

# WU #12 - F test

Math 158 - Jo Hardin

Tuesday 3/8/2022

Name: \_\_\_\_\_

Names of people you worked with: \_\_\_\_\_

Consider the regression model handouts concerning the birth weight data. Carry out an (one!) F test to evaluate whether, when mother's age (`mage`) and weight `gained` are both in the model, the smoking `habit` main effect and `habit*gained` interaction are simultaneously not needed. Note that you need to write out your null and alternative hypotheses, p-value (make a sketch of the appropriate area), conclusion, and summary in the context of the problem.

You might need the following output:

```
anova(lm(weight ~ gained + mage, data = births14))
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## gained     1   33.86   33.860  21.4987 4.042e-06 ***
## mage       1    7.74    7.743   4.9159 0.02685 *
## Residuals 938 1477.35    1.575
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Solution:

The full model is:

$$E[Y] = \beta_0 + \beta_1 \text{gained} + \beta_2 \text{habit} + \beta_3 \text{mage} + \beta_4 \text{gained} \cdot \text{habit}$$

The reduced model is:

$$E[Y] = \beta_0 + \beta_1 \text{gained} + \beta_3 \text{mage}$$

Which is to say, the hypothesis test of interest is:

$$H_0 : \beta_2 = \beta_4 = 0$$

$$H_a : \text{not } H_0$$

The test statistic is calculated from the SSE values obtained from the full and reduced model. Recall, the full model SSE is given in the notes (and reproduced here).

```
anova(lm(weight ~ gained*habit + mage, data = births14))
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## gained      1  33.86  33.860 21.8110 3.449e-06 ***
## habit       1  25.30  25.299 16.2960 5.861e-05 ***
## mage        1   6.42   6.417  4.1333 0.04233  *
## gained:habit 1   0.29   0.286  0.1845 0.66765
## Residuals 936 1453.09   1.552
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\begin{aligned}
 F &= \frac{\frac{SSE(R) - SSE(F)}{(n-3) - (n-5)}}{\frac{SSE(F)}{n-5}} \\
 &= \frac{1477 - 1453}{\frac{2}{936}} \\
 &= 7.73 \\
 \text{p-value} &= P(F_{2,936} \geq 7.73) \\
 &= 1 - \text{pf}(7.73, 2, 936) \\
 &= 0.000468
 \end{aligned}$$

```
1 - pf(7.73, 2, 936)
```

```
## [1] 0.0004680873
```

There is strong evidence that  $\beta_2$  and  $\beta_4$  are not simultaneously zero. That is, we should not remove both habit and the gained \* habit interaction from the model that predicts baby's birth weight in ounces conditional on gained and mage being in the model.