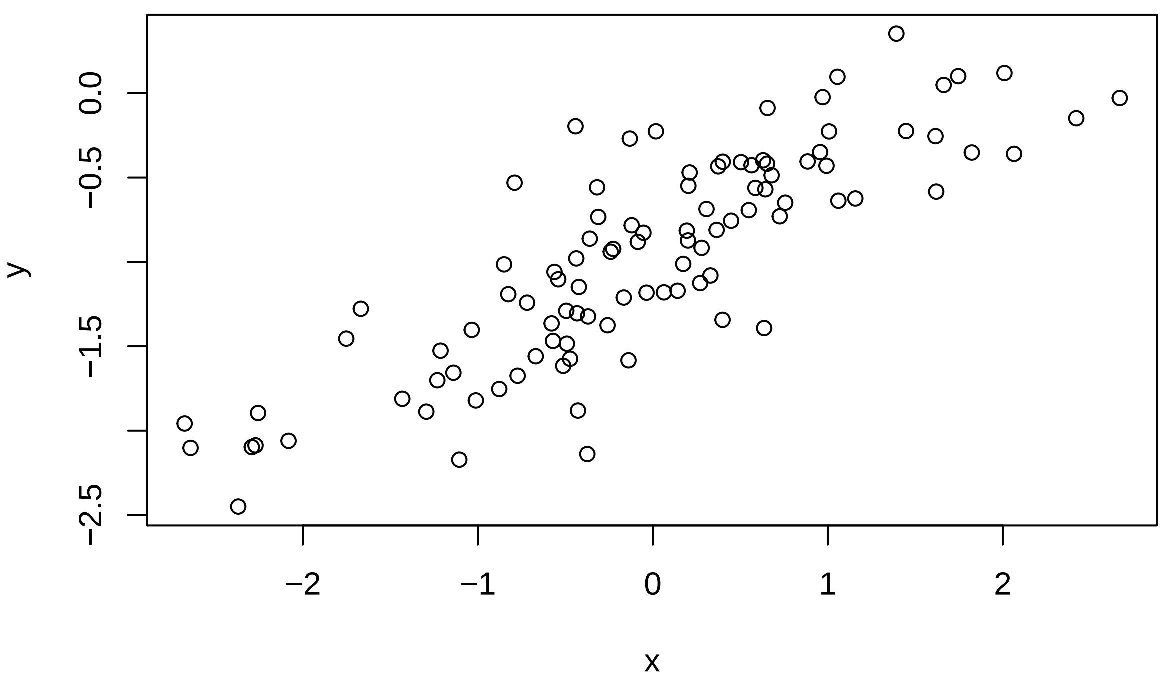
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1288 on 97 degrees of freedom
## Multiple R-squared:  0.9422, Adjusted R-squared:  0.941
## F-statistic: 790.3 on 2 and 97 DF,  p-value: < 2.2e-16
```

Comment: After noise being reduced, we see a decrease in residual standard error and increase in R-squared value. This means that the model fits better than the original one, as dots are closer to the fitted line in the plot. However, the quadratic term still has a large p-value. This means that the quadratic term again is not significant. There is no evidence that the polynomial regression model improves the result.

(i)

```
x2 <- rnorm(100, mean = 0, sd = 1)
eps2 <- rnorm(100, mean = 0, sd = 0.35)
y2 <- -1 + 0.5*x2 + eps2
```

```
plot(x2, y2, xlab = "x", ylab = "y")
```

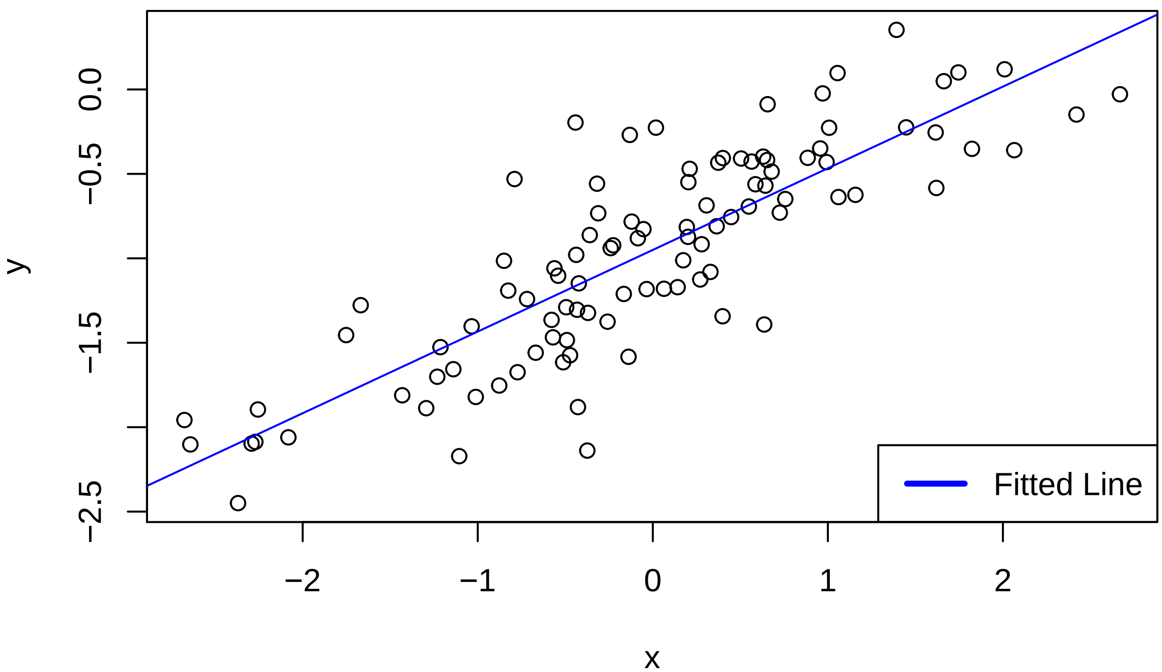



```
model5 <- lm(y2~x2)
summary(model5)
```

```
##
## Call:
## lm(formula = y2 ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.00744 -0.24670  0.00863  0.20950  0.96800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.95035    0.03511  -27.07   <2e-16 ***
## x2           0.48358    0.03215   15.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3508 on 98 degrees of freedom
## Multiple R-squared:  0.6978, Adjusted R-squared:  0.6947
## F-statistic: 226.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
plot(x2, y2, xlab = "x", ylab = "y")
abline(model5, col = "blue")
legend("bottomright",c("Fitted Line"), col = "blue", lwd=3)
```




```
model6 <- lm(y2~poly(x2, 2, raw = TRUE))
summary(model6)
```

```
##
## Call:
## lm(formula = y2 ~ poly(x2, 2, raw = TRUE))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.02243 -0.25052  0.00968  0.21296  0.95353
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -0.93463    0.04264 -21.917   <2e-16 ***
```

```
## poly(x2, 2, raw = TRUE)1  0.48055    0.03257  14.754   <2e-16 ***
## poly(x2, 2, raw = TRUE)2 -0.01328    0.02032  -0.653    0.515
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3519 on 97 degrees of freedom
## Multiple R-squared:  0.6991, Adjusted R-squared:  0.6929
## F-statistic: 112.7 on 2 and 97 DF,  p-value: < 2.2e-16
```

Comment: After noise being increased, we see an increase in residual standard error and decrease in R-squared value. This means that the model fits worse than the original one, as dots are further from the fitted line in the plot. However, the quadratic term still has a large p-value. This means that the quadratic term again is not significant. There is no evidence that the polynomial regression model improves the result.

(j)

Original data set:

β_0:

```
summary(model1)$coefficients[1,1]+c(-1,1)*summary(model1)$coefficients[1,2]*qnorm(0.975)
```

```
## [1] -1.0123473 -0.9222147
```

β_1:

```
summary(model1)$coefficients[2,1]+c(-1,1)*summary(model1)$coefficients[2,2]*qnorm(0.975)
```

```
## [1] 0.4078625 0.5000915
```

Less noisy data set:

β_0:

```
summary(model3)$coefficients[1,1]+c(-1,1)*summary(model3)$coefficients[1,2]*qnorm(0.975)
```

```
## [1] -1.0321767 -0.9818188
```

β_1:

```
summary(model3)$coefficients[2,1]+c(-1,1)*summary(model3)$coefficients[2,2]*qnorm(0.975)
```

```
## [1] 0.4718135 0.5204829
```

Noisier data set:

β_0:

```r
summary(model5)$coefficients[1,1]+c(-1,1)*summary(model5)$coefficients[1,2]*qnorm(0.975)
```

```
## [1] -1.0191633 -0.8815415
```

β_1:

```r
summary(model5)$coefficients[2,1]+c(-1,1)*summary(model5)$coefficients[2,2]*qnorm(0.975)
```

```
## [1] 0.4205748 0.5465847
```