Lab 8

AA 501

Kristen Fitzgerald

2024-07-08

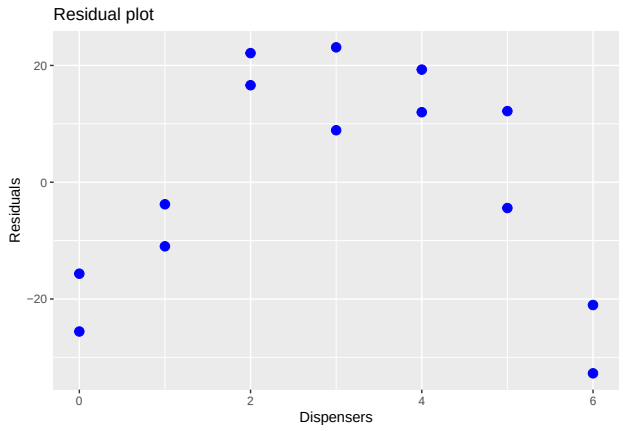
Question 1

Perform a simple linear regression with `sales` as the response and `dispensers` as the predictor variable. What do you see in the residual versus predicted plot? What would you do to fix this problem?

Solution: In the SLR model there appears to be a quadratic relationship between `dispensers` and the residuals. We can likely fix this problem by fitting a quadratic regression equation between `dispensers` and `sales`.

```
slr.sales <- lm(sales ~ dispensers)

ggplot(slr.sales, aes(x = dispensers, y = resid(slr.sales)))  +
  geom_point(color="blue",size=3) +
  labs(title="Residual plot", x="Dispensers", y="Residuals")
```



Question 2

Perform a forward selection (by hand) using the AIC criteria (you will need to use the command AIC(model) to get the AIC values for each model). The "smallest" model should be the just the intercept. The "biggest" model should be Dispensers up to the power of 4 (be sure to follow model hierarchy). What was the best degree for the polynomial based on AIC?

Solution: The best degree for the polynomial was 2 based on the AIC.

```
int.model <- lm(sales ~ 1)

paste0("Intercept Model AIC:", " ", round(AIC(int.model), 2))
```

```
## [1] "Intercept Model AIC: 176.07"
```

```
model1 <- lm(sales ~ dispensers)

paste0("Model 1 AIC:", " ", round(AIC(model1), 2))
```

```
## [1] "Model 1 AIC: 126.88"
```

```
model2 <- lm(sales ~ dispensers + I(dispensers^2))

paste0("Model 2 AIC:", " ", round(AIC(model2), 2))
```

```
## [1] "Model 2 AIC: 101.98"
```

```
model3 <- lm(sales ~ dispensers + I(dispensers^2) + I(dispensers^3))

paste0("Model 3 AIC:", " ", round(AIC(model3), 2))
```

```
## [1] "Model 3 AIC: 102.6"
```

```
model4 <- lm(sales ~ dispensers + I(dispensers^2) + I(dispensers^3) +
             I(dispensers^4))

paste0("Model 4 AIC:", " ", round(AIC(model4), 2))
```

```
## [1] "Model 4 AIC: 104.55"
```

Question 3

Run the model you selected in question 2 and look at the residual versus predicted plot. What do you see?

Solution: The residuals quadratic pattern has nearly disappeared and it appears to hold the assumption of homoscedasticity.

```
slr.sales.quad <- lm(sales ~ dispensers + I(dispensers^2))

ggplot(slr.sales.quad, aes(x = dispensers, y = resid(slr.sales.quad)))  +
  geom_point(color="blue",size=3) +
  labs(title="Residual plot", x="Dispensers", y="Residuals")
```