R Notebook

Code

Harrison Webb MATH 3080 5 May, 2022

Chapter 13-2 homework

Problem 32

a. What is the estimated regression function for the "centered" model?

$$y = .3463 - 1.2933(x - 4.3456) + 2.3964(x - 4.3456)^2 - 2.3968(x - 4.3456)^3$$

b. What is the estimated value of the coefficient β_3 in the 'uncentered' model with regression function $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$? What is the estimate of β_2?

$y = .3463 - 1.2933(x - 4.3456) + 2.3964(x - 4.3456)^2 - 2.3968(x - 4.3456)^3$ $y = 247.91 - 157.906x + 33.643x^2 - 2.3968x^3$
$\beta_3 = -2.3968, \beta_2 = 33.643$

c. Using the cubic model, what value of y would you predict when soil pH is 4.5?

$y = .3463 - 1.2933(x - 4.3456) + 2.3964(x - 4.3456)^2 - 2.3968(x - 4.3456)^3$
$y = .3463 - 1.2933(4.5 - 4.3456) + 2.3964(4.5 - 4.3456)^2 - 2.3968(4.5 - 4.3456)^3$
$y = 0.1949$

d. Carry out a test to decide whether the cubic term should be retained in the model.

$H_0 : \hat{\beta}_3 = 0$
$H_a : \hat{\beta}_3 \neq 0$
$t = \frac{\hat{\beta}_3}{s_{\hat{\beta}_3}} = \frac{-2.3968}{2.4590} = -0.9747$
$df = n - (k + 1) = 16 - (3 + 1) = 12$

Hide

```
2*pt(-0.9747, df=12)
```

```
[1] 0.3489487
```

With a p-value of 0.35, we do not reject H_0 and do not have evidence that $\hat{\beta}_3$ should be retained.

Problem 36

a. Interpret β_1 and β_3.

β_1 : a 1kg increase in weight correlates to an increase of VO2 max of 0.1
β_3 : a 1 min increase in time taken to walk 1 mile correlates to a decrease of VO2 max of 0.13

b. What is the expected value of VO2max when weight is 76 kg, age is 20 yr, walk time is 12 min, and heart rate is 140 b/m?

$Y = 5.0 + 0.01x_1 - 0.05x_2 - 0.13x_3 - 0.01x_4 + \epsilon$ $Y = 5.0 + 0.01 * 76 - 0.05 * 20 - 0.13 * 12 - 0.01 * 140$ $Y = 1.8$

c. What is the probability that VO2max will be between 1.00 and 2.60 for a single observation made when the values of the predictors are as stated in part (b)?

$P(1 \leq Y \leq 2.6) = P(\frac{1-\mu_Y}{\sigma} \leq \frac{Y-\mu_Y}{\sigma} \leq \frac{2.6-\mu_Y}{\sigma}) = P(\frac{1-1.8}{0.4} \leq Z \leq \frac{2.6-1.8}{0.4})$

Hide

```
pnorm(2.6, mean=1.8, sd=0.4) - pnorm(1, mean=1.8, sd=0.4)
```

```
[1] 0.9544997
```

Problem 39

y = sales at a fast-food outlet (1000s of $)
x1 = number of competing outlets in a 1-mile radius
x2 = population within a 1 mile radius (1000s of people)
x3 = 1 if outlet has a drive up window, 0 otherwise
True regression model is $Y = 10.00 - 1.2x_1 + 6.8x_2 + 15.3x_3 + \epsilon$

a. What is the mean value of sales when the number of competing outlets is 2, there are 8000 people within a 1-mile radius, and the outlet has a drive-up window?

$mean = 10 - 1.2 * 2 + 6.8 * 8 + 15.3 * 1$ $mean = 77.3$

b. What is the mean value of sales for an outlet without a drive-up window that has three competing outlets and 5000 people within a 1-mile radius?

$x_3 = 0, x_1 = 3, x_2 = 5$ $mean = 10 - 1.2 * 3 + 6.8 * 5 + 15.3 * 0$ $mean = 40.4$

c. Interpret β_3.

All else equal, the average sales of an outlet with a drive through window is $153000 more than one without.

Problem 49

y = ultimate deflection in mm of composite beams
x1 = shear span ratio
x2 = splitting tensile strength in MPa
$y = 17.3 - 6.37x_1 - 3.66x_2 + 1.71x_1x_2$

b. Should the interaction predictor be retained in the model? Carry out a test of hypotheses using a significance level of .05.

Yes, from the minitab output we see that the p-value for x1x2 is $0.002 \leq 0.05$

c. The estimated standard deviation of $\hat{Y}$ when x1=3 and x2=6 is $s_{\hat{Y}} = 0.555$. Calculate and interpret a confidence interval with a 95% confidence level for true average deflection under these circumstances.

$\hat{y} \pm t_{\alpha/2, n-(k+1)} s_{\hat{y}}$ $17.3 - 6.37(3) - 3.66(6) + 1.71(3)(6) \pm t_{.025,11} * 0.555$

Hide

```
qt(.025, 11)
```

```
[1] -2.200985
```

$7.01 \pm 2.201 * 0.555$ $(5.788, 8.232) $

d. Using the information in (c), calculate and interpret a prediction interval using a 95% confidence level for a future value of ultimate deflection to be observed when x1=3 and x2=6.

$ t_{/ a, n-(k+1)} * $ $7.01 * $ $(3.028, 10.992) $

Problem 67

y = VO2 max
x1 = gender (female=0, male=1)
x2 = weight in lbs
x3 = 1 mile walk time in min
x4 = HR at end of walk in bpm
$y = 3.5959 + 0.6566x_1 + 0.0096x_2 - 0.0996x_3 - 0.0080x_4 $

a. How would you interpret the estimated coefficient $\hat{\beta}_3 = -0.0996$?

a 1 minute increase in mile walk time corresponds to a decrease of -0.0996 in VO2 max on average

b. How would you interpret the estimated coefficient $\hat{\beta}_1 = 0.6566$?

on average, males have a VO2 max that is 0.6566 higher than females

c. Suppose that an observation made on a male whose weight was 170 lb, walk time was 11 min, and heart rate was 140 beats/min resulted in VO2max = 3.15. What would you have predicted for VO2max in this situation, and what is the value of the corresponding residual?

$E[y] = 3.5959 + 0.6566(1) + 0.0096(170) - 0.0996(11) - 0.0080(140) = 3.6689$
$residual = observed - expected = 3.15 - 3.6689 = -0.5189$

d. Using SSE = 30.1033 and SST = 102.3922, what proportion of observed variation in VO2max can be attributed to the model relationship?

$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{30.1033}{102.3922} = 0.706$

e. Assuming a sample size of n = 20, carry out a test of hypotheses to decide whether the chosen model specifies a useful relationship between VO2max and at least one of the predictors.

$f = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}} = \frac{\frac{0.706}{4}}{\frac{1-0.706}{20-4-1}} = 9.005$

Hide

```
qf(.05, df1 = 3, df2 = 16)
```

```
[1] 0.1150445
```

since 9.005 is much larger than 0.115, we reject H_0 and there appears to be a useful relationship between VO2 max and at least one of the predictors