Homework 2

Brian Cervantes Alvarez
April 17, 2024
ST553 Statistical Methods

1.

Derivation of the formula for the sum of squares for a contrast $C = \sum_{i=1}^{g} w_i \mu_i$ as described by Oehlert.

1.1

Oehlert asserts that either $t = \frac{\hat{C}}{SE(\hat{C})}$ or $F = \frac{SS_w}{MSE}$ may be used to test $H_0 : C = 0$. Write down the t-statistic in terms of the contrast coefficients, sample means, and sample sizes:

The estimated contrast $\hat{C}$ is given by:

$$\hat{C} = \sum_{i=1}^{g} w_i \bar{x}_i$$

where w_i are the contrast weights and $\bar{x}_i$ are the sample means of each group.

The variance of $\hat{C}$ is:

$$\text{Var}(\hat{C}) = \sum_{i=1}^{g} w_i^2 \frac{\sigma_i^2}{n_i}$$

where σ_i^2 are the variances of each group and n_i are the sample sizes.

The standard error of $\hat{C}$ is the square root of the variance:

$$SE(\hat{C}) = \sqrt{\text{Var}(\hat{C})} = \sqrt{\sum_{i=1}^{g} w_i^2 \frac{\sigma_i^2}{n_i}}$$

The t-statistic is then given by:

$$t = \frac{\hat{C}}{SE(\hat{C})} = \frac{\sum_{i=1}^{g} w_i \bar{x}_i}{\sqrt{\sum_{i=1}^{g} w_i^2 \frac{\sigma_i^2}{n_i}}}$$

1.2

Recall that if the numerator degrees of freedom is 1, as it is with $F = \frac{SS_w}{MSE}$, then the F-statistic is the square of the t-statistic. Use this fact to write an equation relating F and t:

$$F = t^2 = \left(\frac{\sum_{i=1}^{g} w_i \bar{x}_i}{\sqrt{\sum_{i=1}^{g} w_i^2 \frac{\sigma_i^2}{n_i}}} \right)^2$$

1.3

Solve your equation in part 2 for SS_w:

$$F = \frac{SS_w}{MSE} = t^2$$

$$SS_w = MSE \cdot t^2$$

$$SS_w = MSE \cdot \left(\frac{\sum_{i=1}^{g} w_i \bar{x}_i}{\sqrt{\sum_{i=1}^{g} w_i^2 \frac{\sigma_i^2}{n_i}}} \right)^2$$

2.

Analysis of the French fry experiment using different types of fats to understand how fat type affects the amount of fat absorbed by French fries.

Report

In exploring how different fats influence fat absorption in French fries, our analysis suggests that fries cooked in vegetable oils may absorb more fat compared to those in animal fats, with no apparent difference between canola and coconut oils, either alone or combined. Although these findings are suggestive, they are inconclusive ($p = 0.0586$), indicating a need for further research to clarify these initial observations.

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	764.666667	191.166667	3.27	0.0586
Error	10	584.666667	58.466667		
Corrected Total	14	1349.333333			

R-Square	Coeff Var	Root MSE	Y Mean
0.566700	18.49923	7.646350	41.33333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
fat	4	764.6666667	191.1666667	3.27	0.0586

Source	DF	Type III SS	Mean Square	F Value	Pr > F
fat	4	764.6666667	191.1666667	3.27	0.0586

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Lard vs Beef Tallow	1	24.0000000	24.0000000	0.41	0.5361
Vegetable Oils vs Animal Fats	1	616.3333333	616.3333333	10.54	0.0088
Canola vs Coconut	1	42.6666667	42.6666667	0.73	0.4130
Mixed Vegetable vs Pure	1	3.5555556	3.5555556	0.06	0.8102

3.

For the statistical analysis of the fabric tensile strength data as described, here is the full process to perform in SAS, including reading in the data, transforming it, and conducting the appropriate statistical analysis.

3.1

```
DATA WORK.FABRIC;
    INFILE '/home/u63846470/Datasets/fabric.csv' DLM=',' FIRSTOBS=2;
    INPUT cotton strength;
    cotton2 = cotton**2;
    cotton3 = cotton**3;
RUN;

PROC PRINT DATA=WORK.FABRIC;
RUN;
```

Obs	cotton	strength	cotton2	cotton3
1	1576.5	70	2485352.25	3918157822.13
2	1596.4	70	2548492.96	4068414161.34
3	1418.3	80	2011574.89	2853016666.49
4	1362.9	80	1857496.41	2531581857.19
5	1242.0	90	1542564.00	1915864488.00
6	1167.1	90	1362122.41	1589733064.71
7	819.8	100	672072.04	550964658.39
8	759.8	100	577296.04	438629531.19

3.2

The model parametrization can be expressed as:

$$Y_{ij} = \mu + \gamma_i + \epsilon_{ij}$$

Where:

- Y_{ij} is the tensile strength of the j^{th} sample for the i^{th} cotton percentage.
- μ is the overall mean tensile strength.
- γ_i is the effect of the i^{th} cotton percentage level.
- ϵ_{ij} is the random error, and assuming $\sim N(0, \sigma^2)$

3.3

Error sum of squares is approximately 51.4385.

```
PROC GLM DATA=WORK.FABRIC;
    MODEL strength = cotton;
RUN;
```

The GLM Procedure

Dependent Variable: strength

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	948.561501	948.561501	110.64	<.0001
Error	6	51.438499	8.573083		
Corrected Total	7	1000.000000			

3.4

The polynomial model is expressed as:

$$Y_{ij} = \beta_0 + \beta_1 \text{cotton} + \beta_2 \text{cotton}^2 + \beta_3 \text{cotton}^3 + \epsilon_{ij}$$

```
PROC GLM DATA=WORK.FABRIC;
    MODEL strength = cotton cotton2 cotton3;
RUN;
```

The GLM Procedure

Dependent Variable: strength

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	988.881848	329.627283	118.59	0.0002
Error	4	11.118152	2.779538		
Corrected Total	7	1000.000000			

3.5

Source	DF	Type I SS	Mean Square	F Value	Pr > F
cotton	1	948.5615014	948.5615014	341.27	<.0001
cotton2	1	39.2336700	39.2336700	14.12	0.0198
cotton3	1	1.0866766	1.0866766	0.39	0.5657

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cotton	1	1.67791947	1.67791947	0.60	0.4806
cotton2	1	1.68497878	1.68497878	0.61	0.4797
cotton3	1	1.08667657	1.08667657	0.39	0.5657

3.6

Type I

1. **cotton:**
   - Full model: $Y = \beta_0 + \beta_1 \cdot \text{cotton}$
   - Reduced model: $Y = \beta_0$
2. **cotton2:**
   - Full model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_2 \cdot \text{cotton}^2$
   - Reduced model: $Y = \beta_0 + \beta_1 \cdot \text{cotton}$
3. **cotton3:**
   - Full model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_2 \cdot \text{cotton}^2 + \beta_3 \cdot \text{cotton}^3$
   - Reduced model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_2 \cdot \text{cotton}^2$

Type III

1. **cotton:**
   - Full model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_2 \cdot \text{cotton}^2 + \beta_3 \cdot \text{cotton}^3$
   - Reduced model: $Y = \beta_0 + \beta_2 \cdot \text{cotton}^2 + \beta_3 \cdot \text{cotton}^3$
2. **cotton2:**
   - Full model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_2 \cdot \text{cotton}^2 + \beta_3 \cdot \text{cotton}^3$
   - Reduced model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_3 \cdot \text{cotton}^3$
3. **cotton3:**
   - Full model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_2 \cdot \text{cotton}^2 + \beta_3 \cdot \text{cotton}^3$
   - Reduced model: $Y = \beta_0 + \beta_1 \cdot \text{cotton} + \beta_2 \cdot \text{cotton}^2$

3.7

Our test reveals that the fabric's strength does not just simply increase or decrease steadily with more cotton. Instead, how much it changes depends on the amount of cotton already in the fabric.

- Null, H_0: $\beta_2 = 0$
- Alternative, H_1: $\beta_2 \neq 0$

```
PROC GLM DATA=WORK.FABRIC;
    MODEL strength = cotton cotton2;
    TEST cotton2;
RUN;
```