chapter 13 the effects

Emilio Horner

2023-03-01

1.

a. 1
b. 9-8.1=.9

2.

$Y= \beta_0+\beta_1 X+\beta_2 A+ \varepsilon$

3.

a. a one unit change in x is associated with a change of 3 in y
b. 3/1.3= 2.31 because its above 1.96 it is significant
c. it means that its not likely the relationship between x and y is caused by random chance.

4.

a. 76.185
b. 19.693
c. 306.553
d. 3382
e. Yes, p is less than .05 in both models

5.

a. 110.230YearsEducation-1.581YearsEducation^2

Because the relationshop is not linear the association depends on how much education one starts with.

b. (110.230(17)-1.581(17^2)) -(110.230(16)-1.581(16^2))= 58.057

This is the relationship of a one year increase from 16 years of education to 17

c. more negative because the square term is subtracted
d. adding another term is unnecessary because the relationship isn't cubic. the theoretical model is that one would work more hours and then less as they increased in education. As education increases one works more up to a point.

6.

a. 50.174
b. -923.904 - -773.412

150.49 hours

c. it depends on the difference between people who don't have kids. this is a reference category for the variable.

7.

a. a one unit increase leads to an increase in 110.073 annual hours worked

b. They rise quicker for homeowning families. It is statistically significant.
c. its an interaction term. one would take the derivative. It would drop out the treatment variable (homeownership). that leaves the beta for education. so a one unit increase in the interaction term would lead to a 110.073 increase in hours worked. thats how much stronger the effect it.
d. A one unit change in ln(x) is associated with a 0.067 unit change in y.
e. a 10% increase in x is associated with a .1*832.347 unit change in y.
f. some of the models have an interactive term and others do not. that impacts the sample size.

8. b When error terms are correlated across time, such that knowing the error term in one period gives us some information about the error term in the next period

9. Checking if the R^2 value of the regression is particularly low

10. weighing, or weighted least squares.

11.

a. a. You're doing research on unusual sexual practices. You ask people whether they've ever engaged in these weird practices, which many people might prefer to keep secret, even if they've actually done them.

12.
13.

```
dengue <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/DAAG/dengue.csv")
```

```
library(tidyverse); library(modelsummary)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.4.0      ✓ purrr   0.3.5
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.5.0
## ✓ readr   2.1.3      ✓ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

2.

```
m2 <- lm(NoYes ~ humid, data = dengue)
modelsummary(m2)
```

	(1)
(Intercept)	−0.416
	(0.018)
humid	0.050
	(0.001)

	(1)
Num.Obs.	1998
R2	0.548
R2 Adj.	0.548
AIC	1258.9
BIC	1275.7
Log.Lik.	−626.466
RMSE	0.33

3. The intercept on the regression line is saying that at zero humidity the line would hit the y axis at 0.416. The slope is .050 meaning a one unit increase in humidity leads to a .050 increase in the dengue being present.

4.

```
summary(m2)
```

```
##
## Call:
## lm(formula = NoYes ~ humid, data = dengue)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08989 -0.14813  0.00041  0.18129  1.37152
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.416113   0.018433  -22.57   <2e-16 ***
## humid        0.049758   0.001011   49.24   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3312 on 1996 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.5485, Adjusted R-squared:  0.5482
## F-statistic:  2424 on 1 and 1996 DF,  p-value: < 2.2e-16
```

The summary statistics show standard error and the median. This can reveal how accurate the model is to the dataset.

5.

```
m2 <- lm(NoYes ~ humid + temp, data = dengue)
modelsummary(m2)
```

	(1)
(Intercept)	−0.407
	(0.019)
humid	0.053
	(0.002)
temp	−0.003
	(0.002)
Num.Obs.	1998
R2	0.549
R2 Adj.	0.549
AIC	1257.5
BIC	1279.9
Log.Lik.	−624.758
RMSE	0.33

This model controls for temperature

6.

```
m6 <- glm(NoYes ~ humid + temp, data = dengue,
       family = binomial(link = 'logit'))
modelsummary(m6)
```

	(1)
(Intercept)	−6.592
	(0.303)
humid	0.305
	(0.020)
temp	0.040
	(0.019)

	(1)
Num.Obs.	1998
AIC	1363.0
BIC	1379.8
Log.Lik.	−678.519
RMSE	0.32

7.

```
m7 <- lm(humid ~ temp, data = dengue)
modelsummary(m7)
```

	(1)
(Intercept)	2.362
	(0.211)
temp	0.779
	(0.011)
Num.Obs.	1998
R2	0.733
R2 Adj.	0.733
AIC	10998.5
BIC	11015.3
Log.Lik.	−5496.246
RMSE	3.79

I used resid(m7) to see the residuals, but left it out of the code because the chart it produced was 100 pages of data.

8.

```
m8 <- glm(humid ~ temp, data = dengue)
modelsummary(m8)
```

	(1)
(Intercept)	2.362
	(0.211)
temp	0.779
	(0.011)
Num.Obs.	1998
R2	0.733
AIC	10998.5
BIC	11015.3
Log.Lik.	−5496.246
RMSE	3.79

This model uses the logit link.