SGupta_HW04Question3

Quarto

Simulate data from a linear model: y = 5 + 3*x + error, error ~ N(0, 2^2) Fit the linear model

```
# Load necessary libraries
set.seed(123)
n <- 100
x <- runif(n, 0, 10)
y <- 5 + 3*x + rnorm(n, 0, 2) # true model: y = 5 + 3x + error
lm_fit <- lm(y ~ x)
```

Choose a new predictor value (e.g., x_new = 5)

```
newdata <- data.frame(x = 5)
```

Get the 95% confidence interval for the mean response at x = 5

```
conf_int <- predict(lm_fit, newdata, interval = "confidence", level = 0.95)
```

Get the 95% prediction interval for a new observation at x = 5

```
pred_int <- predict(lm_fit, newdata, interval = "prediction", level = 0.95)
```

```
print("Confidence Interval for E(y_new):")
```

```
[1] "Confidence Interval for E(y_new):"
```

```
print(conf_int)
```

```
       fit      lwr      upr
1 19.89225 19.50754 20.27696
```

```
print("Prediction Interval for a new y_new:")
```

```
[1] "Prediction Interval for a new y_new:"
```

```
print(pred_int)
```

```
       fit      lwr      upr
1 19.89225 16.02601 23.75849
```

For the linear model y = 5 + 3*x + error, when x = 5 the 95% confidence interval for the mean response is very narrow (approximately [19.51, 20.28]). In contrast, the 95% prediction interval for an individual new observation at x = 5 is much wider (approximately [16.03, 23.76]). This shows that while the estimate of the mean is quite same, a new observation includes additional random error, making its interval wider.

Simulate data from a quadratic model: y = 1 + 2x + $3x^2$ + error, error ~ N(0, 1)

```
n <- 100
x <- runif(n, -2, 2)
y_true <- 1 + 2*x + 3*x^2 + rnorm(n, 0, 1)
```
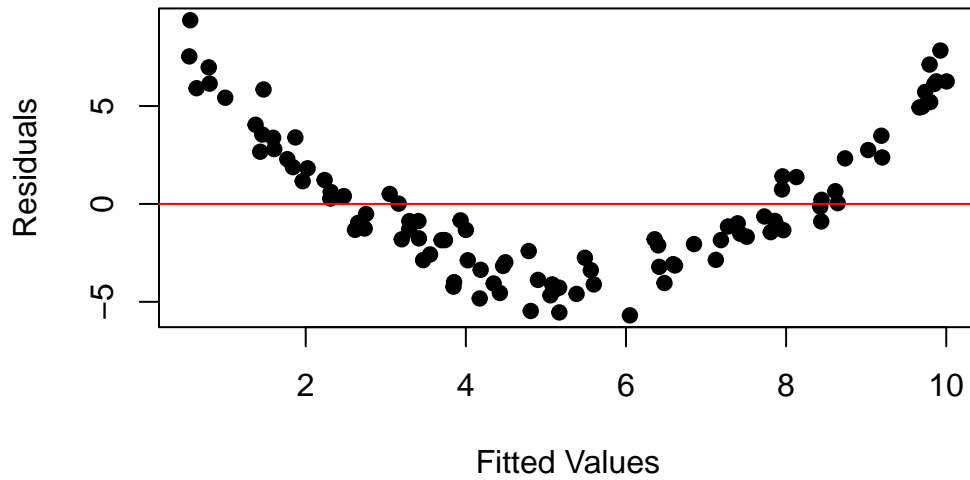
Fit a linear model that omits the quadratic term (i.e., y ~ x)

```
lm_wrong <- lm(y_true ~ x)
```

Plot residuals versus fitted values to reveal nonlinearity

```
par(mfrow = c(1,1))
plot(fitted(lm_wrong),
     resid(lm_wrong),
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted Values (Misspecified Model)",
     pch = 19)
abline(h = 0, col = "red")
```

Residuals vs Fitted Values (Misspecified Model)

The resulting plot of residuals versus fitted values shows a clear curved (nonlinear) pattern. This systematic curvature indicates that the model is misspecified because the true relationship is quadratic, the residuals absorb the unmodeled curvature.