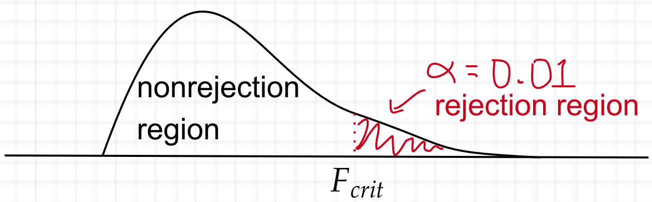
Homework 5 - Camila Alvarez (Z23400244)

Question 1

(a) Write the null and alternative hypotheses and show the rejection and nonrejction regions on the F-distribution curve for $\alpha = 0.01$.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$\mathrm{H_a}: \text{ at least two means are not equal}$$



(b) Calculate SSA, SSE, and SST. Determine the degrees of freedom for the numerator and denominator, and calculate the between-samples and within-samples variances.

$$\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 = 2,378 + 1,275 + 1,507 + 2,454 = 7,614$$

$$\frac{\left(\sum\sum y_{ij}\right)^2}{\sum n_i} = \frac{(398)^2}{22} = 7,200.18$$

$$\sum \frac{Y_i^2}{n_i} = \left(\frac{11,664}{5} + \frac{7,569}{6} + \frac{8,649}{6} + \frac{12,100}{5}\right) = 7,455.8$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{\left(\sum\sum y_{ij}\right)^2}{\sum n_i} = 7,614 - 7,200.18 = 413.82$$

$$SSA(\text{Between}) = \sum \frac{Y_i^2}{n_i} - \frac{\left(\sum\sum y_{ij}\right)^2}{\sum n_i} = 7,455.8 - 7,200.18 = 255.62$$

$$SSE(\text{Within}) = SST - SSE = 413.82 - 255.62 = 158.2$$

Degrees of Freedom:

$$v_1 = 4 - 1 = 3$$
$$v_2 = 22 - 4 = 18$$

Mean Squares :

$$s_1^2 = \frac{SSA}{k-1} = \frac{255.62}{3} = 85.21$$

$$s^2 = \frac{SSE}{k(n-1)} = \frac{158.2}{18} = 8.79$$

(c) Determine the critical value of F for $\alpha = 0.01$ and the F test statistic.

$$F_{Stat} = \left(\frac{SSA}{SSE}\right)\left(\frac{v_2}{v_1}\right) = \left(\frac{255.62}{158.2}\right)\left(\frac{18}{13}\right) = 9.69 \quad \Bigg| \quad F_{Crit} = F_{0.01}(3, 18) = 5.09$$

(d) Give the ANOVA table and determine whether the null should be rejected at a significance level of 1%.

Source	DF	Sum of Squares	Mean Square	F-Statistic	P-Value
Treatment	3	255.62	85.21	9.69	0.0005
Error	18	158.2	8.79		
Total	21	413.82			

The null hypothesis is rejected as the p-value of 0.0005 is less than the significance level of 0.001. As further support for rejecting the null hypothesis, the test statistic is greater than the critical value of 5.09, which puts it outside of the region of acceptance.

(e) Use Python to verify your results in your ANOVA table.

```python
import doex

exp = doex.OneWayANOVA(
    [19, 21, 26, 24, 18],
    [14, 16, 14, 13, 17, 13],
    [11, 14, 21, 13, 16, 18],
    [24, 19, 21, 26, 20],
)

print(exp)
```

```
+----------------------+-----+-----------------+---------------------+-------------+---------+
| Source of Variation  | DOF | Sum of Squares  | Mean Sum of Squares | F statistic | p value |
+----------------------+-----+-----------------+---------------------+-------------+---------+
|      Treatments      |  3  |    255.6182     |       85.2061       |    9.6947   |  0.0005 |
|        Error         |  18 |    158.2000     |        8.7889       |             |         |
|        Total         |  21 |    413.8182     |                     |             |         |
+----------------------+-----+-----------------+---------------------+-------------+---------+
```

Question 2

(a) Construct an ANOVA table.

$$\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 = 71.22$$

$$\frac{\left(\sum\sum y_{ij}\right)^2}{\sum n_i} = \frac{(37.2)^2}{20} = 69.192$$

$$\sum \frac{Y_i^2}{n_i} = \left(\frac{96.04}{5} + \frac{79.21}{5} + \frac{86.49}{5} + \frac{84.64}{5}\right) = 69.276$$

$$\sum \frac{Y_j^2}{n_i} = \left(\frac{34.81}{4} + \frac{50.41}{4} + \frac{59.29}{4} + \frac{54.76}{4} + \frac{82.81}{4}\right) = 70.52$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{\left(\sum\sum y_{ij}\right)^2}{\sum n_i} = 71.22 - 69.192 = 2.028$$

$$SSA = \sum \frac{Y_i^2}{n_i} - \frac{\left(\sum\sum y_{ij}\right)^2}{\sum n_i} = 69.276 - 69.192 = 0.084$$

$$SSB = \sum \frac{Y_j^2}{n_i} - \frac{\left(\sum\sum y_{ij}\right)^2}{\sum n_i} = 70.52 - 69.192 = 1.328$$

$$SSE = SST - SSA - SSB = 2.028 - 1.328 - 0.084 = 0.616$$

Degrees of Freedom :

$v_1 = 4 - 1 = 3$
$v_2 = 5 - 1 = 4$
$v_3 = (3)(4) = 12$

Mean Squares :

$$s_1^2 = \frac{SSA}{k-1} = \frac{0.084}{3} = 0.028$$

$$s_2^2 = \frac{SSB}{b-1} = \frac{1.328}{4} = 0.332$$

$$s^2 = \frac{SSE}{(b-1)(n-1)} = \frac{0.616}{12} = 0.513$$

$\textbf{F - Statistics:}$

$$F_{Stat(SSA)} = \left(\frac{SSA}{SSE}\right)\left(\frac{v_3}{v_1}\right) = \left(\frac{0.084}{0.616}\right)\left(\frac{12}{3}\right) = 0.5454$$

$$F_{Stat(SSB)} = \left(\frac{SSB}{SSE}\right)\left(\frac{v_3}{v_2}\right) = \left(\frac{1.328}{0.616}\right)\left(\frac{12}{4}\right) = 6.4675$$

Source	DF	Sum of Squares	Mean Square	F-Statistic	P-Value
Amounts Absorbed	3	0.084	0.028	0.5454	0.6605
Duration	4	1.328	0.332	6.4675	0.0052
Error	12	0.616	0.513		
Total	19	2.028			

(b) Can you conclude that the amount in skin varies with time?

$H_0' : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$
$H_a' :$ At least one of the α_i's are not equal to zero.

The P – value is approx 0.0052 which is less than the α level 0.05, therefore the null hypothesis is rejected. Therefore the differences between α_i 's is deemed signigicant at $\alpha = 0.05$. This points towards the conclusion that the average amount absorbed varies by duration instead of all of the averages being the same.

(c) Use Python to verify your results in your ANOVA table.

```
import doex

exp = doex.RandomizedCompleteBlockDesign(
    [
        [1.7, 1.8, 1.9, 2.3, 2.1],
        [1.5, 1.6, 1.7, 1.9, 2.2],
        [1.2, 1.8, 2.1, 1.7, 2.5],
        [1.5, 1.9, 2.0, 1.5, 2.3],
    ]
)

print(exp)
```

```
+---------------------+-----+----------------+----------------------+-------------+---------+
| Source of Variation | DOF | Sum of Squares | Mean Sum of Squares | F statistic | p value |
+---------------------+-----+----------------+----------------------+-------------+---------+
|      Treatments     |  3  |     0.0840     |        0.0280        |    0.5455   |  0.6605 |
|        Blocks       |  4  |     1.3280     |        0.3320        |    6.4675   |  0.0052 |
|        Error        |  12 |     0.6160     |        0.0513        |             |         |
|        Total        |  19 |     2.0280     |                      |             |         |
+---------------------+-----+----------------+----------------------+-------------+---------+
```

Question 3

(a) Determine the number of observations made at factor setting.

$SSA \text{ df } = a - 1 \Rightarrow 3 = a - 1$

$\boldsymbol{a = 4}$

$SSB \text{ df } = b - 1 \Rightarrow 1 = b - 1$

$\boldsymbol{b = 2}$

$SSC \text{ df } = c - 1 \Rightarrow 2 = c - 1$

$\boldsymbol{c = 3}$

$SSE \text{ df } = abc(n-1) \Rightarrow 24 = (4)(2)(3)(n-1)$

$\boldsymbol{n = 2}$

(b) Complete the table by computing the F-ratio and P-Value Columns

Source	Sum of Squares	DF	Mean Square	F-Statistic	P-Value
A	SSA	3	140	$F_A = \frac{s_A^2}{s^2} = \frac{140}{5} = 28$	≈ 0
B	SSB	1	480	$F_B = \frac{s_B^2}{s^2} = \frac{480}{5} = 96$	≈ 0
C	SSC	2	325	$F_C = \frac{s_C^2}{s^2} = \frac{325}{5} = 65$	≈ 0
AB	SS(AB)	3	15	$F_{AB} = \frac{s_{AB}^2}{s^2} = \frac{15}{5} = 3$	≈ 0.05
AC	SS(AC)	6	24	$F_{AC} = \frac{s_{AC}^2}{s^2} = \frac{24}{5} = 4.8$	≈ 0.0024
BC	SS(BC)	2	18	$F_{BC} = \frac{s_{BC}^2}{s^2} = \frac{18}{5} = 3.6$	≈ 0.043

ABC	SS(ABC)	6	2	$F_{ABC} = \frac{s^2_{ABC}}{s^2} = \frac{2}{5} = 0.4$	≈ 0.87
Error	SSE	24	5		
Total	SST	47			

Question 4

With hand calculation, verified by using Python, use a 0.05 significant level of significance to test the hypotheses:
1. Different temperatures have no effect on the cake texture.
2. Different temperatures have no effect on the cake texture.
3. The type of overn and temperature do not interact.

$$\sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 = 1,126,087$$

$$\frac{\left(\sum\sum\sum y_{ijk}\right)^2}{\sum n_i} = \frac{(5159)^2}{24} = 1,108,970.0417$$

$$\sum_{i=1}^{a}\frac{Y_i^2}{n_i} = \left(\frac{(1,871)^2}{8} + \frac{(1,704)^2}{8} + \frac{(1,584)^2}{8}\right) = 1,114,164.125$$

$$\sum_{j=1}^{b}\frac{Y_j^2}{n_i} = \left(\frac{(1,219)^2}{6} + \frac{(1,225)^2}{6} + \frac{(1,282)^2}{6} + \frac{(1,433)^2}{6}\right) = 1,113,933.17$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij}^2}{n_i} = \frac{(448)^2}{2} + \frac{(473)^2}{2} + \frac{(461)^2}{2} + \frac{(489)^2}{2} + \frac{(395)^2}{2} + \frac{(360)^2}{2} + \frac{(430)^2}{2} + \frac{(519)^2}{2} + \frac{(376)^2}{2} + \frac{(392)^2}{2} + \frac{(391)^2}{2} + \frac{(425)^2}{2} = 1,122,253.5$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{\left(\sum\sum\sum y_{ijk}\right)^2}{\sum n_i} = 1,126,087 - 1,108,970.0417 \approx 17,116.96$$

$$SSA = \sum_{i=1}^{a}\frac{Y_i^2}{n_i} - \frac{\left(\sum\sum\sum y_{ijk}\right)^2}{\sum n_i} = 1,114,164.125 - 1,108,970.0417 \approx 5,194.08$$

$$SSB = \sum_{j=1}^{b}\frac{Y_j^2}{n_i} - \frac{\left(\sum\sum\sum y_{ijk}\right)^2}{\sum n_i} = 1,113,933.17 - 1,108,970.0417 \approx 4,963.125$$

$$SS(AB) = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij}^2}{n_i} - \frac{\left(\sum\sum\sum y_{ijk}\right)^2}{\sum n_i} - SSA - SSB = 1,122,253.5 - 1,108,970.0417 - 5,194.08 - 4,963.125 = \approx 3,126.25$$

$$SSE = SST - SSA - SSB - SS(AB) = 17,116.96 - 5,194.08 - 4,963.125 - 3,126.25 \approx 3,833.5$$

Degrees of Freedom :

$v_1 = 3 - 1 = 2$
$v_2 = 4 - 1 = 3$
$v_3 = (3)(2) = 6$
$v_4 = (3)(4)(1) = 12$

Mean Squares :

$$s_1^2 = \frac{SSA}{a-1} = \frac{5,194.08}{2} \approx 2,597.04$$

$$s_2^2 = \frac{SSB}{b-1} = \frac{4,963.125}{3} \approx 1,654.375$$

$$s_3^2 = \frac{SS(AB)}{(a-1)(b-1)} = \frac{3,126.25}{6} \approx 521.0417$$

$$s^2 = \frac{SSE}{ab(n-1)} = \frac{3,833.5}{12} \approx 319.458$$

F – Statistics :

$$F_1 = \frac{s_1^2}{s^2} = \frac{2,597.04}{319.458} = 8.13$$

$$F_2 = \frac{s_2^2}{s^2} = \frac{1,654.375}{319.458} = 5.18$$

$$F_3 = \frac{s_3^2}{s^2} = \frac{521.0417}{319.458} = 1.63$$

Source	DF	Sum of Squares	Mean Square	F-Statistic	P-Value
Temperature	2	5, 194.08	2, 597.04	8.13	0.0059
Oven	3	4, 963.125	1, 654.375	5.18	0.0159
Interactions	6	3, 126.25	521.0417	1.63	0.2215
Error	12	3, 833.5	319.458		
Total	23	17, 116.96			

```
> A <- c("A-Category-1","A-Category-1","A-Category-2","A-Category-2","A-Category-3","A-Category-3","A-Category-1","A-Category-1","A-Category-2","A-Category-2","A-Category-3","A-Category-3","A-Category-1","A-Category-1","A-Category-2","A-Category-2","A-Category-3","A-Category-3","A-Category-1","A-Category-1","A-Category-2","A-Category-2","A-Category-3","A-Category-3")

> B <- c("B-Category-1","B-Category-1","B-Category-1","B-Category-1","B-Category-1","B-Category-1","B-Category-2","B-Category-2","B-Category-2","B-Category-2","B-Category-2","B-Category-2","B-Category-3","B-Category-3","B-Category-3","B-Category-3","B-Category-3","B-Category-3","B-Category-4","B-Category-4","B-Category-4","B-Category-4","B-Category-4","B-Category-4")
> DV <-
c(227,221,187,208,174,202,214,259,181,179,198,194,225,236,232,198,178,213,260,229,246,273,206,219)
> ID <- c(0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,4,5)
> df1 <- data.frame(A,B,DV,ID)
> Model1 <- aov(DV ~ A + B + A:B, data=df1)
> summary(Model1)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
A            2   5194  2597.0   8.130 0.00586 **
B            3   4963  1654.4   5.179 0.01589 *
A:B          6   3126   521.0   1.631 0.22152
Residuals   12   3834   319.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> res=residuals(object = Model1)
```

(a) Different temperatures have no effect on the cake texture.

$H_0' : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$
$H_a' :$ At least one of the α_i's are not equal to zero.

The P – value is approx 0.0059 which is less than the α level 0.05, therefore the null hypothesis is rejected. Therefore the differences between α_i 's is deemed signigicant at $\alpha = 0.05$.

(b) Different ovens have no effect on the cake texture.

$H_0'' : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
$H_a'' :$ At least one of the β_j's are not equal to zero.

The P – value is approx 0.0159 which is less than the α level 0.05, therefore the null hypothesis is rejected. Therefore the differences between β_j 's is deemed signigicant at $\alpha = 0.05$.

(c) The type of oven and temperature do not interact.

$H_0''' : \alpha\beta_{11} = \beta_{12} = ... = \alpha\beta_{34} = 0$
$H_a''' :$ At least one of the $\alpha\beta_{ij}$'s are not equal to zero.

The P – value is approx 0.2215 which is more than the α level 0.05, therefore the null hypothesis cannot be rejected. Therefore the differences between $\alpha\beta_{ij}$'s is deemed not signigicant at $\alpha = 0.05$.