ST553 HW1

Nick Sun

April 6, 2019

Question 1

```
filename fats 'Fats.csv';
data dat;
    infile fats firstobs=2 dlm=',';
    input absorb fat $;
run;

proc glm data=dat;
    class fat;
    model absorb=fat;
run;
```

Given this small p-value, we can reject H_0 at the $\alpha = .05$ level that there is no difference in the amount of fat absorption between the different kinds of fats. There is evidence to say that the model is better at explaining the data than using just the overall mean and at least one group mean that is significantly different than the others. In other words, fat absorption differs between fats.

The ANOVA table and boxplot are provided in Figures 1 and 2 respectively.

Question 2

The model will look like the following:

$$y_{ij} = \mu_i + \epsilon_{ij}$$

Our necessary assumptions are that the errors are independently distributed and $N(0, \sigma^2)$. Our estimates $\hat{y_{ij}}$ will simply be the calculated sample means for each group.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1258.133333	629.066667	6.91	0.0101
Error	12	1093.200000	91.100000		
Corrected Total	14	2351.333333			

Figure 1: ANOVA table for Fat data

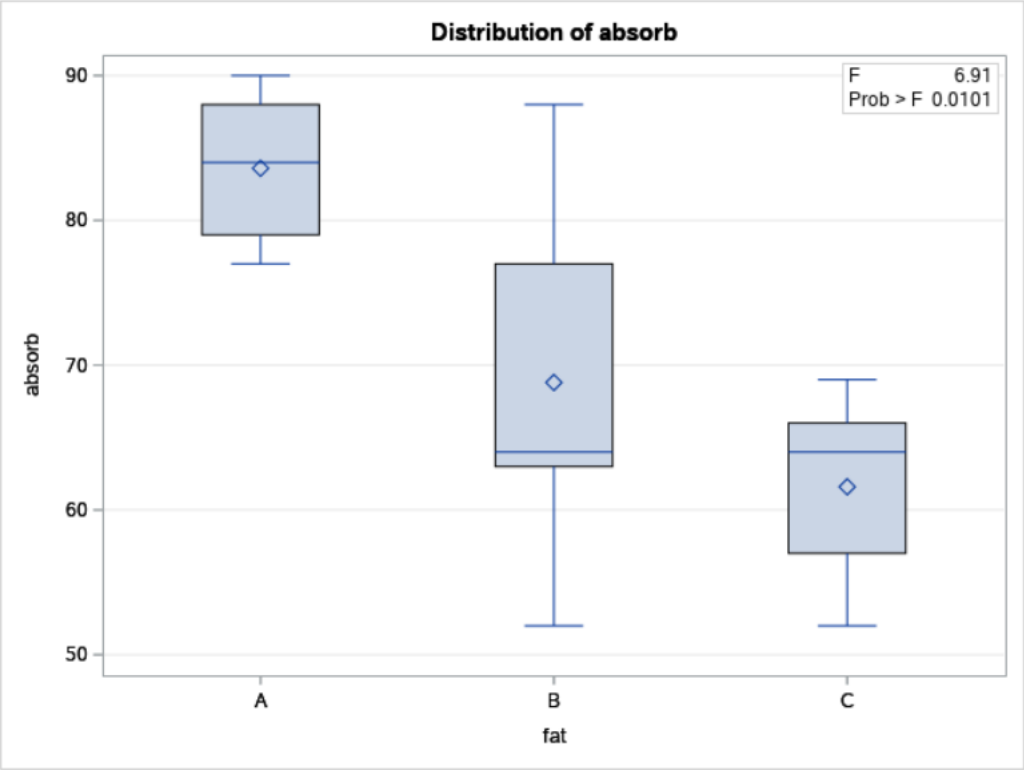

Figure 2: Fat absorption between different fats

We can write out and reparameterize the model using matrix algebra:

$$\begin{pmatrix} 77 \\ 90 \\ 88 \\ 84 \\ 79 \\ 88 \\ 64 \\ 52 \\ 77 \\ 63 \\ 66 \\ 64 \\ 69 \\ 52 \\ 57 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

where the matrices are our response vector Y, our design matrix X, and our parameter vector μ respectively.

```
(XTX <- t(X) %*% X)
```

```
##      col1 col2 col3
## col1    5    0    0
## col2    0    5    0
## col3    0    0    5
```

```r
(XTX_inv <- solve(XTX))
```

```
##      col1 col2 col3
## col1  0.2  0.0  0.0
## col2  0.0  0.2  0.0
## col3  0.0  0.0  0.2
```

```r
(XTY <- t(X) %*% Y)
```

```
##      [,1]
## col1  418
## col2  344
## col3  308
```

```r
(beta_hat <- XTX_inv %*% XTY)
```

```
##      [,1]
## col1 83.6
## col2 68.8
## col3 61.6
```

```
## [1] 83.6 68.8 61.6
```

These estimates agree with our matrix algebra calculations!

In the treatment effects parameterization, Y remains the same. The only thing that changes is the design matrix X which is now:

$$X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix}$$

and the parameter vector which is now:

$$\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

unit	Trt
17	1
13	2
1	3
19	4
4	5
15	1
16	2
5	3
6	4
3	5
14	1
20	2
12	3
18	4
9	5
11	1
7	2
10	3
2	4
8	5

Figure 3: Output of PROC PLAN

Question 3

We can do this in SAS using PROC PLAN

```
proc plan;
  factors unit=20 Trt=1 of 5 comb;
run;
```

We can arrange the experimental units by numbering the the paths from 1 to 20 and then we can assign the treatments to each one according to the above scheme. The 5 treatments in this experiment can be thought of as:

- **1** corresponds to 0 walking passes
- **2** corresponds to 25 walking passes
- **3** corresponds to 75 walking passes
- **4** corresponds to 200 walking passes
- **5** corresponds to 500 walking passes

Question 4

The probability that $P(|t_{stat}| < t_{\alpha/2,N-g})$ when the null hypothesis is correct is just $1 - 2(\alpha/2) = 1 - \alpha$

To get the $(1-\alpha)100\%$ confidence interval, we just have to do some algebra:

$$\begin{aligned}
\left|\frac{\bar{y_i}-\mu_0}{s/\sqrt{n_i}}\right| &< t_{\alpha/2,N-g} \\
-(s/\sqrt{n_i})(t_{\alpha/2,N-g}) < \bar{y_i}-\mu_0 &< (s/\sqrt{n_i})(t_{\alpha/2,N-g}) \\
\bar{y_{i\cdot}} - (s/\sqrt{n_i})(t_{\alpha/2,N-g}) < \mu_0 &< \bar{y_{i\cdot}} + (s/\sqrt{n_i})(t_{\alpha/2,N-g})
\end{aligned}$$