Hw05ST430Yu

Haozhe (Jerry) Yu

2023-10-27

Question 1

```
software <- as_tibble(read_table("Datasets/software.txt",
            col_names = TRUE,
            ))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Rep = col_double(),
##   Software = col_double(),
##   SalesLastQuarter = col_double(),
##   SalesThisQuarter = col_double()
## )
```

1A. Fit a model in which sales last quarter is ignored. We want to know whether software package has any effect on sales.

```
softwareslr <- lm(SalesThisQuarter ~ as.factor(Software), software)
```

i. Write E(y|x).

E(y|x) = 81.5833333 + -2software2 + -7.6666667software3.

ii. What proportion of the variation in sales this quarter is explained by software package?

The proportion of the variation in sales this quarter that can be explained w the software package is 0.0817601.

iii. What is the null hypothesis for testing whether software package has any effect on sales?

H0: The type of software package has no effect on sales this quarter. That is $\beta_{software2} = \beta_{software3} = 0$.

iv. Give the test statistic.

The test statistic is 1.4691611.

v. Give the p-value.

The p value is 0.2447802.

vi. Do you reject H0 at = 0.05?

No, 0.2447802 > 0.05.

vii. Are the results statistically significant at the 0.05 level?

No, the results are not significant at the 0.05 level because out p value is higher than 0.05. We fail to reject the null hypothesis and conclude there is not enough evidence to support the claim that the type of software package has an effect on the sales for this quarter, if we ignore the sales form last quarter.

1B. Fit a model with software package and sales last quarter as the explanatory variables, and sales this quarter as the response variable.

```
softwaremlr <-
  lm(SalesThisQuarter ~ as.factor(Software) + SalesLastQuarter,
     software)
```

i. Write E(y|x).

E(y|x) = -36.442295 + 0.7535141software2 + -1.2835203software3 + 1.5019168Sales Last Quarter.

ii. What is the null hypothesis for testing whether software package has any effect on sales this quarter once you control for sales last quarter?

H0 = $\beta_{software2} = \beta_{software3} = 0$.

iii. Give the test statistic.

```
softf <-
  ftest(softwaremlr, matrix(c(0, 1, 0, 0, 0, 0, 1, 0), nrow = 2, byrow =
                              TRUE))
```

the F statistic for the general linear test is 0.2422432.

iv. Give the p-value.

The p value for the general linear test is 0.7862915.

v. Do you reject H0 at = 0.05?

No, 0.7862915 > 0.05.

vi. Are the results statistically significant at the 0.05 level?

No, 0.7862915 > 0.05, so we fail to reject the null hypothesis and conclude that there is not evidence for the claim that the software package has an effect on sales this quarter once you control for sales last quarter.

vii. What proportion of the remaining variation in sales this quarter is explained by software package once you allow for sales last quarter?

```
softwaremlrr <-
  lm(SalesThisQuarter ~ SalesLastQuarter,
     software)
```

The proportion of the remaining variation in sales this aquarter explained byt he software package controlling for last quarter's sales is 0.0149144. (Calculated using `rsp.partial`)

1C. Fit a full model (with interaction) in which the slopes and intercepts of the regression lines relating sales last quarter to sales this quarter might depend on the kind of software the sales representatives are using.

```
softwaremlri <- lm(SalesThisQuarter~as.factor(Software)*SalesLastQuarter,software)
```

i. Write E(y/x).

E(y|x) = -92.2593254 + 144.6515746software2 + 48.1096684software3 + 2.2122077Sales Last Quarter + -1.8579265software2*Sales Last Quarter + -0.6238716software3*Sales Last Quarter

ii. What is the null hypothesis for testing whether the three slopes are equal?

H0 = $\beta_{software2*SalesLastQuarter} = \beta_{software3*SalesLastQuarter} = 0$.

iii. What is the null hypothesis for testing whether the effect of software program on sales this quarter depends on sales last quarter?

H0 = $\beta_{software2*SalesLastQuarter} = \beta_{software3*SalesLastQuarter} = 0$.

iv. Carry out an F-test to determine whether the effect of software type on sales depends on the representative's performance last quarter.

```
tImatrix <- matrix(c(0,0,0,0,1,0,
                     0,0,0,0,0,1),
                   nrow = 2,
                   byrow = TRUE
                   )
```

A. Give the test statistic.

The test statistic (F) calculated by `ftest()` is 10.3049569.

B. Give the p-value.

The p value (p) calculated by `ftest()` is 3.919989×10^{-4}.

C. Do you reject H0 at = 0.05?

As $3.919989 \times 10^{-4} < 0.05$, we reject H0.

D. Are the results statistically significant at the 0.05 level?

As we reject the null hypothesis, we can conclude that our results are statistically significant at $\alpha = 0.05$ and that there is evidence to support the claim that the effect of software type on sales depends on the representative's performance last quarter.

v. Estimate the slopes and intercepts of the three regression lines.

- E(SalesThisQuarter|Software = 1) = -92.2593254 + 2.2122077*SalesLastQuarter
- E(SalesThisQuarter|Software = 2) = 52.3922493 + 0.3542812*SalesLastQuarter
- E(SalesThisQuarter|Software = 2) = -44.1496569 + 1.5883362*SalesLastQuarter

vi. Test whether the slope is different from zero for software package two.

A. State the null hypothesis.

H0: $\beta_{software2*SalesLastQuarter} + \beta_{SalesLastQuarter} = 0$

B. Give the test statistic.

```
softs42 <- matrix(c(0,0,0,1,1,0),
                  nrow=1,
                  byrow = TRUE)
```

The F value is 1.2518795.

C. Give the p-value.

The p value is 0.2720729.

D. Do you reject H0 at = 0.05?

As $0.2720729 > 0.05$, we fail to reject H0.

E. Are the results statistically significant at the 0.05 level?

As we fail to reject the null hypothesis, we conclude that the results are not statistically significant and there is not evidence to support the claim that the slope is different from zero for software package two.

1D. Test the hypothesis that you would test in order to answer this question: Controlling for sales last quarter, is average expected sales this quarter for software 1 and 3 different from expected sales this quarter for package 2?

Question 2

2.Download the "Explosives dataset" from Moodle. Fit a simple linear regression, relating the deflection of galvonometer (Y) to the area of the wires on the coupling (X). Complete the following parts.

```
pigs <- as_tibble(read_table("Datasets/pig.txt",
col_names = TRUE
))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Drug = col_double(),
##   Momweight = col_double(),
##   Dadweight = col_double(),
##   Pigweight = col_double()
## )
```