Question 1

```
% Question 1
sse = 2000;
msg = 116.3;
df_g = 4; % df = (# of groups = 5) - 1
df_e = 45; % df = (n = 50) – (K = 5)
% msg = ssg/df_g
ssg = msg*df_g; % 465.2
% mse = sse/df_e
mse = sse / df_e; % 44.44
F = msg/mse; % 2.6168
```

Source	Sum of Squares	Df	Mean Squares	F
Group	465.2	4	116.3	2.6168
Error	2000	48	44.44	
Total	2465.2	52		

The critical value for F_(4,48) is 3.0662. Since the test statistic is less than the critical value, we fail to reject the null hypothesis. There is not enough significant evidence to disprove the null hypothesis and claim that at least on the means of the different categories are different.

Question 2

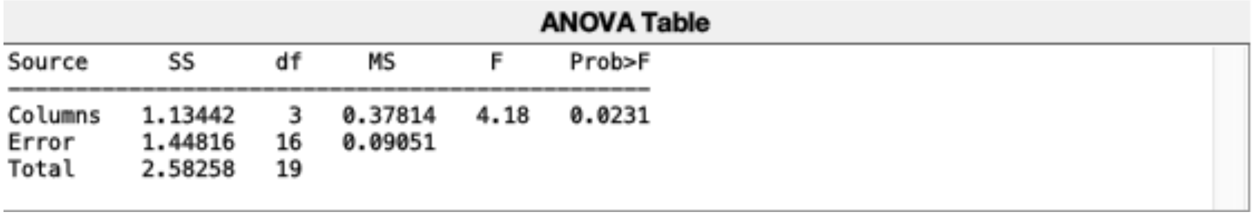
```
fifty = [-0.4, 0.34, 0.19, 0.05, -0.14];
hundred = [0.01, -0.39, -0.08, -0.09, -0.31];
one_fif = [0.65, 0.53, 0.39, -0.15, 0.46];
two_hun = [0.24, 0.44, 0.13, 1.03, 0.05];
groups = {'50 ppm','100 ppm','150 ppm','200 ppm'};
two_anova = anova1([fifty.' hundred.' one_fif.' two_hun.'], groups);
two_anova;
```

a. State the null and alternative hypothesis

```
Ho : u_50 = u_100 = u_150 = u_200
H1 : u_50 != u_100 != u_150 != u_200 (at least one mean is different)
```

b. Calculated the following summary statistics for each group: sample size, sample mean, sample standard deviation.
   - Sample size for every group is 5
   - Sample mean
     - mean(fifty) = 0.0080
     - mean(hundred) = -0.1720
     - mean(one_fif) = 0.3760
     - mean(two_hun) = 0.3780
   - sample standard deviation
     - std(fifty) = 0.2887
     - std(hundred) = 0.1695
     - std(one_fif) = 0.3093
     - std(two_hun) = 0.3928

The following parts, I answered using the calculated ANOVA table below



ANOVA Table

```
Source      SS      df     MS       F      Prob>F
-------------------------------------------------
Columns   1.13442    3   0.37814   4.18   0.0231
Error     1.44816   16   0.09051
Total     2.58258   19
```

c. What is the error mean square?

a. sserror = (std(fifty)^2*4)+(std(hundred)^2*4)+(std(one_fif)^2*4)+(std(two_hun)^2*4);

b. mserror = sserror/16; % 0.09051

c. The error mean square is in the 3rd column 2nd row. We get an MSE of 0.09051. This is calculated by dividing the sum of error squares by the df_error.

d. How many degrees of freedom are associated with error mean square?

a. The df for error MS is found in the 2nd column, 2nd row with a value of 16. It its calculated by taking the total number of observations and subtracting by the number of groups. N = 20 and k = 4 so df_group = 16.

e. Calculate the estimate of the grand mean

a. I used the code below to calculate the grand mean

i. grandmean = (mean(fifty)*5 + mean(hundred)*5 + mean(one_fif)*5 + mean(two_hun)*5)/20; % 0.1475

b. I get a value of 0.1475.

f. Calculate the Group Sum of Squares

a. ssgroups = ((mean(fifty) - grandmean)^2)*5 + ((mean(hundred) - grandmean)^2)*5 + ((mean(one_fif) - grandmean)^2)*5 + ((mean(two_hun) - grandmean)^2)*5; % 1.13442

b. The SS_group is shown in the 1st column, 1st row with a value of 1.13442.

g. Calculate the group degrees of freedom and the Group Mean Squares

a. The group df is shown in the table in the 2nd column, first row with a value of 3. This is calculated by subtracting the number of groups (4) – 1 = 3.

b. msegroups = ssgroups/3; % 0.37814

i. The Group Mean Squares is shown in the 3rd column and first row with a value of 0.37814. The Group Mean Square is calculated by dividing the Group Sum of Squares by the df_groups.

h. What is F for this example?

a. F = msegroups/mserror; % 4.18

b. The F statistic for this example is found in the 4th column and is 4.18 which could be calculated by dividing the MS_group / MS_error.

i. What is the p-value for this test?

a. The calculated p-value for this test is in the 5th column of the ANOVA table and is 0.0231.

j. Report your results

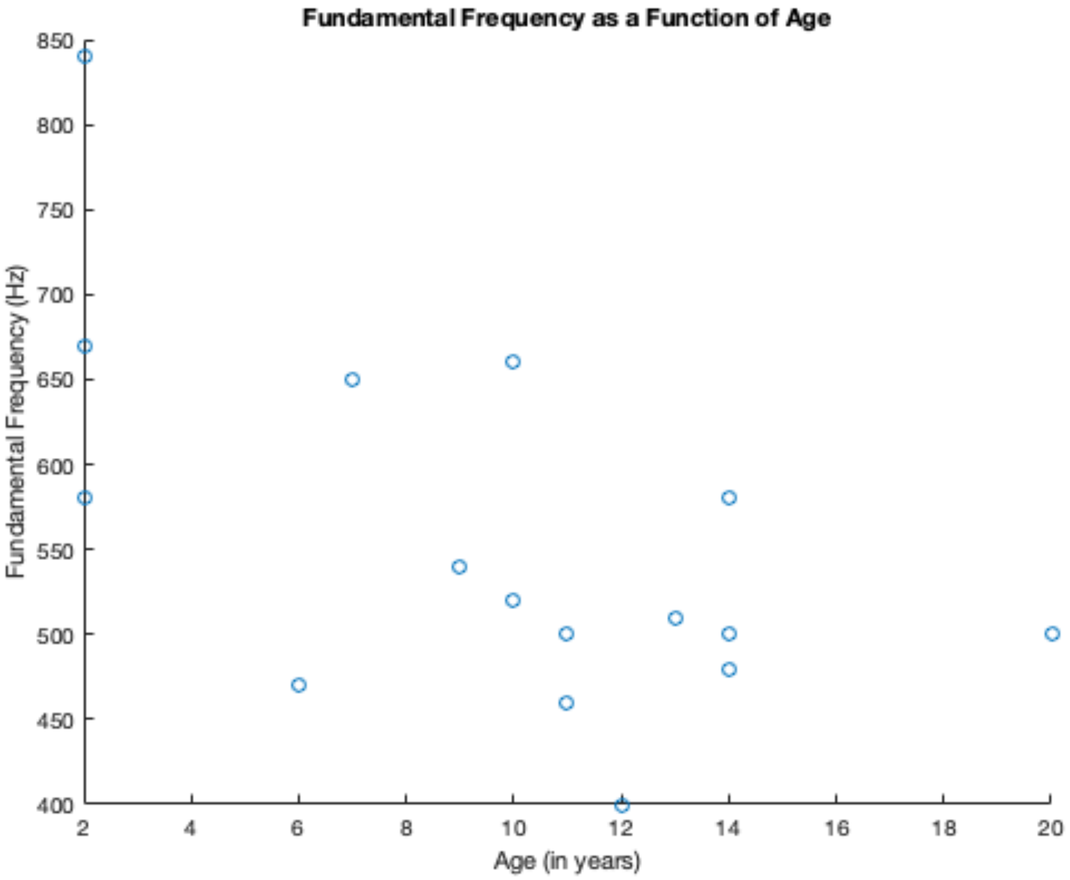
With a p-value of 0.0231 < alpha = 0.05, we reject the null hypothesis. There is statistically significant evidence that at least one of the groups of caffeine concentration has a different difference between the amount of nectar consumed from the caffeine feeders and that removed from the control feeders at the same station.

Question 3

```
% Question 3
age = [2,2,2,6,9,10,13,10,14,14,12,7,11,11,14,20];
hz = [840,670,580,470,540,660,510,520,500,480,400,650,460,500,580,500];

% a
hold on;
scatter(age, hz); % negative
title('Fundamental Frequency as a Function of Age');
```

```
xlabel('Age (in years)');
ylabel('Fundamental Frequency (Hz)');
hold off;
```



Based on the plot above, the relationship suggests a negative association.

```
% b
ssa = sum((age - mean(age)).^2); % 380.4375
```

Using the Equation for the Sum of Squares, I got a Sum of Squares for Age of 380.4375.

```
% c
ssf = sum((hz - mean(hz)).^2); % 174175
```

Using the Equation for the Sum of Squares, I got a Sum of Squares for fundamental frequency of 174175.

```
% d
ssp = sum((hz - mean(hz)).*(age - mean(age))); % -4.89875e+03
```

Using the Equation for the Sum of Products, I got a Sum of products between age and frequency of -4.89875e+03.

```
% e
r = ssp/(sqrt(ssf)*sqrt(ssa)); % -0.601797954619960
```

Using the equartion for the correlation coefficient, we got an r value of -0.601798.

```
% f
z = 0.5*log((1+r)/(1-r)); % -0.695961235390619
```

Using the Equation for Fisher's z-transformation, we got a transformed correlation coefficient of -0.69596.

```
% g
sdz = sqrt(1/(16-3)); % 0.277350098112615
```

Using the equation for an approximate standard error of the z-transformed correlation, we got an error of about 0.27735.

```
% h
z_crit = 1.96;
```

Based off what I already know from previous experience, the two tailed critical value of the standard normal distribution corresponding to alpha = 0.05 is 1.96.

```
% i
lower_untrans = z - 1.96*sdz; % -1.239567427691344
upper_untrans = z + 1.96*sdz; % -0.152355043089895
```

Using the formula for the confidence interval for the z-transformed population correlation, we are 95% confident that the true epsilon for the z-transformed population correlation is between the values -1.24 and -0.15.

```
% j
lower_r = (exp(2*lower_untrans) - 1)/(exp(2*lower_untrans) + 1); % -0.845332178884088
upper_r = (exp(2*upper_untrans) - 1)/(exp(2*upper_untrans) + 1); % -0.151187061636697
```

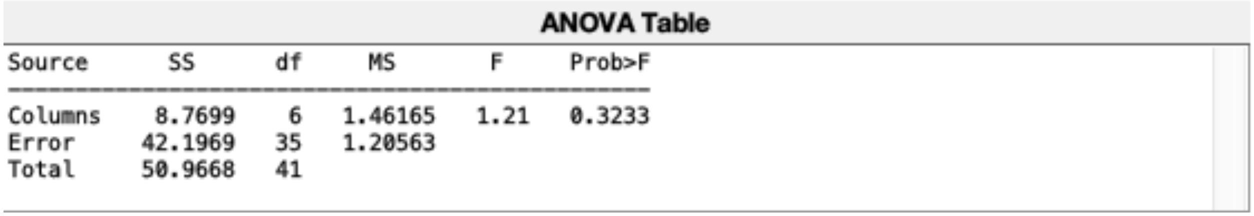
After using the formula for untransforming the z transform to get the actual population correlation coefficient confidence interval, we are 95% confident that the true population correlation coefficient is between the values -0.845 and -0.151.

Question 4

```
% Question 4

% a
sat_meanb = mean(Q4Boston(:,2)); % 2.9567
sun_meanb = mean(Q4Boston(:,3)); % 3.9517
mon_meanb = mean(Q4Boston(:,4)); % 3.26
tue_meanb = mean(Q4Boston(:,5)); % 2.3783
wed_meanb = mean(Q4Boston(:,6)); % 3.0233
thu_meanb = mean(Q4Boston(:,7)); % 3.5627
fri_meanb = mean(Q4Boston(:,8)); % 3.345
sat_stdb = std(Q4Boston(:,2)); % 1.295
sun_stdb = std(Q4Boston(:,3)); % 0.9726
mon_stdb = std(Q4Boston(:,4)); % 1.0068
tue_stdb = std(Q4Boston(:,5)); % 0.9203
wed_stdb = std(Q4Boston(:,6)); % 1.2214
thu_stdb = std(Q4Boston(:,7)); % 1.067
fri_stdb = std(Q4Boston(:,8)); % 1.1513

% b
groups = {'Sat','Sun','Mon','Tues','Wed','Thurs','Fri'};
four_anova1 = anova1(Q4Boston(:,2:8), groups);
four_anova1;
```



```
                    ANOVA Table
Source      SS      df     MS       F     Prob>F
-------------------------------------------------
Columns    8.7699    6   1.46165   1.21   0.3233
Error     42.1969   35   1.20563
Total     50.9668   41
```

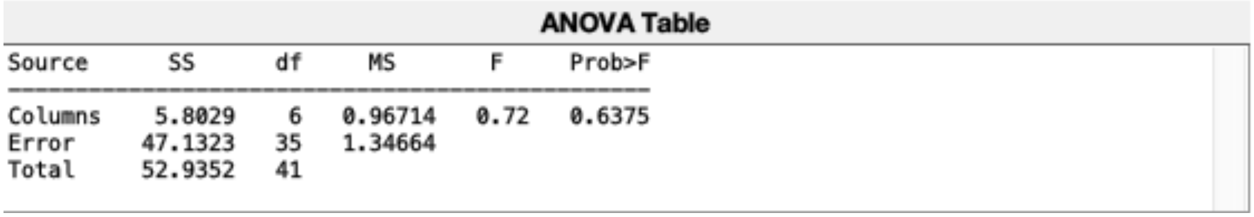
```
Ho: the mean precipitation in Boston for each day are equal
H1: at least one mean for any day is different
```

Group Sum of Squares = 8.7699
Error Sum of Squares = 42.1969
Observed F Statistic = 1.21
Critical F value (F_6,35) = 2.7961

With a p-value > alpha = 0.05 and a F statistic of 1.21 less than the critical value 2.7961, we fail to reject the null hypothesis. There is not enough statistically significant evidence to determine any difference in mean precipitation in Boston for any of the days of the week.

```
% c
sat_meanp = mean(Q4Pittsburg(:,2)); % 2.905
sun_meanp = mean(Q4Pittsburg(:,3)); % 3.435
mon_meanp = mean(Q4Pittsburg(:,4)); % 2.4817
tue_meanp = mean(Q4Pittsburg(:,5)); % 2.3783
wed_meanp = mean(Q4Pittsburg(:,6)); % 2.7233
thu_meanp = mean(Q4Pittsburg(:,7)); % 3.2117
fri_meanp = mean(Q4Pittsburg(:,8)); % 2.465
sat_stdp = std(Q4Pittsburg(:,2)); % 1.2128
sun_stdp = std(Q4Pittsburg(:,3)); % 1.3397
mon_stdp = std(Q4Pittsburg(:,4)); % 1.0773
tue_stdp = std(Q4Pittsburg(:,5)); % 0.6018
wed_stdp = std(Q4Pittsburg(:,6)); % 1.1076
thu_stdp = std(Q4Pittsburg(:,7)); % 1.5411
fri_stdp = std(Q4Pittsburg(:,8)); % 1.0178

four_anova2 = anova1(Q4Pittsburg(:,2:8), groups);
four_anova2;
```



```
                    ANOVA Table
Source      SS      df     MS       F     Prob>F
-------------------------------------------------
Columns    5.8029    6   0.96714   0.72   0.6375
Error     47.1323   35   1.34664
Total     52.9352   41
```

```
Ho: the mean precipitation in Pittsburg for each day are equal
H1: at least one mean for any day is different
```

Group Sum of Squares = 5.8029
Error Sum of Squares = 47.1323
Observed F Statistic = 0.72
Critical F value (F_6,35) = 2.7961

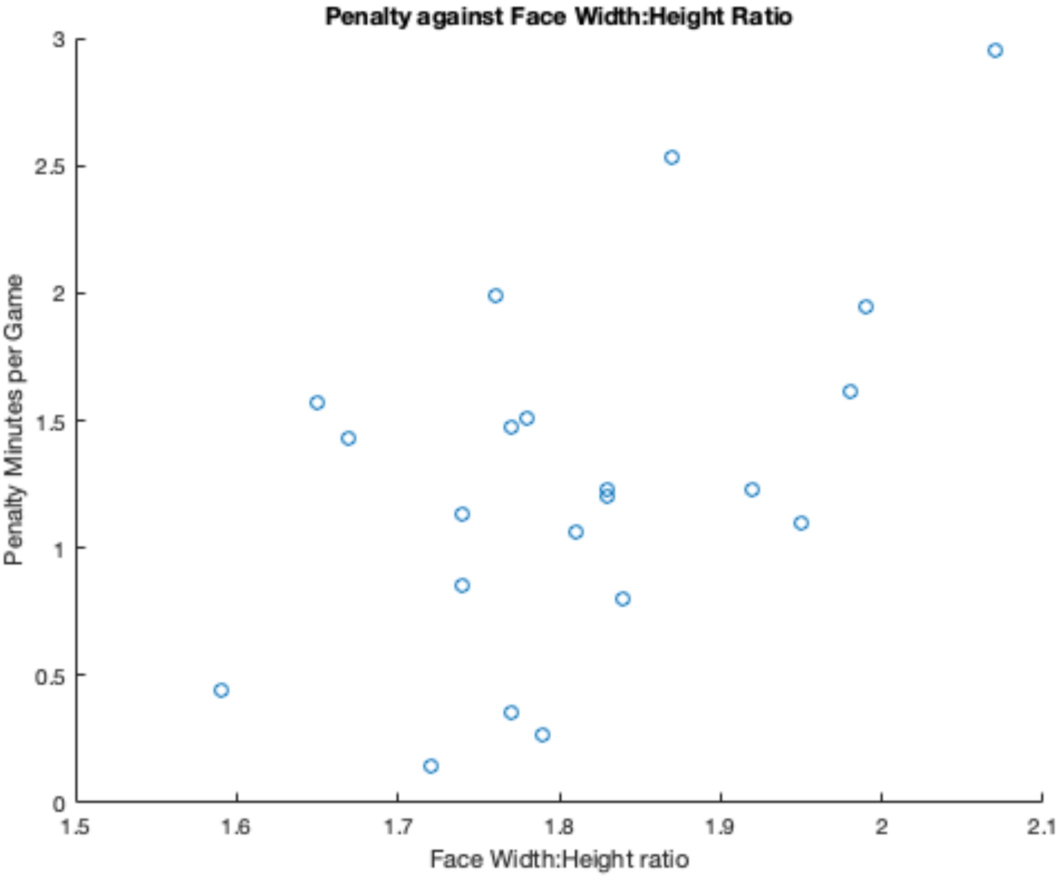
With a p-value > alpha = 0.05 and a F statistic of 0.72 less than the critical value 2.7961, we fail to reject the null hypothesis. There is not enough statistically significant evidence to determine any difference in mean precipitation in Pittsburg for any of the days of the week.

Question 5

```
% Question 5
```

```
face =
[1.59,1.67,1.65,1.72,1.79,1.77,1.74,1.74,1.77,1.78,1.76,1.81,1.83,1.83,1.84,1.87,1.92,1.9
5,1.98,1.99,2.07];
penalty =
[0.44,1.43,1.57,0.14,0.27,0.35,0.85,1.13,1.47,1.51,1.99,1.06,1.2,1.23,0.8,2.53,1.23,1.1,1
.61,1.95,2.95];

% a
hold on;
scatter(face,penalty);
xlabel('Face Width:Height ratio');
ylabel('Penalty Minutes per Game');
title('Penalty against Face Width:Height Ratio');
hold off;
```



```
% b
```

Based on the plot above, the plot looks like the assumptions of linear regression are met. Based on the direction it looks like the data is going, I expect the future analysis results to show that the slope is positive.

```
% c
mean(face); % 1.8129
mean(penalty); % 1.2767
```

The mean face width:height ration is 1.8129 and the mean penalty per game in minutes is 1.2767 minutes.

```
% d
```

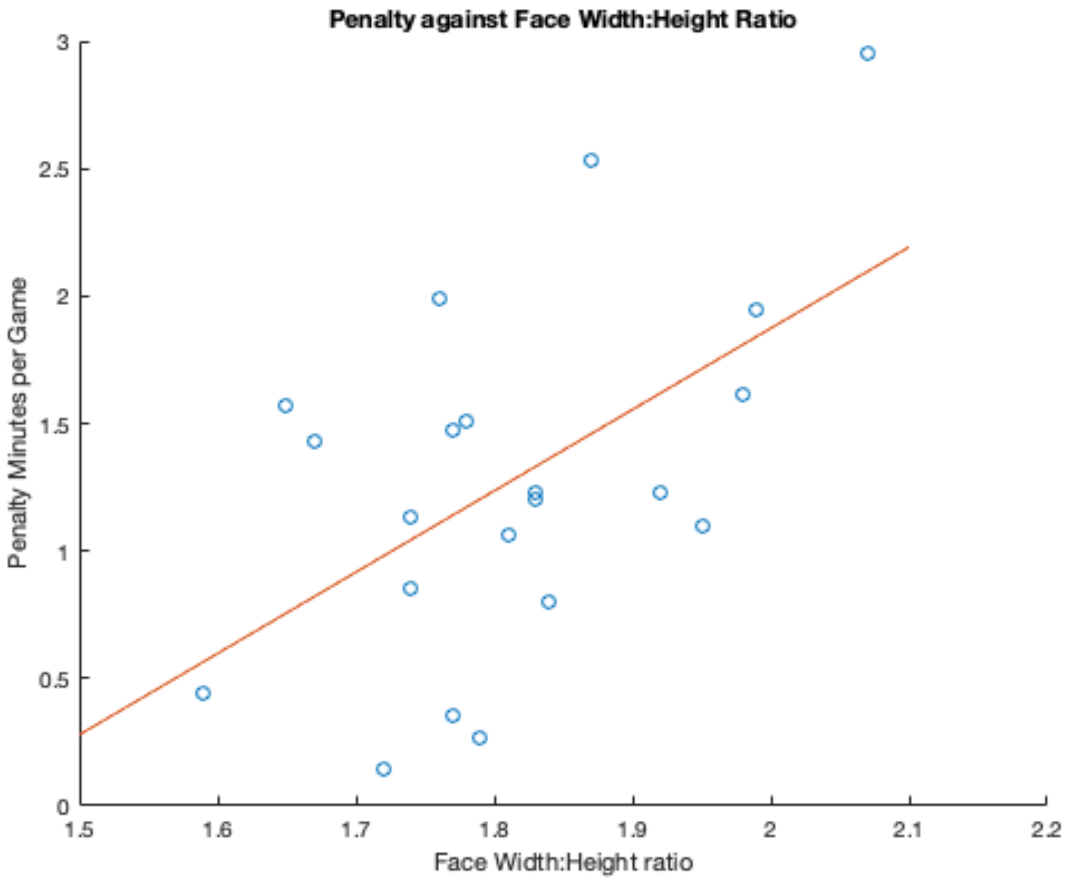
```
b = sum((face - mean(face)).*(penalty - mean(penalty)))/sum((face - mean(face)).^2); %
3.189
```

Using the equation for the regression slope estimate, for every unit increase in face ratio, the penalty minutes increase about 3.189 minutes.

```
% e
a = mean(penalty) - b*mean(face); % -4.5045
```

Using the equation for the estimate of the intercept, when the face ratio is 0, we expect the number of penalty minutes to be about -4.5045. Since it is impossible to have negative time, and the x limit of this plot does not go less than 1.5, it is unlikely this phenomena would happen.

```
% f
% y = 3.189x - 4.5045
x = 1.5:0.05:2.1;
y = 3.189*x - 4.5045;
hold on;
scatter(face,penalty);
xlabel('Face Width:Height ratio');
ylabel('Penalty Minutes per Game');
title('Penalty against Face Width:Height Ratio');
plot(x,y);
hold off;
```


Penalty against Face Width:Height Ratio
Penalty Minutes per Game
Face Width:Height ratio