

hw_8

lisa liubovich

2024-03-28

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(broom)
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##   recode
##
## The following object is masked from 'package:purrr':
##
##   some
```

1.

```
brand <- read_csv("https://dcgerard.github.io/stat\_415\_615/data/brand.csv")
```

```
## Rows: 16 Columns: 3
## -- Column specification -----
## Delimiter: ","
## dbl (3): like, moisture, sweetness
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
brand
```

```
## # A tibble: 16 x 3
##   like moisture sweetness
##   <dbl>      <dbl>      <dbl>
## 1    64         4         2
## 2    73         4         4
## 3    61         4         2
## 4    76         4         4
## 5    72         6         2
## 6    80         6         4
## 7    71         6         2
## 8    83         6         4
## 9    83         8         2
## 10   89         8         4
## 11   86         8         2
## 12   93         8         4
## 13   88        10         2
## 14   95        10         4
## 15   94        10         2
## 16  100        10         4
```

```
lmbrand <- lm(like ~ moisture + sweetness, data = brand)
Anova(lmbrand)
```

```
## Anova Table (Type II tests)
##
## Response: like
##           Sum Sq Df F value    Pr(>F)
## moisture 1566.45  1 215.947 1.778e-09 ***
## sweetness  306.25  1  42.219 2.011e-05 ***
## Residuals   94.30 13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.

For moisture:

Reduced model: $Y_i = \beta_0 + \beta_2 X_{i2} + \varepsilon_i$ (including just sweetness as a predictor)

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$ (including both moisture and sweetness as predictors)

H0: $\beta_1 = 0$ (there is no relationship between like and moisture given sweetness)

HA: $\beta_1 \neq 0$ (there is a relationship between like and moisture given sweetness)

For sweetness:

Reduced model: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$ (including just moisture as a predictor)

Full model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$ (including both moisture and sweetness as predictors)

H0: $\beta_2 = 0$ (there is no relationship between like and sweetness given moisture)

HA: $\beta_2 \neq 0$ (there is a relationship between like and sweetness given moisture)

3.

$$\text{SSR}(X_1|X_2) = \text{SSE}(X_2) - \text{SSE}(X_1, X_2) = (1566.45 + 94.30) - (94.30) = 1566.45$$

$$\text{SSR}(X_2|X_1) = \text{SSE}(X_1) - \text{SSE}(X_1, X_2) = (306.25 + 94.30) - (94.30) = 306.25$$

$$\text{SSE}(X_1, X_2) = 94.30$$

4.

$$F^* = ((\text{SSE}(R) - \text{SSE}(F)) / (\text{dfr} - \text{dff})) / (\text{SSE}(F)/\text{dff})$$

Where:

$$\text{SSE}(R) = \text{SSE}(X_2) = 1566.45 + 94.30 = 1660.75$$

$$\text{SSE}(F) = \text{SSE}(X_1, X_2) = 94.30$$

$$df_r = n - q = 16 - 2 = 14$$

$$df_f = n - p = 16 - 3 = 13$$

$$F^* = ((1660.75 - 94.30) / (14 - 13)) / (94.30 / 13)$$

$$= (1566.45 / 1) / (94.30 / 13)$$

$$= 1566.45 / 7.353846$$

$$= 215.9475$$

```
94.30/13
```

```
## [1] 7.253846
```

```
1566.45 / 7.253846
```

```
## [1] 215.9475
```

5.

In order to see if we can run it a lack-of-fit test for the data, we need to check if there are repeat observations at the same level of a predictor.

```
brand %>%  
  group_by(moisture, sweetness) %>%  
  count()
```

```
## # A tibble: 8 x 3  
## # Groups:   moisture, sweetness [8]  
##   moisture sweetness     n  
##   <dbl>      <dbl> <int>  
## 1      4          2     2  
## 2      4          4     2  
## 3      6          2     2
```

```
## 4      6      4      2
## 5      8      2      2
## 6      8      4      2
## 7     10      2      2
## 8     10      4      2
```

There are repeat observations, so we can run an F-test LOF.

```
lm_reduced <- lm(like ~ moisture + sweetness, data = brand)
brand1 <- mutate(brand,
                 moist_fac = factor(moisture),
                 sweet_fac = factor(sweetness))
lm_full <- lm(like ~ moist_fac * sweet_fac, data = brand1)
anova(lm_reduced, lm_full)
```

```
## Analysis of Variance Table
##
## Model 1: like ~ moisture + sweetness
## Model 2: like ~ moist_fac * sweet_fac
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      13 94.3
## 2       8 57.0  5      37.3 1.047  0.453
```

p-value is large \rightarrow we do not have evidence to suggest a lack of fit (p-value = 0.453)

6.

```
glance(lmbrand)
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic      p.value    df logLik   AIC   BIC
##   <dbl>      <dbl> <dbl>    <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    0.952      0.945  2.69    129. 0.00000000266     2  -36.9  81.8  84.9
## # i 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ (none of the predictors are associated with like)

H_A : at least one β is non-zero (at least one of the predictors is associated with like)

F-statistic: 129.0832

df extra: 2

p-value: 2.658261e-09

We have evidence to suggest that at least one of moisture and sweetness are associated with like.

7.

The overall F-test assesses the overall significance of the regression model by evaluating whether there is a significant linear relationship between the predictors and response variable as a whole. That is, the overall

F-test tests whether at least one of the predictors in the model is associated with the response variable. The lack-of-fit test assesses whether the regression model adequately fits the observed data by evaluating whether the model adequately captures the functional form of the relationship between the predictors and the response variable. That is, the lack-of-fit test focuses on assessing whether the model adequately captures the variability in the data that is not explained by the model.

Assistance from ChatGPT:

<https://chat.openai.com/c/d1a5222f-6403-4f12-b79f-3d0b9d83b927>