Alex Jones

Burr 4th

11/28/2021

Analysis of Iron Contents in Various Ethiopian Dishes

With iron-deficiency anemia being incredibly prevalent in developing countries, a study was done to examine various iron levels found in various dishes (Legumes, Meat, and Vegetables) cooked in different kinds of pots (Iron, Clay, and Aluminum). With the found data, we will conduct a two-way ANOVA test to examine if a cross-interaction effect exists between a dish/pot combination, as well as making comparisons between each factor-level combination to find how best to combat malnutrition.

1. Using the following code to input the Cook Data into R, making sure to convert the predictor variables into factor objects:

   ```
   > cook <- read.csv("cook.csv", header= TRUE)
   > cook$pot <- factor(cook$pot)
   > cook$dish <- factor(cook$dish)
   ```

   And then creating a table to check the first observations of each factor-level combination:

   ```
   > cook[c(1,5,9,13,17,21,25,29,33),]
   ```

   The following table is produced:

OBS	POT	DISH	IRON
1	Aluminum	Meat	1.77
5	Clay	Meat	2.27
9	Iron	Meat	5.27
13	Aluminum	Legumes	2.40
17	Clay	Legumes	2.41
21	Iron	Legumes	3.69
25	Aluminum	Vegetable	1.03
29	Clay	Vegetable	1.55
33	Iron	Vegetable	2.45

2. Using the “tapply()” function to calculate the means and standard deviations for each factor-level combination:

   ```
   > tapply(cook$iron, list(cook$pot,cook$dish), mean)
   > tapply(cook$iron, list(cook$pot,cook$dish), sd)
   ```

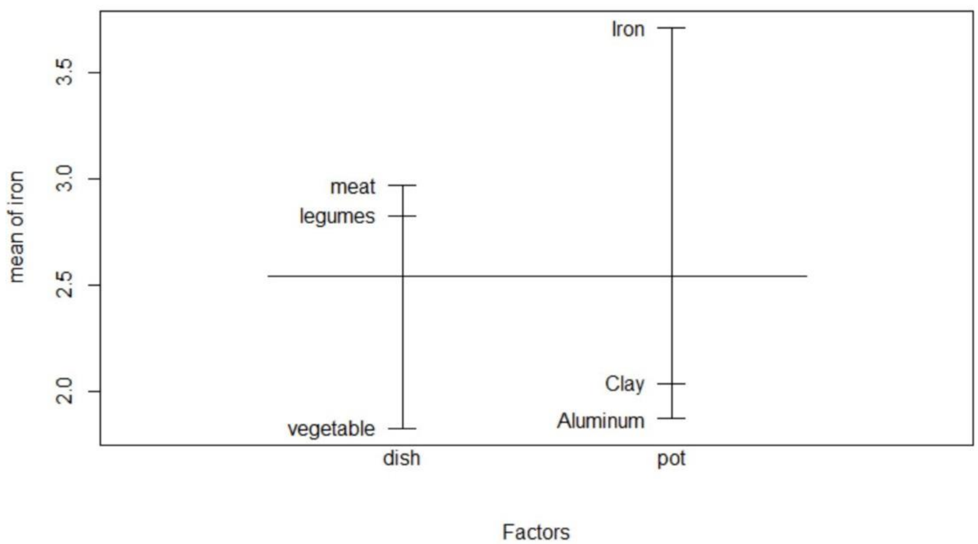
   The following values are produced, organized in the below table:

	Legumes	Meat	Vegetable	Row Average
Aluminum	Mean: 2.330 SD: 0.111	Mean: 2.058 SD: 0.252	Mean: 1.233 SD: 0.231	1.874
Clay	Mean: 2.473 SD: 0.0714	Mean: 2.178 SD: 0.621	Mean: 1.460 SD: 0.460	2.037
Iron	Mean: 3.670 SD: 0.173	Mean: 4.680 SD: 0.628	Mean: 2.790 SD: 0.240	3.713
Column Average	2.824	2.972	1.828	2.541

Then, making a plot to graph the factor level means:

```
> plot.design(iron ~ dish*pot, data=cook)
```

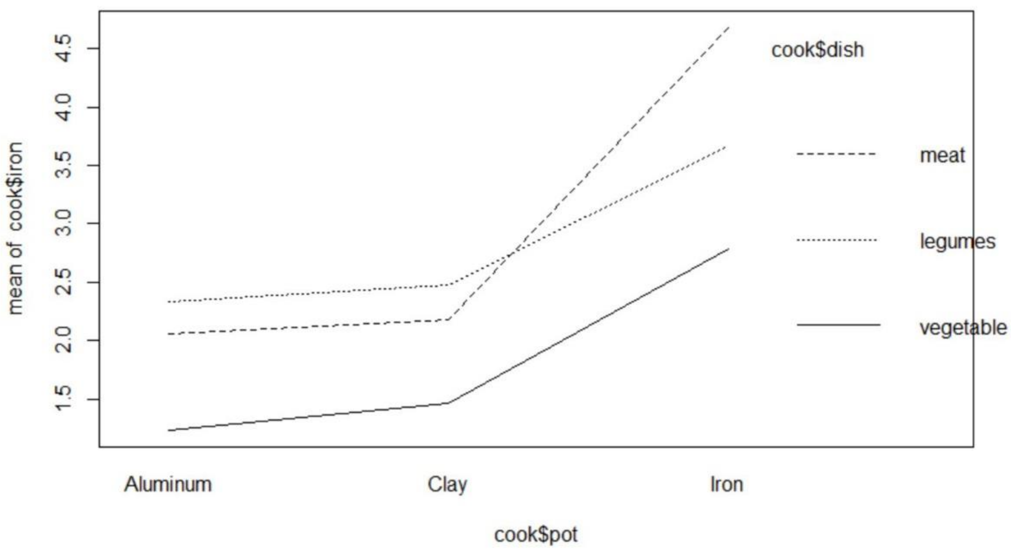
The following plot is created:



Inputting code to produce an interaction plot:

```
> interaction.plot(cook$pot,cook$dish,response=cook$iron)
```

The following plot is produced:



3. It seems very clear from the interaction plot that there exists some interaction between Pot Type and Dish Type. If there was no interaction, we would not observe the "Iron, Meat" combination as much higher than everything else; it would instead be between "Legumes, Iron" and "Vegetable, Iron" like it had in other factor levels. This can be verified by checking if μ_{32} (4.68) is equivalent to $\mu_{..} + \alpha_3 + \beta_2$, where α_2 and β_3 are the corresponding main row and column effects. Calculating these effects:
$$\alpha_3 = 2.972 - 2.541 = 0.431$$
$$\beta_2 = 3.713 - 2.541 = 1.172$$
And inputting them into the above equation:
$$2.541 + 0.431 + 1.172 = 4.144 \neq 4.68$$
It is apparent that there exists a cross-interaction value $(\alpha\beta)_{32}$ that shifts the mean.
4. Using the following code to fit the two-way ANOVA model with interaction, and then producing the ANOVA table:
```
> CoolBeans <- aov(iron ~ dish + pot + dish:pot, data=cook)
> anova(CoolBeans)
```
The following table is produced:

```
Analysis of Variance Table

Response: iron
          Df  Sum Sq Mean Sq F value    Pr(>F)    
dish       2  9.2969  4.6484  34.456 3.699e-08 ***
pot        2 24.8940 12.4470  92.263 8.531e-13 ***
dish:pot   4  2.6404  0.6601   4.893  0.004247 ** 
Residuals 27  3.6425  0.1349                      
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Testing to see if there exists an interaction effect between Dish and Pot Type, we will use the following hypotheses:

$$H_0\text{: all } (\alpha\beta)_{ij} = 0 \text{ vs. } H_A\text{: there exists some } (\alpha\beta)_{ij} \neq 0$$

Something to note is that the null distribution for the significance test is $F_{4,27}$.

Using the MSAB and MSE values found in the ANOVA table listed above, the F-statistic is found:

$$F_{obs} = \frac{MSAB}{MSE} = \frac{0.6601}{0.1349} = 4.893$$

Finding the corresponding p-value:

```
> pf(4.893, 4, 27, lower.tail = FALSE)
```

We find that the p-value is equal to 0.004, verified by the above ANOVA table. Given such a low p-value, we choose to reject H_0; there exists an interaction effect between Dish and Pot Types. With this conclusion, we will go ahead with an analysis of the cell means.

6. Performing the Scheffé method for the group of intervals, we will need to find three values: The critical value, and the standard error for two kinds of intervals we will be making ($\mu_{ij} - \mu_{kl}$ and $\mu_{3.} - \frac{\mu_{1.}+\mu_{2.}}{2}$)

Finding the critical constant will use the following formula, with $\alpha = 0.1$:

$$\sqrt{(ab-1)F(1-\alpha; ab-1, abn-ab)}$$

Which, after using R:

```
> S <- sqrt(8*qf(0.9,8,27))
```

Comes out to be 3.908

Calculating the standard error for the intervals, we will use the formula:

$$SE(\bar{Y}_i - \bar{Y}_j) = \sqrt{MSE \sum \frac{{c_i}^2}{n_i}}$$

Plugging the corresponding values for each standard error we will be calculating:

$$SE(\bar{Y}_{ij} - \bar{Y}_{kl}) = \sqrt{0.1349(2\left(\frac{1}{4}\right))}$$

and

$$SE\left(\bar{Y}_{3.} - \frac{\bar{Y}_{1.} + \bar{Y}_{2.}}{2}\right) = \sqrt{0.1349(\frac{1}{12} + 2\left(\frac{1}{2^2}\left(\frac{1}{12}\right)\right))}$$

We get 0.260 and 0.130, respectively.
Taking the collected values and plugging them into the equation:

$$\hat{L} \pm S * SE(\bar{Y}_i - \bar{Y}_j)$$

The following intervals are produced, organized by the below table:

PARAMETER	**CONTRAST**	**ESTIMATE**	**STANDARD ERROR**	**CONFIDENCE INTERVAL (90%)**
$\boldsymbol{D_1}$	$\mu_{31} - \mu_{32}$	-1.010	0.260	(-2.026, 0.006)
$\boldsymbol{D_2}$	$\mu_{31} - \mu_{33}$	0.880	0.260	(-0.136, 1.896)
$\boldsymbol{D_3}$	$\mu_{32} - \mu_{33}$	1.890	0.260	(0.874, 2.906)
$\boldsymbol{D_4}$	$\mu_{3.} - \frac{\mu_{1.} + \mu_{2.}}{2}$	1.758	0.130	(1.250, 2.266)

Looking at the first three intervals that are family-wise comparisons, we can say with 90% confidence that all 3 of the intervals contain the true parameter being estimated, that is, the true difference in iron levels between each dish type, given that an iron pot is used. We can see that there is generally an increase in iron levels when meat is cooked in an iron pot, as D_3 is above zero and D_1 narrowly contains it. Looking at D_4, there is a significant difference between cooking with iron and cooking with clay or aluminum, as the interval lies well above zero, meaning that there is an increase in iron levels when cooking with an iron pot.

Closing Remarks:

Throughout the study, it has seemed clear that to attain the highest iron levels, it is best to cook in an iron pot. Adding onto this, a meat dish cooked in an iron pot also exhibits significantly higher iron levels than otherwise. However, the main point of this study was to combat malnutrition in developing countries, and while we have figured out how to combat malnutrition, these results do not take into account other external factors such as cost effectiveness of each dish/pot combination. It seems that if we truly wanted to solve this issue, cost-effectiveness should also be taken into consideration.

Appendix: One-Way Analysis of the Cook Data

To use the Cook Data in a one-way setting, the table was changed, so each factor-level combination became its own factor. Doing this in excel, the changes in the table are illustrated below:

Obs.	Dish	Pot	Iron
1	Aluminum	Meat	1.77
5	Clay	Meat	2.27
9	Iron	Meat	5.27
13	Aluminum	Legumes	2.40
17	Clay	Legumes	2.41
21	Iron	Legumes	3.69
25	Aluminum	Vegetable	1.03
29	Clay	Vegetable	1.55
33	Iron	Vegetable	2.45

Obs.	Treatment	Iron
1	AM	1.77
5	CM	2.27
9	IM	5.27
13	AL	2.40
17	CL	2.41
21	IL	3.69
25	AV	1.03
29	CV	1.55
33	IV	2.45

After this, we go ahead and read in the data:

```
> newcook <- read.csv("newcook.csv", header = TRUE)
```

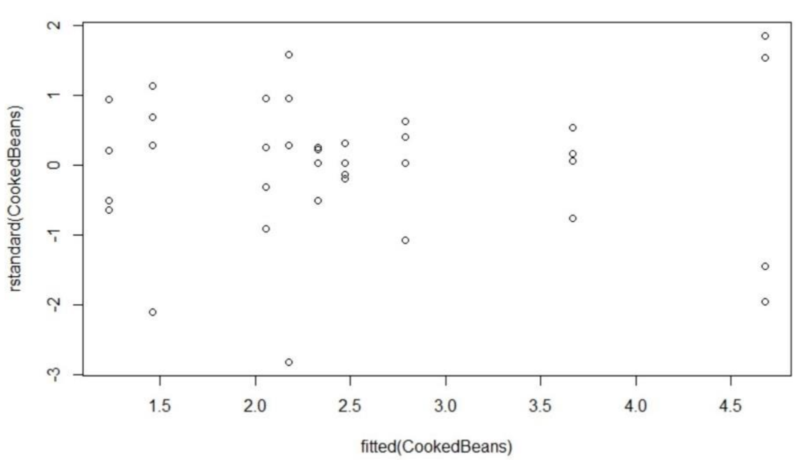
```
> newcook$Treatment <- factor(newcook$Treatment)
```

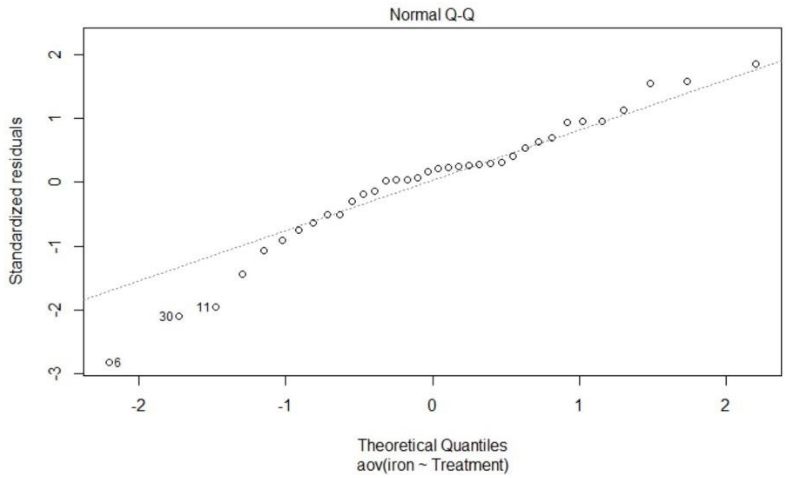
Doing a quick check on the assumptions needed for a one-way test (constant variance and normality):

```
> CookedBeans <- aov(iron ~ Treatment, data = newcook)
```

```
> plot(CookedBeans, which=1:2)
```

The following graphs are produced:





Looking at both graphs, there are questions about the validity of these assumptions. However, further studying shows that these assumptions can still be made, and the test can be carried out.

Carrying out the significance test, we will make the hypotheses H_0: all μ_i are equal vs. H_A: $\mu_i \neq \mu_j$ for some $i \neq j$ and set $\alpha = 0.05$. Printing out the ANOVA table:

```
> anova(CookedBeans)
```

```
Analysis of Variance Table

Response: iron
          Df Sum Sq Mean Sq F value    Pr(>F)
Treatment  8 36.831  4.6039  34.126 3.623e-12 ***
Residuals 27  3.643  0.1349
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

With such a low p-value, we can reject H_0: there appears to be at least one difference in iron levels between the nine treatments. This prompts further comparison, specifically pairwise comparisons.

Using code to generate Tukey’s Studentized Range Intervals at 95% confidence:

```
> TukeyHSD(CookedBeans)
```

We generate 36 intervals. Since that is a bit much to look at, I will only list the 21 that were found to be significant:

Contrast	Interval
AL-AV	(0.224, 1.971)
IL-AL	(0.466, 2.214)
IM-AL	(1.476, 3.224)
IL-AM	(0.739, 2.486)
IM-AM	(1.749, 3.496)
CL-AV	(0.366, 2.114)
CM-AV	(0.071, 1.819)
IL-AV	(1,564, 3.311)
IM-AV	(2.574, 4.321)
IV-AV	(0.684, 2.431)
CL-CV	(0.139, 1.886)
IL-CL	(0.324, 2.071)
IM-CL	(1.334, 3.081)
IL-CM	(0.619, 2.366)
IM-CM	(1.629, 3.376)
IL-CV	(1.336, 3.084)
IM-CV	(2.346, 4.094)
IV-CV	(0.456, 2.204)
IM-IL	(0.136, 1.884)
IV-IL	(0.006, 1.754)
IM-IV	(1.016, 2.764)

Looking at the table, there are some things to note. First, all the highlighted rows are comparisons between iron pots and non-iron pots. As you can see, the table is full of these significant differences, affirming a result that we found in the main body of this study: iron-potted dishes have significantly more iron levels in them than others. Another thing to note: every comparison between the Iron/Meat combination and another combination lead to a significant difference found in favor of the Iron/Meat combination.

As the final part of this analysis, we will be looking at the overall difference between one-way and two-way ANOVA tests. These can be neatly ordered into three major differences:

1. SSTR vs. SSA/SSB/SSAB: In a one-way ANOVA setting, the total sums of squares are split into sums square of treatment (SSTR) and sums square of error (SSE). However, in the two-way setting, the SSTR is split into three parts: sums square of A, B and the interaction effect AB.
2. Degrees of Freedom: In a one-way setting, the $dF(SSTR) = r - 1$, where $r =$ total number of treatments. In a two-way setting, we have three different degrees of freedom: $dF(SSA) = a - 1, dF(SSB) = b - 1,$ and $dF(SSAB) = (a-1)(b-1)$. Note that since those three sums are derived from SSTR, the sums of the degrees of freedoms is equal to $ab - 1$, or the total number of treatments minus one.
3. Null Distribution: In a one-way setting, there is only one significance test that can be done, with a null distribution of $F_{r-1,r(n-1)}$. However, in a two-way setting there are

three tests that you can do: Test for Factor A main effects, Factor B main effects, and cross-interaction effects, with degrees of freedom of $F_{a-1,ab(n-1)}, F_{b-1,ab(n-1)}, F_{(a-1)(b-1),ab(n-1)}$ respectively.