University at Buffalo
The State University of New York

ISE 507: Design & Analysis of Experiments

Group Mini Project 1

Instructions

This project is open-book and open-notes. It is best to attempt the work on your own and then ask questions of your group mates. Try to focus on understanding the reasons for any differences between your answer and someone else's approach, rather than just copying someone else's approach. Please refer to the syllabus to determine when this mini-project is due. Late mini-projects will not be accepted, except in extenuating circumstances (e.g., family emergency, illness, etc.), with official documents and instructors permission.

Provide your full name and student ID number on your submission.

Problem:
You are the engineer on a team that is formulating a "Brand New" cake mix. Actually, the team is formulating two new mixes. The overall goal of is to bring to market a mix that produces a moister cake than the current "super-moist" version of the competitor.

<u>The objectives of the current experiment</u>: A small part of the overall program is to decide 1) whether we can be reasonably sure that either of the two mixes is moister than the current "super-moist" one by an average of 15%, and if so 2) which of the two new mixes is moister. The reasoning for these objectives is that less than 15% average difference would probably not be distinguishable by enough cake eaters to economically justify bringing a new product to market. However, if this 15% difference does exist among either of the new mixes the cost of bringing either to market would be the same. (Past work shows that this 15% figure is a reasonable one.)

Your team has developed this protocol for the experiment:

a. Mix up one batch of each of the 3 mixes. (You had asked that numerous batches be made up for each mix, but you were told that the results would be the same mix after mix.)
b. Pour each of the 3 mixes into one of the cupcake molds for 8 pans. Each pan has 4 mold positions , but only 3 of the 4 are being used in the experiment. This makes a total of 24 cupcakes.
c. Number the pans 1-8 at random. Insert small thermometers into each of the 3 mixes for pans 2, 3, 6, and 7 (for the total of 12 thermometers).
d. Preheat the oven to 350°F. When the oven has reached this temperature, arrange the 8 pans into the oven as follows:

	Left Side		Right Side	
Top	1	3	5	7
Bottom	2	4	6	8

(It was anticipated that any oven location effects will be of the nature "top → bottom" or "left → right".)

e. Let the cupcakes cook until the average temperature of the small thermometers is 180°F. Then remove the pans from the oven.
f. Measure the amount of moisture in each of the 24 cupcakes by 1) taste on a standardized 1-10 scale, 2) an analytical technique.

   The taste test is done first and will use ½ of each cupcake (the half chosen will be decided at random.) This test will be done first and will take about ½ hour total. The remaining ½ cupcake will be set aside for the analytical measurement at a later time. The analytical measurement will use one machine, will require 15 minutes per cupcake and will be run on a random order of the cupcakes.

This protocol was accepted and followed, except that the thermometers that were to be placed into the mixes for pan 7 were not available when the experiment was actually run.

You have been assigned to analyze only the analytical measurements from this experiment and answer the questions found after the dataset.

Data: (Mix – new mixes are #1 and #2, current mix is #3)

Moisture	Mix	Oven Location	Random Order Of Measure
4.50	1	1	11
8.61	1	2	13
8.55	1	3	24
5.39	1	4	6
8.00	1	5	15
8.55	1	6	23
6.77	1	7	10
7.39	1	8	2
3.72	2	1	9
8.11	2	2	12
6.72	2	3	3
6.05	2	4	19
5.33	2	5	7
6.72	2	6	1
6.72	2	7	17
7.55	2	8	20
3.77	3	1	22
7.55	3	2	21
7.27	3	3	18
3.22	3	4	14
4.11	3	5	8
6.27	3	6	16
4.00	3	7	5
5.50	3	8	4

Questions:

a. Paying close attention to the experimental protocol described above, what is Experimental Design was used in this experiment? Provide an explanation for your answer. (5 points)
b. Referencing the paper found on UBLearns, is this study enumerative or analytic? Explain why with 1-2 sentences. (5 points)
c. Analyze this data using the most appropriate ANOVA model with Minitab. Be sure to make plots of the raw data and include Minitab Session window output. Provide a descriptive analysis of all figures and output. (10 points)
d. Examine the assumptions (normality, independence, and so on). Show me (words, graphs, and/or tables) how you examined all the assumptions. Which of the assumptions, if any, are violated? (5 points)
e. If any of the assumptions have been violated, provide a reasonable, physical explanation for how it could have happened . If you did not find any violations, explain what was done in the experiment to minimize potential violations. (5 points)
f. If the block effects are important, interpret these effects in light of the experimental protocol. Interpretations should be as physically realistic as possible. (5 points)
g. Now that you have analyzed this dataset, think about how the experiment was setup and how the data was collected. What concerns do you have about your ability to make a prediction regarding the next batch of cupcakes made? Hint: think about data collection and the error term. (10 points)

h. Answer the two questions raised in the objectives of the experiment. Use the appropriate method of comparison of means here and defend its use. (If you wish, you may pretend that none of the assumptions have been violated, since the course is not gone far enough to show you how to correct them.) (5points)

Problem statement

The current project aims to develop two new cake mixes that produce moister cakes than the current "super-moist" version of a competitor. To achieve this, we determine if either of the two new cake mixes is significantly moister than the current "super-moist" mix with a minimum difference of 15%, such that it is distinguishable by enough cake eaters to economically justify bringing a new product to market. If this difference exists, we want to determine which of the two new mixes produces the moister cake.

a) **Paying close attention to the experimental protocol described above, what is Experimental Design was used in this experiment? Provide an explanation for your answer.**

It is determined that the experimental design used in this experiment is the randomized complete block design (RCBD). In this experiment, the oven locations are considered as the blocks or the nuisance factors. It is also given that the 8 pans containing the 3 samples of the cake mixes are arranged randomly over eight locations in the oven. The cupcakes are also randomly assigned to the oven locations and the order in which they were measured analytically was also random. The cupcakes are also assigned randomly within each block so that the effect of blocking variables is evenly spread across the different treatments. The Randomized Complete Block Design is especially useful when the blocking variable is a significant source of variability and when there is a limited number of experimental units available. This is shown in the problem statement that the effect of the location of pans in the oven (blocking variable) varies with respect to the arrangement of the pans (top to bottom and left to right).

b) **Referencing the paper found on UBLearns, is this study enumerative or analytic? Explain why with 1-2 sentences.**

The given study is analytic by nature. The study is focused on measuring the amount of moisture in the cupcakes using an analytical technique, and to improve the moisture content in the new mixes by at least 15 percent more than the current super moist cupcake, which involves running a machine-based measurement on a random order of the cupcakes. This suggests that the study was designed to analyze a sample of the cupcakes rather than enumerate all possible cupcakes that could be produced.

c) **Analyze this data using the most appropriate ANOVA model with Minitab. Be sure to make plots of the raw data and include Minitab Session window output. Provide a descriptive analysis of all figures and output.**

After arranging the data in Minitab, we have conducted an ANOVA analysis of the experiment. Using Stats>ANOVA>General Linear model, we have performed two-way ANOVA analysis. Two-way ANOVA analysis is considered here because there are two factors in this experiment-Cake mixes and Oven locations. We have obtained the following results:

Factor Information

Factor	Type	Levels	Values
Oven Location	Random	8	1, 2, 3, 4, 5, 6, 7, 8
Mix	Random	3	1, 2, 3

Fig-1

Figure 1 shows the factors that are considered in the experiment. Levels represent the number of treatments that are done at random. Three cake mixes are experimented with 8 different treatments (locations).

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Oven Location	7	40.396	5.7708	8.70	0.000
Mix	2	16.259	8.1297	12.25	0.001
Error	14	9.290	0.6636		
Total	23	65.945			

Fig-2

Figure-2 shows ANOVA table consisting of each factor.

Null hypothesis:

For the Oven Location factor, the null hypothesis is that the means of the dependent variable (moisture) are equal across all levels of Oven Location. The alternative hypothesis is that there is a significant difference between the means across all levels of the location.

For the Mix factor, the null hypothesis is that the means of the dependent variable are equal across all levels of Mix. The alternative hypothesis is that there is a significant difference between at least one pair of means.

F-Values interpretation

The F-value is the test statistic used to determine whether the term is associated with the response.
In this case, it can be interpreted that F-value is that the variability between the groups due to the Oven Location factor and the Mix factor, respectively, is statistically significant. This means that the differences in the dependent variable across different levels of Oven Location and Mix are not likely due to chance and suggest that these factors are important in explaining the variation in the dependent variable.

ANOVA Table Interpretation

Oven Location: The F-value of 8.70 and p-value of 0.000 indicate that the location of the oven has a statistically significant effect on the response variable.
Mix: The F-value of 12.25 and p-value of 0.001 indicate that the mix also has a statistically significant effect on the response variable.
We reject the null hypothesis for the both factors.
Error: The mean square error is 0.6636, which represents the unexplained variation in the response variable.
Total: The adjusted sum of squares for the Total row represents the total variation in the dependent variable that is explained by the model.

Adjusted sum of squares

Adj SS values represent the amount of variability in the response variable that is explained by the Oven Location factor and the Mix factor, after adjusting for the other factors in the model.

The Adj SS for Oven Location is 40.396, which means that this factor explains 40.396 units of variability in the response variable, after adjusting for the effects of the Mix factor. Similarly, the Adj SS for Mix is 16.259, which means that this factor explains 16.259 units of variability in the response variable, after adjusting for the effects of the Oven Location factor.

Considering, F-value and P-value for each factor we have determined that these factors have a significant effect on the response variable.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.814591	85.91%	76.86%	58.60%

Fig-3

This table provides information on the goodness of fit of the regression model from the ANOVA analysis.

- R-sq is 85.91%. This implies that 85.91% of the variation in the response variable is explained by the factors in the model.
- R-sq(adjusted) is 76.86%. This is slightly lower than R-sq as the model includes multiple independent variables.
- R-sq and R-sq (adj) are closer to each other. This implies that the model is good in explaining the significant independent variables.
- R-sq(pred) is 58.60%, which means that the model is expected to explain 58.60% of the variation in the new data. This suggests that the model might not explain the variation in the new data completely.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	6.265	0.166	37.68	0.000
Oven Location				
1	-2.269	0.440	-5.16	0.000
2	1.825	0.440	4.15	0.001
3	1.248	0.440	2.84	0.013
4	-1.379	0.440	-3.13	0.007
5	-0.452	0.440	-1.03	0.322
6	0.915	0.440	2.08	0.056
7	-0.435	0.440	-0.99	0.339
8	0.548	0.440	1.25	0.233
Mix				
1	0.955	0.235	4.06	0.001
2	0.100	0.235	0.42	0.678
3	-1.054	0.235	-4.48	0.001

Fig-4

The coefficients table (fig-4) from ANOVA analysis shows the coefficients and statistics related to factors, mix, and oven location. The coefficient represents the estimated value of the regression coefficient for each factor (oven location and mix) in the linear regression model.

These coefficients represent the change in the response variable associated with a unit change in the factor, holding all other independent variables (factors) constant, given that the p-value is less than the significance level. For example, for location 1 the response value is expected to decrease by 2.269 units when the level is increased from the reference or baseline (in the regression model).

Null hypothesis: Here the null hypothesis can be stated as the factors (oven locations and mix) has a significant effect on the response variable (moisture level).
If P<0.05, we reject the null hypothesis and state that the corresponding oven location or mix is statistically significant and has a significant effect on the moisture levels of the cupcakes.

Interpreting p-values from the table, **these factors have a significant effect on the response variable**.
Locations: 1, 2, 3, 4
Mixes: 1 & 3

Regression Equation

$$\text{Moisture} = 6.265 - 2.269\ \text{Oven Location\_1} + 1.825\ \text{Oven Location\_2} + 1.248\ \text{Oven Location\_3} - 1.379\ \text{Oven Location\_4} - 0.452\ \text{Oven Location\_5} + 0.915\ \text{Oven Location\_6} - 0.435\ \text{Oven Location\_7} + 0.548\ \text{Oven Location\_8} + 0.955\ \text{Mix\_1} + 0.100\ \text{Mix\_2} - 1.054\ \text{Mix\_3}$$

Equation treats random terms as though they are fixed.

Fig-5

From this regression equation, the response is Moisture, which is predicted by several independent variables from oven location and mix levels. Similar to coefficients table, the coefficients in the equation represent the estimated effect of each independent variable on the moisture, while holding the other independent variables constant. However, significance of each factor effect on the response is considered when corresponding p-values are less than 0.05. For location 1 the response value is expected to decrease by 2.269 units when the level is increased from the reference or baseline when null hypothesis is rejected to consider its significance.

Fits and Diagnostics for All Observations

Obs	Moisture	Fit	Resid	Std Resid
1	4.500	4.951	-0.451	-0.73
2	3.720	4.096	-0.376	-0.60
3	3.770	2.943	0.827	1.33
4	8.610	9.045	-0.435	-0.70
5	8.110	8.190	-0.080	-0.13
6	7.550	7.036	0.514	0.83
7	8.550	8.468	0.082	0.13
8	6.720	7.613	-0.893	-1.44
9	7.270	6.459	0.811	1.30
10	5.390	5.841	-0.451	-0.73
11	6.050	4.986	1.064	1.71
12	3.220	3.832	-0.612	-0.98
13	8.000	6.768	1.232	1.98
14	5.330	5.913	-0.583	-0.94
15	4.110	4.759	-0.649	-1.04
16	8.550	8.135	0.415	0.67
17	6.720	7.280	-0.560	-0.90
18	6.270	6.126	0.144	0.23
19	6.770	6.785	-0.015	-0.02
20	6.720	5.930	0.790	1.27
21	4.000	4.776	-0.776	-1.25
22	7.390	7.768	-0.378	-0.61
23	7.550	6.913	0.637	1.02
24	5.500	5.759	-0.259	-0.42

Fig-6

This table shows the results of fitting the ANOVA model to the observed data.

For each observation, the moisture, the fitted value of moisture, the residual, and the standardized residual. The fitted values represent the predicted moisture for each observation. The residuals represent the differences between the observed moisture content and the predicted moisture content. The standardized residuals show how many standard deviations each residual is from the mean residual, which can help identify outliers in the data.

From this table, we can see that the residuals range from -1.054 to 1.232, and the standardized residuals range from -1.71 to 1.98. We observe that some of the standardized residuals are quite large; there does not appear to be any clear pattern in the residuals, and they are not consistently positive or negative. Therefore, it seems that the model fits the data reasonably well, which can also be supported by the R-sq value.

Variance Components, using Adjusted SS

Source	Variance	% of Total	StDev	% of Total
Oven Location	1.70242	51.60%	1.30477	71.83%
Mix	0.933262	28.29%	0.96605	53.19%
Error	0.663559	20.11%	0.81459	44.85%
Total	3.29924		1.81638	

Fig-7

The first column shows the sources of variation, which are oven Location and mix. The largest source of variation is Oven Location, which accounts for 51.60% of the total variance, followed by Mix, which accounts for 28.29% of the total variance. The remaining 20.11% of the total variance is attributed to Error, which represents the unexplained variation in the model.

All-in-all, from ANOVA model we can interpret that

- Locations 2 and 3 have positive significance on the moisture, whereas location 1 and 4 have negative significant effect on the moisture.
- From the company's 2 mixes (mix 1 and mix 2), mix 1 has significant association with the moisture response.
- From R-sq and adj R-sq of the model, we can interpret that the model explains significant variation. However, Pred R-sq implies that model is relatively not good in explaining the new data.
- Oven location accounts slightly more that 50% of the variation, whereas mix accounts for approximately 30% of the variation. Error of about 21% represents the unexplained variation.

d) Examine the assumptions (normality, independence, and so on). Show me (words, graphs, and/or tables) how you examined all the assumptions. Which of the assumptions, if any, are violated?

The following are the assumptions made for the ANOVA analysis:

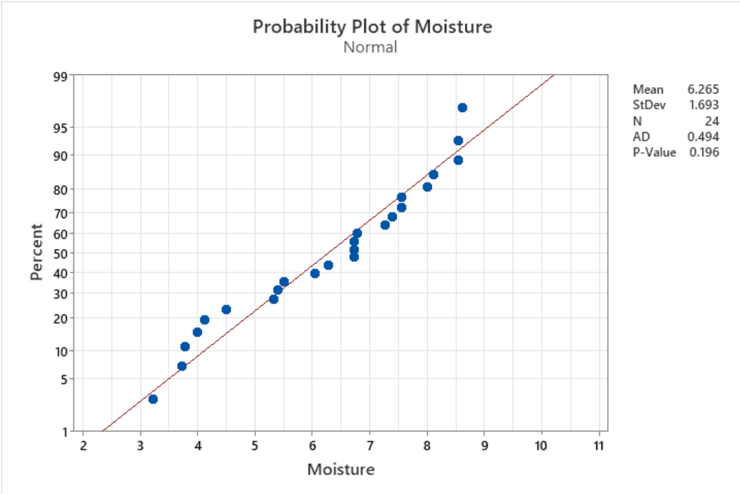
1. **Normality:** The data within each group should be normally distributed. Normality can be assessed through a histogram or a normal probability plot.
2. **Homogeneity of variance:** The variance of the data within each group should be approximately equal. This assumption is known as homogeneity of variance or homoscedasticity. This can be checked using a plot of the residuals (the differences between the observed values and the predicted values) against the predicted values.
3. **Independence:** The observations within each group should be independent of each other. This means that the value of one observation should not be affected by the value of another observation within the same group.

Normality:

To test the normality of the data (moisture levels in the cupcakes) the probability plot is drawn along with the graphical summary of the given data.

Null Hypothesis: The null hypothesis for the normality test is that the data is normally distributed.

Alternative Hypothesis: The Alternative Hypothesis is that the data is not normally distributed.



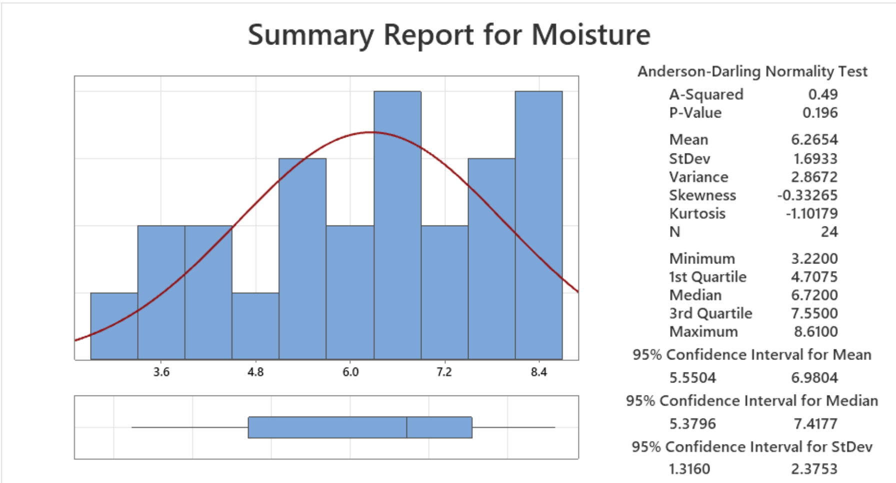


Fig-8

After the normality test is conducted (Anderson Darling Normality Test), the results show that the p value of the normality plot is 0.196 which is greater than the significance value of 0.05. This shows that we fail to reject the null hypothesis that the data of the moisture levels form a normal distribution. Hence the assumption that the data is normally distributed is not violated.

Homogeneity of variance:

To check the validity of this assumption, the test for equal variances has been conducted in Minitab. This test has been done separately. One for moisture levels as the response and the Oven Location as the Factor and Two, moisture levels as the response and the mixes as the factors.

Equivalence of Variance Test Results:
Moisture versus Oven Location

Method

Null hypothesis	All variances are equal
Alternative hypothesis	At least one variance is different
Significance level	α = 0.05

Bartlett's method is used. This method is accurate for normal data only.

95% Bonferroni Confidence Intervals for Standard Deviations

Oven Location	N	StDev	CI
1	3	0.43662	(0.181792, 7.8043)
2	3	0.53028	(0.220792, 9.4786)
3	3	0.93895	(0.390948, 16.7834)
4	3	1.48062	(0.616480, 26.4654)
5	3	1.98953	(0.828373, 35.5620)
6	3	1.20760	(0.502804, 21.5853)
7	3	1.58502	(0.659950, 28.3316)
8	3	1.14019	(0.474736, 20.3804)

Individual confidence level = 99.375%

Fig-9

Tests

Method	Test Statistic	P-Value
Bartlett	5.18	0.638

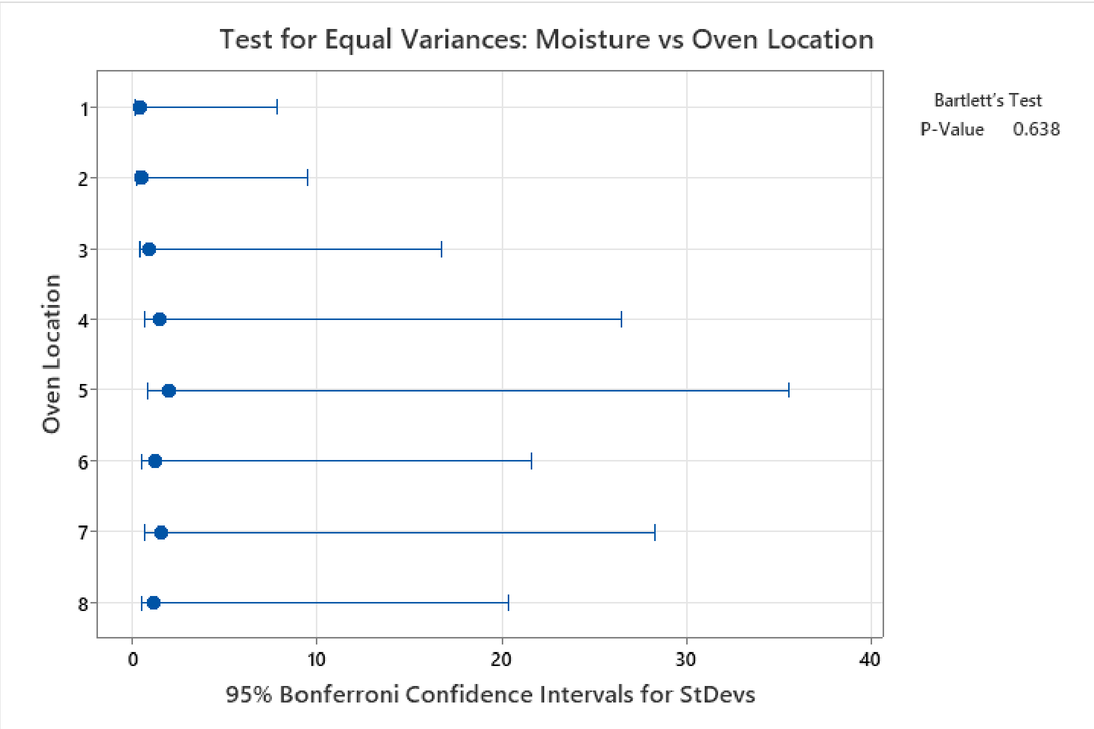


Fig-10

The p-Value for the equivalence test is obtained as 0.638 (Fig-10). This is greater than the significance level value of 0.05. This shows that we fail to reject the null hypothesis that there is homogeneity in the variances of the factors and responses. This proves that the assumption of equal variances is not violated. The Bonferroni Confidence Interval the individual confidence level indicates how confident we can be that an individual confidence interval contains the population standard deviation of that specific group. For example, we can be 99.375% confident that the standard deviation for the moisture levels with oven location 1 as the factor is within the confidence interval (0.181792, 7.8043).

Moisture versus Mixes

Method

Null hypothesis	All variances are equal
Alternative hypothesis	At least one variance is different
Significance level	α = 0.05

Bartlett's method is used. This method is accurate for normal data only.

95% Bonferroni Confidence Intervals for Standard Deviations

Mix	N	StDev	CI
1	8	1.56100	(0.94864, 3.82308)
2	8	1.36245	(0.82798, 3.33680)
3	8	1.67480	(1.01780, 4.10180)

Individual confidence level = 98.3333%

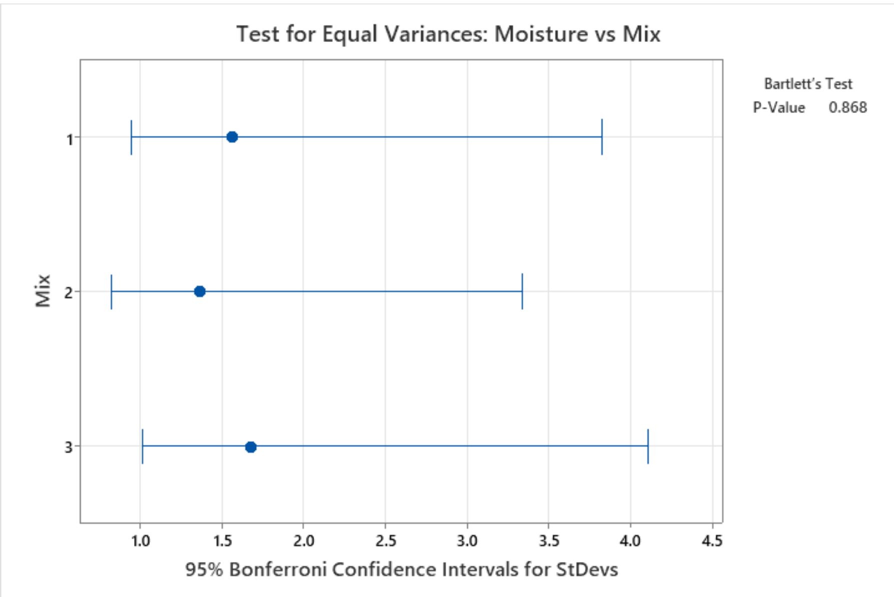


Fig-11

The p-Value for the equivalence test is obtained as 0.868 (fig-11). This is greater than the significance level value of 0.05. This shows that we fail to reject the null hypothesis that there is homogeneity in the variances of the factors and responses. This proves that the assumption of equal variances is not violated. The Bonferroni Confidence Interval the individual confidence level indicates how confident we can be that an individual confidence interval contains the population standard deviation of that specific group. For example, we can be 98.3333% confident that the standard deviation for the moisture levels with mix 1 as the factor is within the confidence interval (0.94864, 3.82308).

Independence:

To check the independence assumption in Minitab, we used a time-series plot or residuals vs order plot. These plots can help to detect any patterns or trends in the data that may violate the independence assumption. In this case, the time series plot and the residuals vs order of observations plots are used to verify the independence of the data. The plots are shown here.

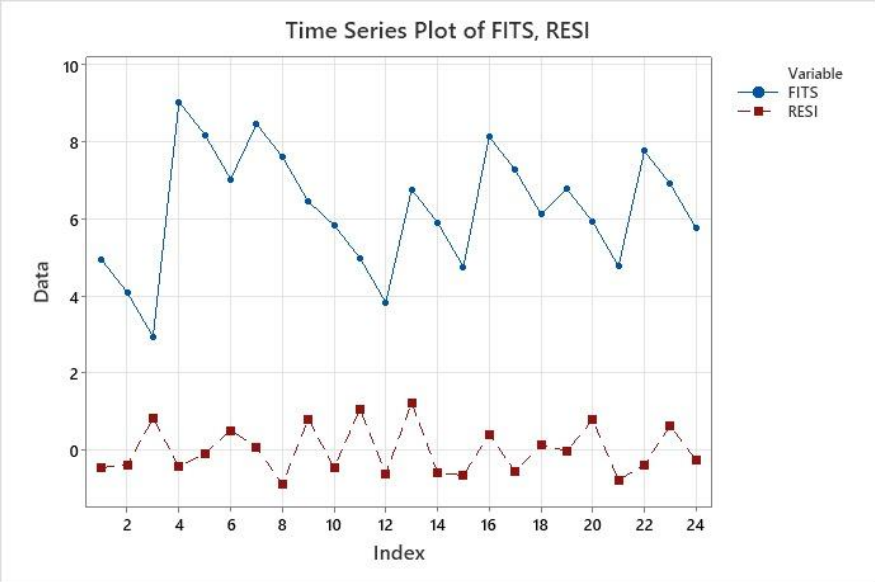


Fig-12

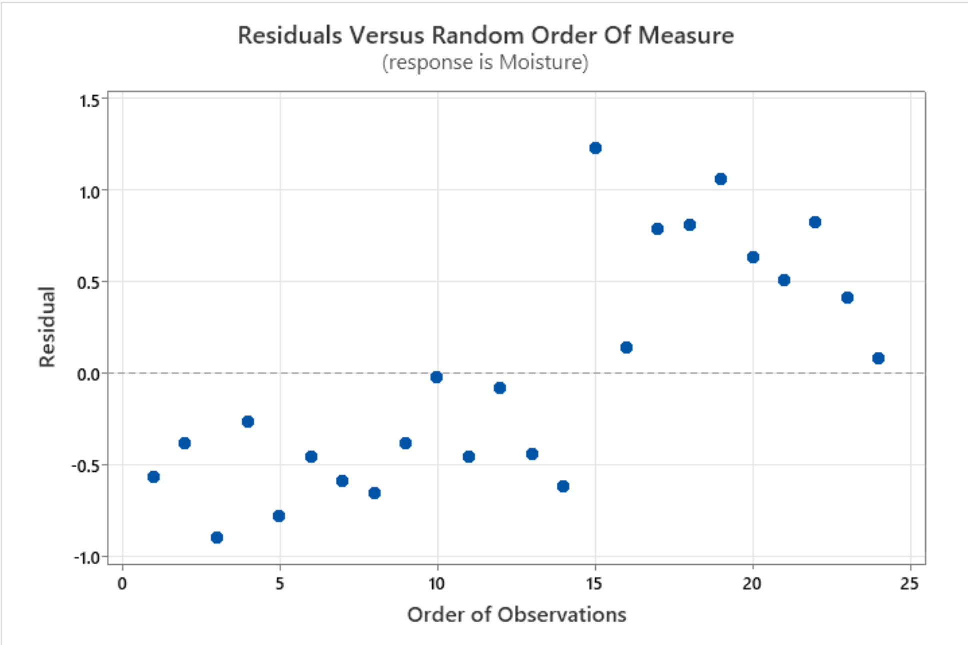


Fig-13

The time series plot (fig-12) for the residuals shows that there is a cyclic pattern for the observation orders 8 to 14 which shows that the data is not independent. The Residuals vs order of measure (fig-13) shows a sudden shift in the values of the residuals. This also indicates that the residual data is not independent. It proves that the assumption of ANOVA that the data is independent is violated.

e) If any of the assumptions have been violated, provide a reasonable, physical explanation for how it could have happened. If you did not find any violations, explain what was done in the experiment to minimize potential violations.

The independence assumption of ANOVA has been violated in this experiment; it means that the observations are not independent of each other. This ANOVA assumption requires that each observation is independent of all other observations, which means that the response variable is not influenced by any other factors other than the factors being studied.

One potential explanation for the violation of the independence assumption in this experiment is that certain cakes were baked together in the same oven or in close proximity to each other. This could have caused their moisture content to be affected by variations in baking time, temperature, or airflow, which could have resulted in correlated observations that go against the assumption of independence. It is given that cupcakes are cooked until the average temperature is 180 degrees, which means that there might be variations in the heat distribution of the oven. Considering that the thermometer from pan 7 is not available, we can infer that it may affect the independency of the observations. Moreover, the analytical measurement technique might also be the cause for the dependency in the response. We are not sure whether the equipment is validated and has bias. Therefore, we can attribute the violation of independency assumption with the experimental set up.

f) If the block effects are important, interpret these effects in light of the experimental protocol. Interpretations should be as physically realistic as possible.

- Possible block effects include differences in oven temperature owing to inconsistent heat distribution and airflow, humidity, variations in ingredient quality or quantity, or differences in mixing or baking techniques.
- If these block effects are not accounted for, they may lead to biased or confounded results, making it difficult to draw valid conclusions from the experiment.
- In this experiment, the locations of the pans with the 3 cupcakes with different mixes in the oven can be considered as the blocks.
- To consider the impacts of oven location on cake moisture, this block arrangement was chosen. It is stated in the problem statement that the location effects will be of the order of top to bottom and left to right.
- Most electric and gas ovens have two heat sources – one at the top and one at the bottom. When the oven is turned on to preheat, both heating elements come on to heat the oven.
- When the oven reaches the desired temperature, the top element switches off and the bottom one cycles on and off to maintain a temperature at an average of the one you set when you turned the oven on.
- It is also seen in practice since the hot rises, the racks at the top of the oven are consistently hotter, while the bottom will swing hotter and then cooler as you bake.
- Overall, following the experimental protocol suggests that cupcakes baked in different parts of the oven will have different moisture levels, even after adjusting the type of mix used. So block effects are important and considered.

g) Now that you have analyzed this dataset, think about how the experiment was setup and how the data was collected. What concerns do you have about your ability to make a prediction regarding the next batch of cupcakes made? Hint: think about data collection and the error term.

The experiment involved baking 24 cupcakes (8 pans of 3 pans each), baking each of the 3 mixes in random order, and placing them in an oven preheated to 350°F. A thermometer was placed in each of the three mixtures in pans 2, 3, 6, and 7 to monitor the temperature. Cupcakes were taken out of the oven when the average temperature on the little thermometer reached 180°F.

The cupcakes were then taste-tested on a standard scale of 1 to 10, followed by the moisture content determined by an analytical measurement. Half of each cupcake was used for taste testing and the other half was set aside for analytical measurements.

Data collected in this experiment included the moisture content of each cupcake, the mix used, oven location, and measurement order. However, it is important to note that a thermometer placed in the pan 7 mixture was not available during the experiment.

The experimental design is a randomized complete block design in which the blocks are assumed to be homogeneous and any variation within blocks is due to chance. However, it is possible that other factors that were not considered in the experimental design could have an impact on the results.

This helps to ensure that any differences observed in the moisture content of the cupcakes can be attributed to the independent variables being tested. The data was collected using a standardized method for measuring moisture content, which should help to reduce measurement error.

Concerns:
The experiment only tested 24 cupcakes, which may not be representative of the larger population of cupcakes that could be made using the same mixes and baking process. Additionally, the experiment only used one machine for the analytical measurements, which could introduce variability if the machine is not perfectly calibrated or if there are any other issues with the machine. A larger sample size would provide more confidence in the results.
Only two variables, oven location and mix, were tried in the experiment, leaving out other variables that might influence how moist cupcakes are. The moisture content may also be impacted by additional elements like baking time, and temperature.

Another concern is the error term is variability in the data (Fig-7), suggesting that there is a lot of variation in the data that is not explained by the tested factors. Around 21% of the variation is attributed to error. These 21% is not explained by this experiment. Also, there may be some inherent variability in the moisture content of cupcakes made from the same mix, which could make it difficult to make accurate predictions about future batches. Additionally, the lack of thermometers in pan 7 could also introduce variability into the results. Moreover, the test of tasting the cupcakes is subjective as the preferences vary across individuals and may include bias in their judgement.

The regression equation provides a useful summary of the relationships between the predictor variables and the moisture content of the cupcakes. There are limitations to its ability to predict the moisture content of future batches of cupcakes, as it is shown by R-sq(pred) of around 58% (Fig-3), in the ANOVA analysis.

Overall, these concerns suggest that making accurate predictions about future batches of cupcakes based on this dataset could be challenging. To improve the accuracy of predictions, it may be necessary to conduct additional experiments with larger sample sizes and more rigorous data collection protocols to minimize variability and account for potential sources of error. However, the results of this experiment can still provide valuable insights into the factors that affect moisture content in cupcakes and can inform, experiments with larger sample sizes and more rigorous experimental designs will be conducted in the future.

h) Answer the two questions raised in the objectives of the experiment. Use the appropriate method of comparison of means here and defend its use.
The data given in the problem has been analyzed in Minitab. It is determined that the moisture readings for the various cupcakes are normally distributed. Since the data is normally distributed, the t-test can be used for the comparison of the means of samples.

The normality plot is shown below:

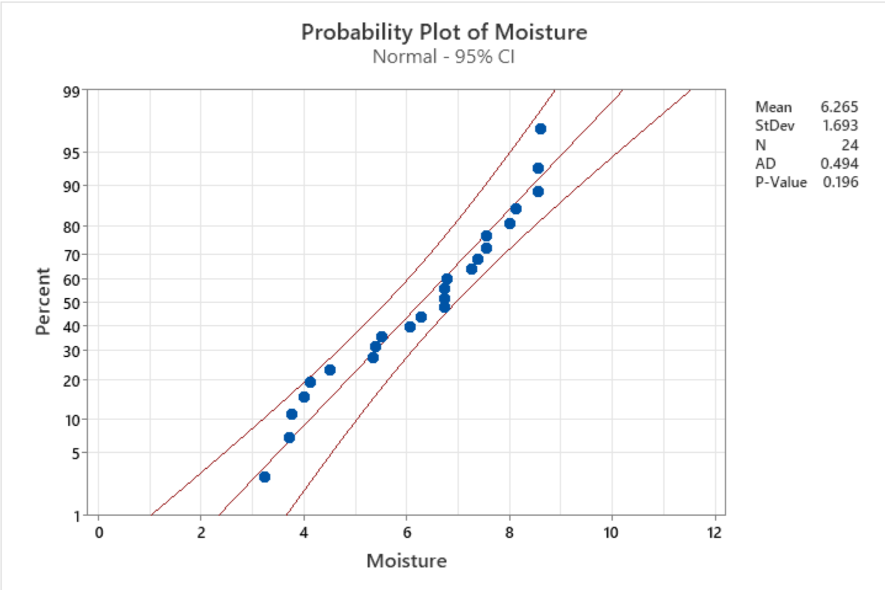


Fig-14

Since the p-value is greater than 0.05, we fail to reject the null hypothesis, where the null hypothesis states that the data is normally distributed. In this experiment, we have done the 2-sample t-test three times to determine if any difference exists between the means of the samples of the moisture values collected. These tests are conducted in the following manner:

2-sample t-test for mixes 1 and 2:

Method

μ_1: population mean of Mix 1
μ_2: population mean of Mix 2
Difference: $\mu_1 - \mu_2$

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Mix 1	8	7.22	1.56	0.55
Mix 2	8	6.36	1.36	0.48

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.17	14	0.263

Fig-15

The 2-sample t-test for mixes 1 and 2 gives a p-value of 0.263 which is greater than 0.05. It shows that we fail to reject the null hypothesis, where the null hypothesis states that there is no statistically significant difference between the means of the 2 samples of cupcake mixes. This implies that there is **no difference between the means of mixes 1 and 2**. This is also true from a practical point of view because both of these mixes are prepared by the same company under identical conditions.

2 sample t-test for mixes 1 and 3:

Method

μ_1: population mean of Mix 1
μ_2: population mean of Mix 3
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

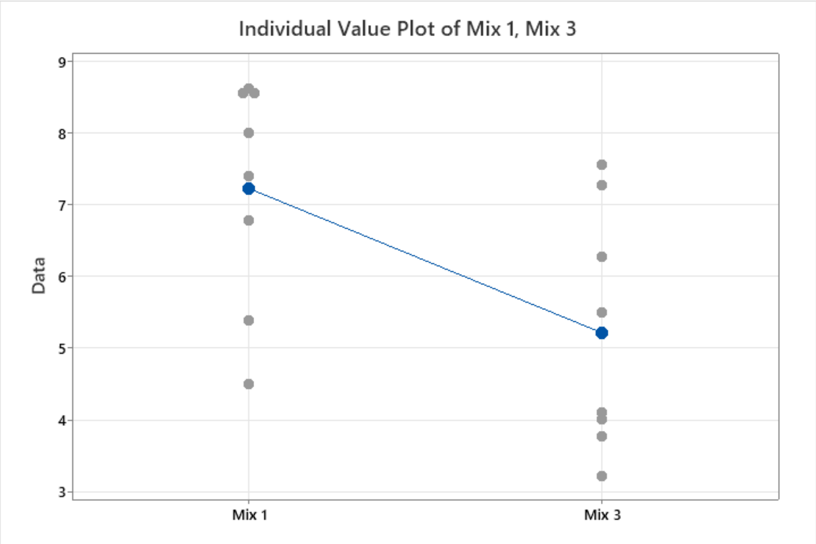
Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Mix 1	8	7.22	1.56	0.55
Mix 3	8	5.21	1.67	0.59

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
2.48	14	0.026



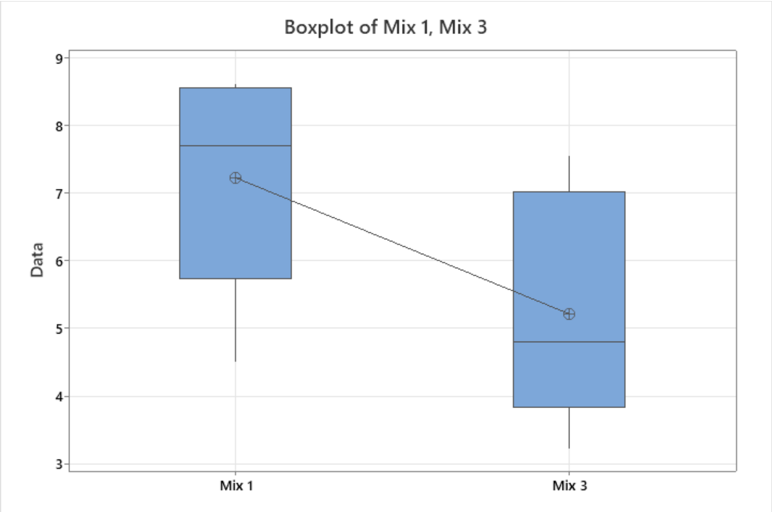


Fig-16

The 2-sample t-test for mixes 1 and 3 gives a p-value of 0.026 which is less than 0.05. It shows that we reject the null hypothesis, where the null hypothesis states that there is no statistically significant difference between the means of the 2 samples of cupcake mixes. This implies that there is a difference between the means of mix 1 and mix 2. This is also shown in the individual and the box plots. Therefore, we can conclude that the mean levels of moisture in both these mixes are **different from each other**. From the t-test, we gather the means of the samples.

Mix 1 - 7.22

Mix 3 - 5.21

It is determined that mix 1 has an average of **38.57 percent** more moisture than mix 3.

2 sample t-test for mixes 2 and 3:

Method

μ_1: population mean of Mix 2

μ_2: population mean of Mix 3

Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Mix 2	8	6.36	1.36	0.48
Mix 3	8	5.21	1.67	0.59

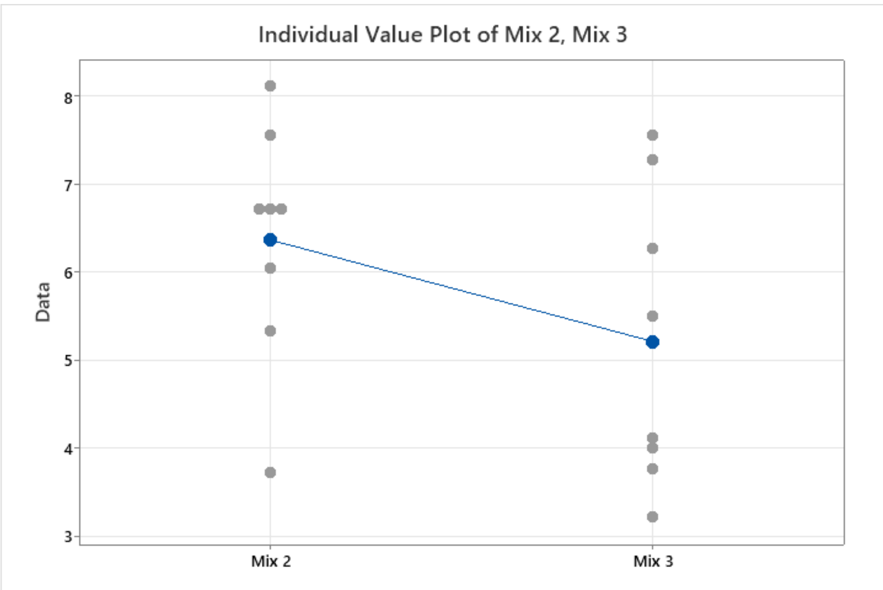
Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$

Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.51	14	0.153

Fig-17



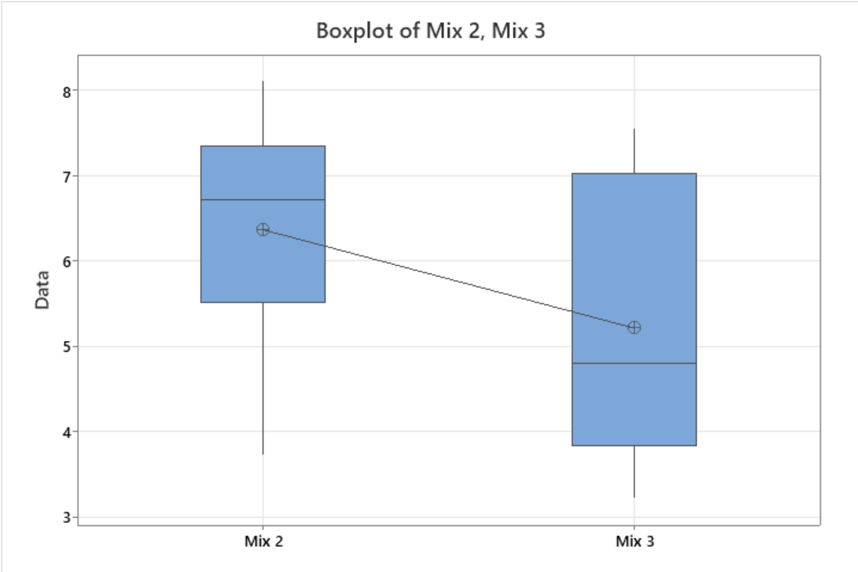


Fig-18

The 2-sample t-test for mixes 2 and 3 gives a p-value of 0.153 which is greater than 0.05. It shows that we fail to reject the null hypothesis, where the null hypothesis states that there is no statistically significant difference between the means of the 2 samples of cupcake mixes. This is also shown in the individual and the box plots. Therefore, we can conclude that the mean levels of moisture in both these mixes are **not statistically different from each other**. From the t-test, we gather the means of the samples.

Mix 2 – 6.36

Mix 3 - 5.21

It is determined that mix 2 has an average of **22.07 percent** more moisture than mix 3.

In conclusion, we can be reasonably sure that Mix 1 is moister than the current super moist one (mix-3) by an average of 15 percent based on the t-tests conducted on the samples collected and also mix 1 is moister than mix 2 and mix 3.