

Day12

Olivia Wu

2024-04-09

Problem 4.13

If the mother is white, then the birth weight will be about $117.87 - 0 + 0 - 0 = 117.87$ oz.

If the mother is black, then the birth weight will be about $117.87 - 7.31 + 0 - 0 = 110.56$ oz.

If the mother is Hispanic, then the birth weight will be about $117.87 - 0 + 0.65 - 0 = 118.52$ oz.

If the mother is other, then the birth weight will be about $117.87 - 0 + 0 - 0.73 = 117.14$ oz.

Problem 4.14

a) Since the p -values of the *Hispanic* and *Other* predictors is greater than 0.05, we can not say that they are significant.

b) $R^2 = 1.9\%$, so this model explains 1.9% of the variability in the birth weight of a baby.

c) $p < 0.05$, so we can reject the null hypothesis and say that at least one of the predictors is important.

Problem 4.16

a) $R^2 = 0.756$, so about 76% of the variability in log nitrogen assimilation is explained by the model.

```
##
## Call:
## lm(formula = LogNassim ~ LogMass, data = Cater)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26089 -0.11558  0.02162  0.16725  0.41862
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.88738    0.01841  -102.53  <2e-16 ***
## LogMass      0.37096    0.01332   27.85  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2501 on 251 degrees of freedom
## (14 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.7555, Adjusted R-squared:  0.7545
## F-statistic: 775.6 on 1 and 251 DF,  p-value: < 2.2e-16
```

b) $R^2 = 0.8348$, so about 83% of the variability is explained in this model. This is more than the model from before.

```
##
## Call:
## lm(formula = LogNassim ~ I(Instar == 2) + I(Instar == 3) + I(Instar ==
##      4) + I(Instar == 5), data = Cater)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08300 -0.06971  0.01542  0.10938  0.51487
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.72292    0.03270  -83.261   < 2e-16 ***
## I(Instar == 2)TRUE  0.15076    0.04625   3.260  0.00127 **
## I(Instar == 3)TRUE  0.30615    0.04495   6.811 7.27e-11 ***
## I(Instar == 4)TRUE  0.69102    0.04169  16.576   < 2e-16 ***
## I(Instar == 5)TRUE  1.24924    0.04169  29.966   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2068 on 248 degrees of freedom
## (14 observations deleted due to missingness)
## Multiple R-squared:  0.8348, Adjusted R-squared:  0.8322
## F-statistic: 313.4 on 4 and 248 DF,  p-value: < 2.2e-16
```

c) When the caterpillar is in stage 2, the log nitrogen assimilation is about -2.572. When it is in stage 3, then the log nitrogen assimilation is about -2.417.

d) $R^2 = 0.8361$, so about 84% of the variability is explained in this model. This is more than the model from before.

```
##
## Call:
## lm(formula = LogNassim ~ LogMass + I(Instar == 2) + I(Instar ==
##      3) + I(Instar == 4) + I(Instar == 5), data = Cater)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09910 -0.06945  0.01555  0.10943  0.49154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.56744    0.11702  -21.939   < 2e-16 ***
## LogMass         0.06148    0.04444   1.384  0.16776
## I(Instar == 2)TRUE  0.10206    0.05805   1.758  0.07998 .
## I(Instar == 3)TRUE  0.21480    0.07983   2.691  0.00761 **
## I(Instar == 4)TRUE  0.54611    0.11270   4.846 2.23e-06 ***
## I(Instar == 5)TRUE  1.04876    0.15077   6.956 3.12e-11 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2065 on 247 degrees of freedom
## (14 observations deleted due to missingness)
## Multiple R-squared:  0.8361, Adjusted R-squared:  0.8328
## F-statistic: 252 on 5 and 247 DF, p-value: < 2.2e-16
```

e) The R^2 is similar for both models, so I would check the p -value for the *logBodyMass* predictor. $p = 0.168 > 0.05$, so there is not enough evidence to show that *logBodyMass* is a significant predictor.

f) There are a lot of predictors to check, so we can use a nested F test using the models found in (a) and (d). The p value is less than 0.05, so the indicators of Instar are important.

```
## Analysis of Variance Table
##
## Model 1: LogNassim ~ LogMass
## Model 2: LogNassim ~ LogMass + I(Instar == 2) + I(Instar == 3) + I(Instar ==
## 4) + I(Instar == 5)
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      251 15.706
## 2      247 10.528  4     5.1776 30.368 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 4.19

a)

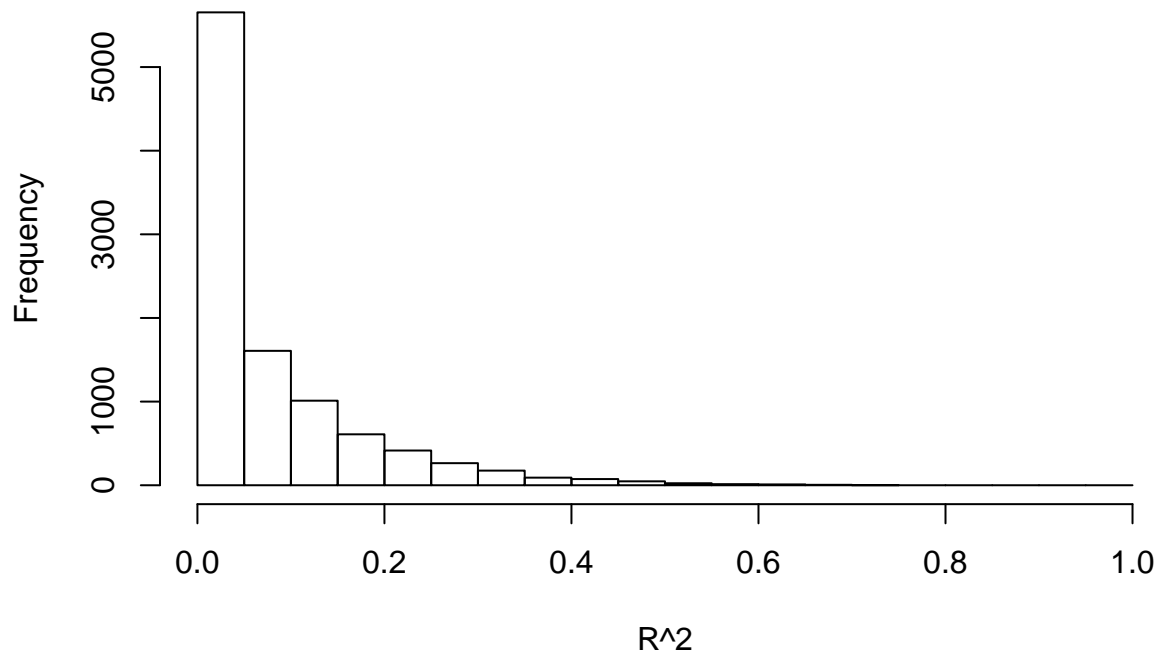
```
##
## Call:
## lm(formula = Time ~ Pitchers + Attendance, data = BB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.604  -8.248  -3.252   5.806  29.074
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.031e+02  2.556e+01   4.032  0.00198 **
## Pitchers      8.109e+00  2.557e+00   3.171  0.00890 **
## Attendance    6.312e-04  3.913e-04   1.613  0.13502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.45 on 11 degrees of freedom
## Multiple R-squared:  0.5307, Adjusted R-squared:  0.4454
## F-statistic: 6.219 on 2 and 11 DF, p-value: 0.0156
```

b) Let's use R^2 . $R^2 = 53.07$, so about 53.1% of the variability is explained by this model.

c)

```
## [1] -0.04047084
```

Histogram of Random Samples



d)

e) Find the number of observations that have an R^2 greater than 0.5307, and this divided by 10,000 is our p -value. This turns out to be $p = 0.0045 < 0.05$. There is enough evidence to suggest a relationship of predicting Time with Pitchers and Attendance.

```
## [1] 0.0034
```

f) The summary output shows a p -value of 0.0156 for the F statistic. This is greater than the p -value from our randomization test, but it is still less than 0.05.