STA 471 - Homework 9

Richard McCormick

2023-12-03

1. Problem T on page 325 of the textbook.

```
X <- c( 18, 20, 24, 28, 30, 33, 36, 48, 60, 40, 42, 45, 62, 71, 75 )
Y <- c( 4.8, 5.5, 5.8, 6.0, 6.5, 6.6, 6.7, 7.0, 7.3, 3.3, 3.8, 4.1, 5.0, 5.5, 6.0 )
D <- c( 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1 )

hw9.data <- data.frame( Y, X, D )
```

a. Define a suitable numerical dummy variable D to seperate the two groups(f and m) of data. Fit the model $Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 DX + \epsilon$ and give (1) the fitted equation, (2) the fitted line for each group (f and m) based on the fitted equation.

```
hw9.model <- lm( Y ~ X + D + I( D * X ) )
summary( hw9.model )
```

```
##
## Call:
## lm(formula = Y ~ X + D + I(D * X))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65980 -0.09839  0.01710  0.15928  0.41248
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.518222   0.292245  15.460 8.29e-09 ***
## X            0.052310   0.008258   6.335 5.56e-05 ***
## D           -3.640652   0.603329  -6.034 8.50e-05 ***
## I(D * X)     0.014659   0.012337   1.188     0.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3168 on 11 degrees of freedom
## Multiple R-squared:  0.9449, Adjusted R-squared:  0.9298
## F-statistic: 62.85 on 3 and 11 DF,  p-value: 3.292e-07
```

1. Give the Fitted Equation

The fitted equation for the model is:
$Y = 4.518222 + 0.052310X - 3.640652D + 0.014659DX$

2. Give the Fitted Line for each group (f and m) based on the fitted equation.

The fitted line for **f** is:
$Y = 4.518222 + 0.052310X$

The fitted line for **m** is:
$Y = 0.877572 + 0.066969X$

b. Test to determine whether the two straight lines are parallel at $\alpha = 0.05$.

I. Hypothesis

$H_0 : \beta_3 = 0.$ The lines are parallel.
$H_A : \beta_3 \neq 0.$ The lines are not parallel.

II. Test Statistics

Test Statistic: $t = \frac{b_3 - \beta_3}{se(b3)}$
Observed Statistic: $t = 1.188$

Critical T-Value ($t_{15-3-1} = t_{11}$): 2.201

III. Conclusion

As $t_{obs} = 1.188 < 2.201 = t_{11}$, we can **accept** the null hypothesis and conclude that the lines are parallel with each other.

c. Test to determine whether the two straight lines are coincident at $\alpha = 0.05$.

I. Hypothesis

$H_0 : \beta_2 = \beta_3 = 0.$ The lines are coincident.
$H_A : At\ least\ one\ of \beta_2, \beta_3 \neq 0.$ The lines are not coincident.

II. Test Statistics

Test Statistic: $F = \frac{MS_{reg}}{MS_{resid}}$ Observed Statistic: $F_{obs} = 62.85$

Critical f-Value ($f_{3,11}$): 2.66

III. Conclusion

As $f_{obs} = 62.85 > 2.66 = f_{3,11}$, we can **reject** the null hypothesis and accept the alternative hypothesis that the lines are not coincident.

d. Fit seperate straight-line regressions of Y on X for f and m, respectively. How do the fitted lines here compare to the fitted lines in part a?

```
hw9.data.f <- subset( hw9.data, D == 0 )
hw9.model.f <- lm( Y ~ X, data=hw9.data.f )
summary( hw9.model.f )
```

```
##
## Call:
## lm(formula = Y ~ X, data = hw9.data.f)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65980 -0.06442  0.01710  0.29863  0.41248
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.518222   0.340607  13.265 3.24e-06 ***
## X           0.052310   0.009624   5.435 0.000971 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3692 on 7 degrees of freedom
## Multiple R-squared:  0.8084, Adjusted R-squared:  0.7811
## F-statistic: 29.54 on 1 and 7 DF,  p-value: 0.0009711
```

The fitted model for *just* **f** is: $Y = 4.518222 + 0.052310X$
The model in part **a** was: $Y = 4.518222 + 0.052310X$

The two models are identical.

```
hw9.data.m <- subset( hw9.data, D == 1 )
hw9.model.m <- lm( Y ~ X, data=hw9.data.m )
summary( hw9.model.m )
```

```
##
## Call:
## lm(formula = Y ~ X, data = hw9.data.m)
##
## Residuals:
##       10       11       12       13       14       15
## -0.25633  0.10974  0.20883 -0.02964 -0.13236  0.09976
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.877570   0.322311   2.723  0.05284 .
## X           0.066969   0.005597  11.966  0.00028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1935 on 4 degrees of freedom
## Multiple R-squared:  0.9728, Adjusted R-squared:  0.966
## F-statistic: 143.2 on 1 and 4 DF,  p-value: 0.0002795
```

The fitted model for *just* **m** is: $Y = 0.877570 + 0.066969X$
The model in part **a** was: $Y = 0.877572 + 0.066969X$

The two models are identical.