ANOVA: Full Factorial Designs

- Introduction to ANOVA: Full Factorial Designs

- Restrictions

Introduction to ANOVA

Analysis of variance (ANOVA) is a statistical technique used to investigate and model the relationship between a response variable and one or more independent variables.

Each explanatory variable (**factor**) consists of two or more categories (**levels**).

ANOVA tests the **null hypothesis** that the population means of each level are equal, versus the **alternative hypothesis** that at least one of the level means are not all equal.

EXAMPLE 1: A 2003 study was conducted to test if there was a difference in attitudes towards science between boys and girls.

- **Factor** : gender with **Levels** : boys and girls
- **Unit** (Experimental Unit or Subject): each individual child
- **Response Variable**: Each child’s score on an attitude assessment.
- **Null Hypothesis**: boys and girls have the same mean score on the assessment.
- **Alternative Hypothesis**: boys and girls have different mean scores on the assessment.

Introduction to ANOVA

Example 1 can be analyzed with ANOVA or a two-sample t-test discussed in introductory statistics courses.

In both methods the experimenter collects sample data and calculates averages. If the means of the two levels are “significantly” far apart, the experimenter will accept the alternative hypothesis. While their calculations differ, **ANOVA and two-sample t-tests always give identical results in hypothesis tests for means with one factor and two levels.**

Unfortunately, modeling real world phenomena often requires more than just one factor. In order to understand the sources of variability in a phenomenon of interest, **ANOVA can simultaneously test several factors each with several levels.**

Although there are situations where t-tests should be used to simultaneously test the means of multiple levels, doing so create a multiple comparison problem. Determining when to use ANOVA or t-tests is discussed in all the suggested texts at the end of this tutorial.

Introduction to ANOVA

Key steps in designing an experiment include:

1) **Identify factors of interest** and a **response variable.**

2) **Determine appropriate levels** for each explanatory variable.

3) **Determine a design** structure.

4) **Randomize** the order in which each set of conditions is run and collect the data.

5) **Organize the results** in order to draw appropriate conclusions.

This presentation will discuss how to organize and draw conclusions for a specific type of design structure, the full factorial design. This design structure is appropriate for fixed effects and crossed factors, which are defined at the end of this tutorial. Other design structures are discussed in the ANOVA: Advanced Designs tutorial.

Introduction to Multivariate ANOVA

EXAMPLE 2: Soft Drink Modeling Problem (Montgomery p. 232): A soft drink bottler is interested in obtaining more uniform fill heights in the bottles produced by his manufacturing process. The filling machine theoretically fills each bottle to the correct target height, but in practice, there is variation around this target, and the bottler would like to understand better the sources of this variability and eventually reduce it. The engineer can control three variables during the filling process (each at two levels):

Factor A: Carbonation with Levels : 10% and 12%

Factor B: Operating Pressure in the filler with Levels : 25 and 30 psi

Factor C: Line Speed with Levels: 200 and 250 bottles produced per minute (bpm)

Unit: Each bottle

Response Variable: Deviation from the target fill height

Six Hypotheses will be simultaneously tested

The steps to designing this experiment include:

1) **Identify factors of interest** and a **response variable.**

2) **Determine appropriate levels** for each explanatory variable.

Introduction to Multivariate ANOVA

3) Determine a design structure: Design structures can be very complicated. One of the most basic structures is called the **full factorial design**. This design tests every combination of factor levels an equal amount of times. To list each factor combination exactly once

1st Column--alternate every other (2^0) row

2nd Column--alternate every 2 (2^1) rows

3rd Column--alternate every 4th ($=2^2$) row

This is called a 2^3 full factorial design (i.e. 3 factors at 2 levels will need 8 runs). Each row in this table gives a specific treatment that will be run. For example, the first row represents a specific test in which the manufacturing process ran with A set at 10% carbonation, B set at 25 psi, and line speed, C, is set at 200 bmp.

test	A	B	C
1	10%	25	200
2	12%	25	200
3	10%	30	200
4	12%	30	200
5	10%	25	250
6	12%	25	250
7	10%	30	250
8	12%	30	250

*If there were four factors each at two levels there would be 16 treatments.

*If factor C had 3 levels there would be 2*2*3 = 12 treatments.

Introduction to Multivariate ANOVA

4) Randomize the order in which each set of test conditions is run and collect the data. In this example the tests will be run in the following order: 7, 4, 1, 6, 8, 2, 3.

If the tests were run in the original test order, time would be **confounded** (**aliased**) with factor C. Randomization doesn’t guarantee that there will be no confounding between time and a factor of interest, however, it is the best practical technique available to protect against confounding.

run order	test	A Carb	B Pressure	C speed	Results
3	1	10%	25	200	-4
7	2	12%	25	200	1
8	3	10%	30	200	-1
2	4	12%	30	200	5
6	5	10%	25	250	-1
4	6	12%	25	250	3
1	7	10%	30	250	2
5	8	12%	30	250	11

In the following slides A^- will represent carbonation at the low level (10% carbonation) and A^+ will represent carbonation at the high level (12% carbonation). In the same manner B^+, B^-, C^+ and C^- will represent factors B and C at high and low levels.

Introduction to Multivariate ANOVA

5) Organize the results in order to draw appropriate conclusions. Results are the data collected from running each of these 8 = 2^3 conditions. For this example the Results column is the observed deviation from the target fill height in a production run (a trial) of bottles at each set of these 8 conditions.

A	B	C	Results
10%	25	200	-4
12%	25	200	1
10%	30	200	-1
12%	30	200	5
10%	25	250	-1
12%	25	250	3
10%	30	250	2
12%	30	250	11
Grand Mean			2

Once we have collected our samples from our 8 runs, we start organizing the results by computing all averages at low and high levels.

To determine what effect changing the level of A has on the results, calculate the average value of the test results for A^- and A+.

While the overall average of the results (i.e. the **Grand Mean**) is 2, the average of the results for A^- (factor A run at low level) is (-4 + -1 + -1 + 2)/4 = -1

5 is the average value of the test results for A+ (factor A run at a high level)

Main Effects

The B and C averages at low and high levels also calculated.

A	B	C	Results
10%	25	**200**	***-4***
12%	25	**200**	***1***
10%	30	**200**	***-1***
12%	30	**200**	***5***
10%	25	250	-1
12%	25	250	3
10%	30	250	2
12%	30	250	11

	A Avg.	B Avg.	C Avg.
low	-1	-.25	.25
high	5	4.25	3.75

The mean for B^- is (- 4 + 1 + -1 + 3)/4 = -.25

The mean for C^- is (- 4 +1+-1+5)/4 = .25

Notice that each of these eight trial results are used multiple times to calculate six different averages. This can be effectively done because the full factorial design is balanced. For example when calculating the mean of C low (C^-), there are 2 A highs (A^+) and 2 A lows (A^-), thus the mean of A is not confounded with the mean of C. This balance is true for all mean calculations.

Main Effects

Often the impact of changing factor levels are described as effect sizes. A Main Effects is the difference between the factor average and the grand mean.

	A Avg.	B Avg.	C Avg.
low	-1	-.25	.25
high	5	4.25	3.75

Subtract the grand mean (2) from each cell

A Effect	B Effect	C Effect
-3	-2.25	-1.75
3	2.25	1.75

Effect of A^+ = average of factor A+ minus the grand mean
= 5 – 2
= 3

Effect of C^- = .25 – 2 = -1.75

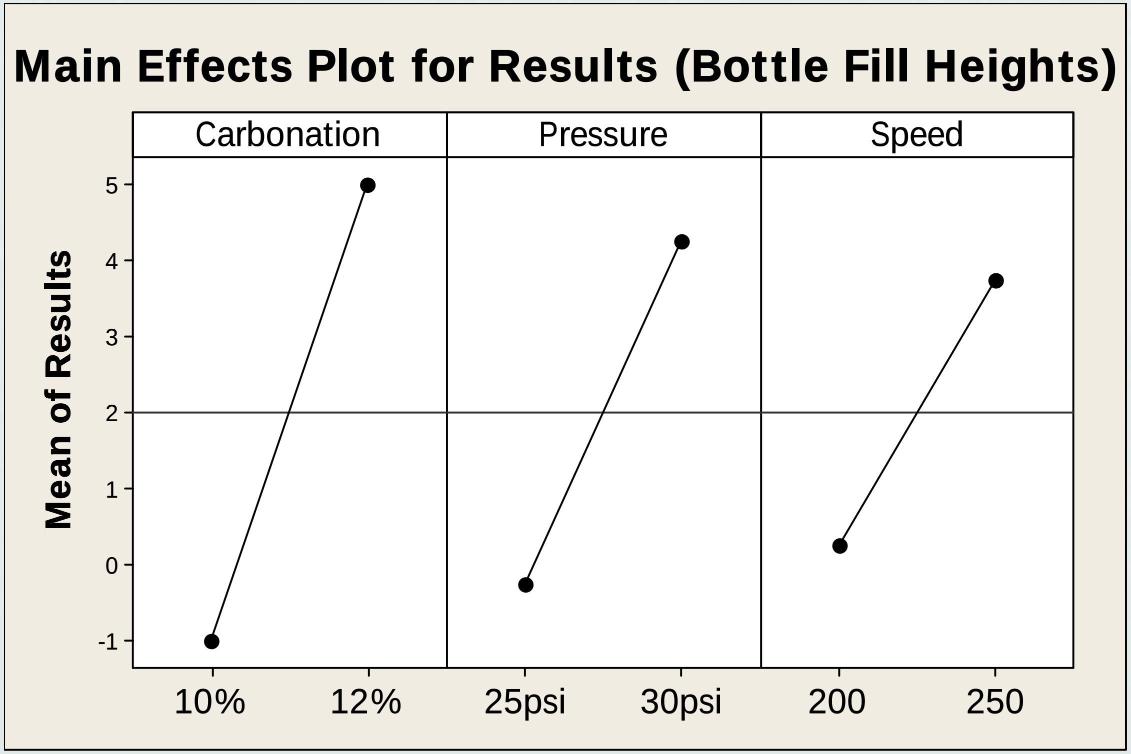
Effect sizes determine which factors have the most significant impact on the results. Calculations in ANOVA determine the significance of each factor based on these effect calculations.

Main Effects

Main Effects Plots are a quick and efficient way to visualize effect size. The grand mean, 2, is plotted as a horizontal line. The average result is represented by dots for each factor level.

The Y axis is always the same for each factor in Main Effects Plots. Factors with steeper slopes have larger effects and thus larger impacts on the results.

	A Avg.	B Avg.	C Avg.
low	-1	-.25	.25
high	5	4.25	3.75



This graph shows that **A$^+$** has a higher mean fill height than **A$^-$**. **B$^+$** and **C$^+$** also have higher means than **B$^-$** and **C$^-$** respectively. In addition, the effect size of factor A, Carbonation, is larger than the other factor effects.

Interaction Effects

In addition to determining the main effects for each factor, it is often critical to identify how multiple factors interact in effecting the results. An interaction occurs when one factor effects the results differently depending on a second factor. To find the AB interaction effect, first calculate the average result for each of the four level combinations of A and B:

A	B	C	Results
10%	25	**200**	***-4***
12%	25	**200**	***1***
10%	30	**200**	***-1***
12%	30	**200**	***5***
10%	25	250	-1
12%	25	250	3
10%	30	250	2
12%	30	250	11

Calculate the average when factors A and B are both at the low level (-4 + -1) / 2 = -2.5

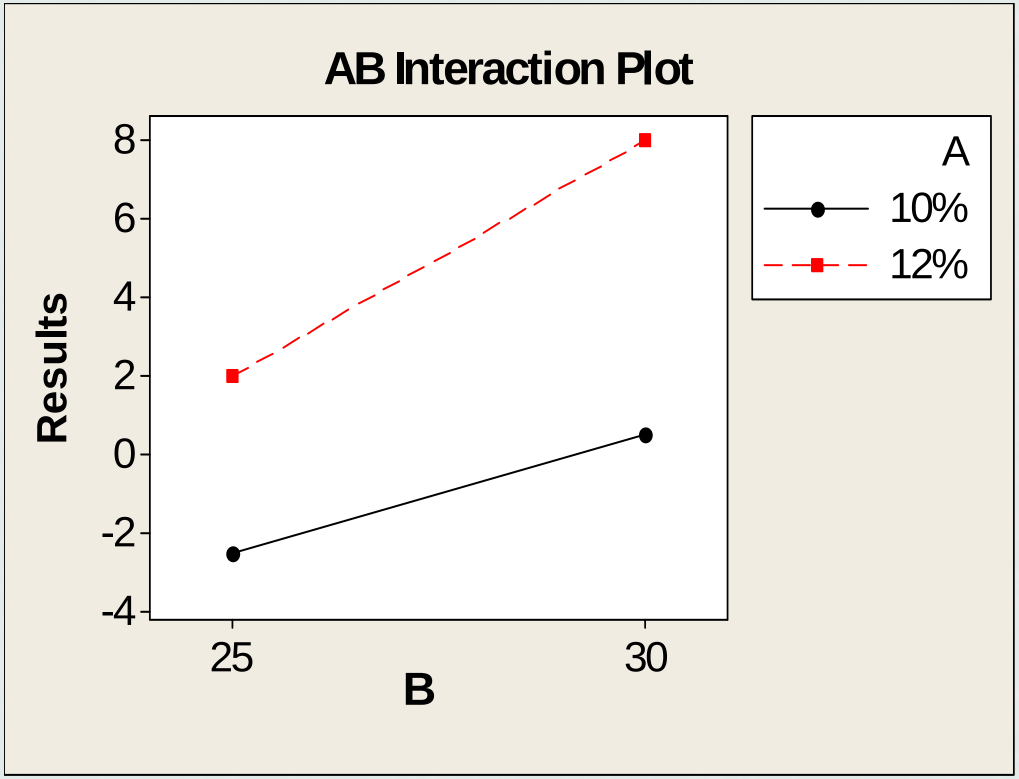
Calculate the mean when factors A and B are both at the high level (5 + 11) / 2 = 8

	AB Avg.
A^-B^-	-2.5
A^+B^-	2.0
A^-B^+	0.5
A^+B^+	8.0

Also calculate the average result for each of the levels of AC and BC.

Interaction Effects

Interaction plots are used to determine the effect size of interactions. For example, the AB plot below shows that the effect of B is larger when A is 12%.

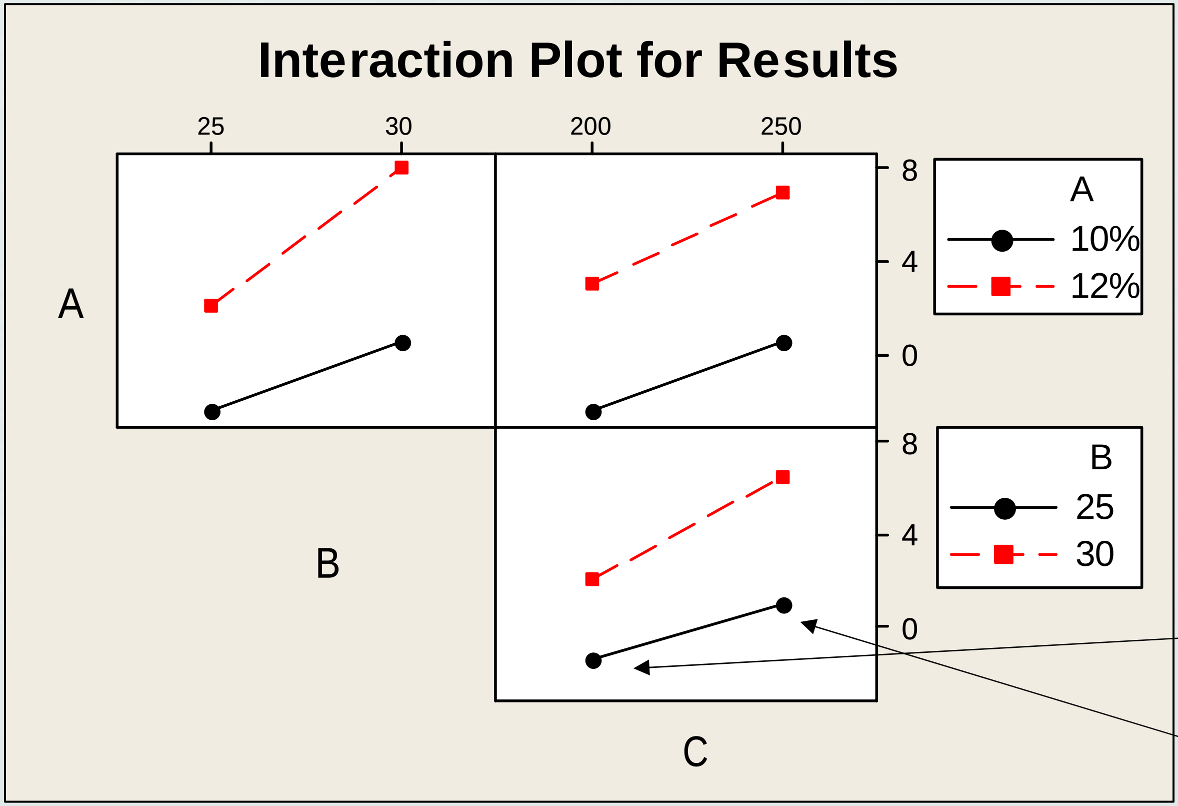


	AB Avg.
A^-B^-	-2.5
A^+B^-	2.0
A^-B^+	0.5
A^+B^+	8.0

This plot shows that when the data is restricted to A^+,the B effect is more steep [the AB average changes from 2 to 8] than when we restrict our data to A^-, [the AB average changes from -2.5 to .5].

Interaction Effects

The following plot shows the interaction (or 2-way effects) of all three factors. When the lines are parallel, interaction effects are 0. The more different the slopes, the more influence the interaction effect has on the results. To visualize these effects, the Y axis is always the same for each combination of factors. This graph shows that the AB interaction effect is the largest.



This plot shows that the B^-C^- average (i.e. B set to 25 and C set to 200) is -1.5.

The B^-C^+ average is 1.

Interaction Effects

To calculate the size of each two-way interaction effect, calculate the average of every level of each factor combination as well as all other effects that impact those averages.

The A effect, B effect, and overall effect (grand mean) influence the AB interaction effect. Factor C is completely ignored in these calculations. Note that these values are placed in rows corresponding to the original dataset.

A	B	C	Results
10%	25	**200**	-4
12%	25	**200**	1
10%	30	**200**	-1
12%	30	**200**	5
10%	25	250	-1
12%	25	250	3
10%	30	250	2
12%	30	250	11

AB Avg.	A Effect	B Effect	Grand Avg.
-2.5	-3	-2.25	*2*
2.0	3	-2.25	*2*
0.5	-3	2.25	*2*
8.0	3	2.25	*2*
-2.5	-3	-2.25	2
2.0	3	-2.25	2
0.5	-3	2.25	2
8.0	3	2.25	2

These two rows show the AB average, the A effect, the B effect, and the grand mean when A^- and B^-.

These two rows show the AB average, the A effect, the B effect, and the grand mean when A^- and B^+.

Interaction Effects

Effect size is the difference between the average and the partial fit.

Partial fit = the effect of all the influencing factors.

For main effects, the partial fit is the grand mean.

Effect of AB = AB Avg.– [effect of A + effect of B + the grand mean]

Effect for A^-B^- = -2.5 – [-3 + -2.25 +2] = .75

Effect for A^-B^+ = 0.5 – [-3 + 2.25 +2] = -.75

A	B	C	Results
10%	25	**200**	-4
12%	25	**200**	1
10%	30	**200**	-1
12%	30	**200**	5
10%	25	250	-1
12%	25	250	3
10%	30	250	2
12%	30	250	11

AB Avg.	A Effect	B Effect	Grand Avg.
-2.5	-3	-2.25	*2*
2.0	3	-2.25	*2*
0.5	-3	2.25	*2*
8.0	3	2.25	*2*
-2.5	-3	-2.25	2
2.0	3	-2.25	2
0.5	-3	2.25	2
8.0	3	2.25	2

subtract the partial fit from each level average

AB Effect
0.75
-0.75
-0.75
0.75
0.75
-0.75
-0.75
0.75

Interaction Effects

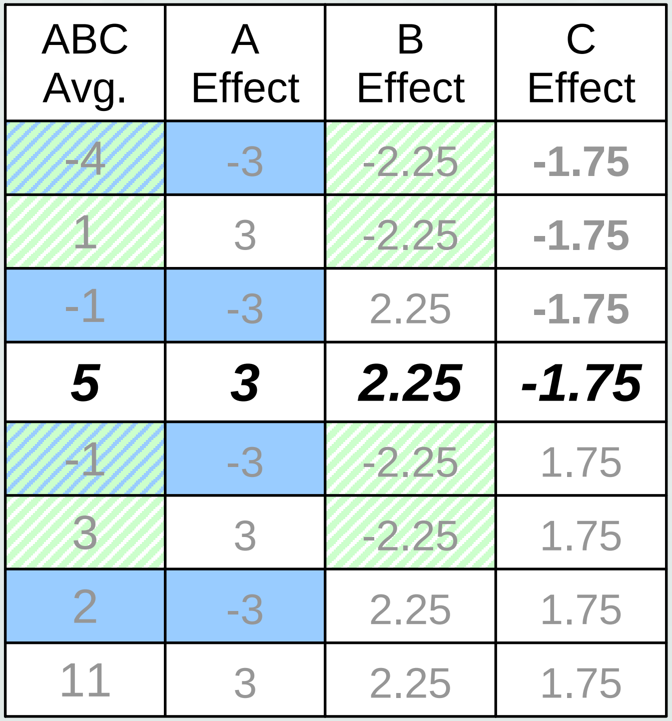
The effect size for 3-way interactions are calculated by finding the appropriate average and subtracting the partial fit.

To calculate the ABC effect when A and B are high and C is low ($A^+B^+C^-$)

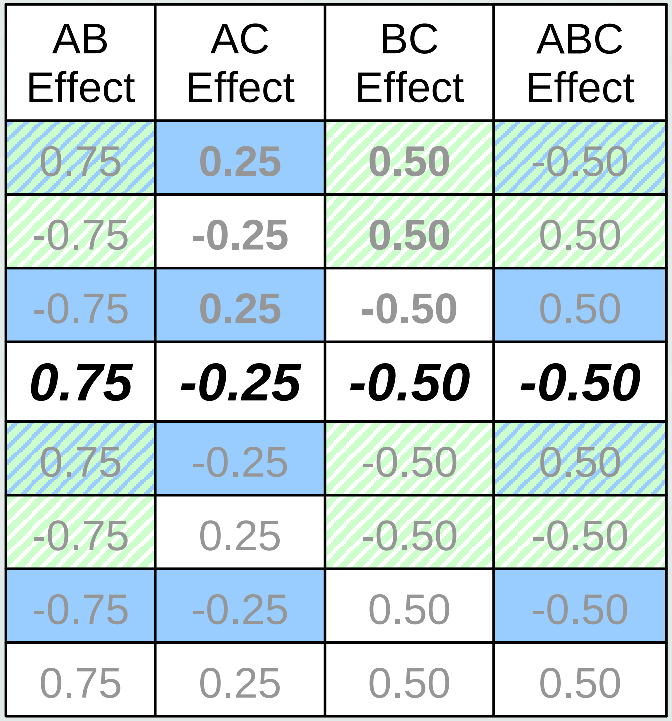
$A^+B^+C^-$ average – [A^+ effect + B^+ effect + C^- effect + A^+B^+ effect + A^+C^- effect + B^+C^- effect + grand mean]

= 5 – [+3 +2.25 -1.75 +.75 -.25 -.5 +2]

= -.5



ABC Avg.	A Effect	B Effect	C Effect
-4	-3	-2.25	**-1.75**
1	3	-2.25	**-1.75**
-1	-3	2.25	**-1.75**
5	***3***	***2.25***	***-1.75***
-1	-3	-2.25	1.75
3	3	-2.25	1.75
2	-3	2.25	1.75
11	3	2.25	1.75



AB Effect	AC Effect	BC Effect	ABC Effect
0.75	**0.25**	**0.50**	-0.50
-0.75	**-0.25**	**0.50**	0.50
-0.75	**0.25**	**-0.50**	0.50
0.75	***-0.25***	***-0.50***	***-0.50***
0.75	-0.25	-0.50	0.50
-0.75	0.25	-0.50	-0.50
-0.75	-0.25	0.50	-0.50
0.75	0.25	0.50	0.50

Interaction Effects

On your own, calculate all the AC, BC and ABC effects and verify your work in the following table.

Notice that each effect column sums to zero. This will always be true whenever calculating effects. This is not surprising since effects measure the unit deviation from the observed value and the mean.

ABC Avg.	A Effect	B Effect	C Effect
-4	-3	-2.25	**-1.75**
1	3	-2.25	**-1.75**
-1	-3	2.25	**-1.75**
5	3	2.25	**-1.75**
-1	-3	-2.25	1.75
3	3	-2.25	1.75
2	-3	2.25	1.75
11	3	2.25	1.75

AB Effect	AC Effect	BC Effect	ABC Effect
0.75	**0.25**	**0.50**	-0.50
-0.75	**-0.25**	**0.50**	0.50
-0.75	**0.25**	**-0.50**	0.50
0.75	**-0.25**	**-0.50**	-0.50
0.75	-0.25	-0.50	0.50
-0.75	0.25	-0.50	-0.50
-0.75	-0.25	0.50	-0.50
0.75	0.25	0.50	0.50

Interaction Effects

Also note that the ABC Average column is identical to the results column. In this example, there are 8 runs (observations) and 8 ABC interaction levels. There are not enough runs to distinguish the ABC interaction effect from the basic sample to sample variability. In factorial designs, each run needs to be repeated more than once for the highest-order interaction effect to be measured. However, this is not necessarily a problem because it is often reasonable to assume higher-order interactions are not significant.

A	B	C	Results
10%	25	**200**	**- 4**
12%	25	**200**	**1**
10%	30	**200**	**-1**
12%	30	**200**	**5**
10%	25	250	-1
12%	25	250	3
10%	30	250	2
12%	30	250	11

AB Avg.	AC Avg.	BC Avg.	ABC Avg.
-2.5	**-2.5**	**-1.5**	**- 4**
2.0	**3.0**	**-1.5**	**1**
0.5	**-2.5**	**2.0**	**-1**
8.0	**3.0**	**2.0**	**5**
-2.5	0.5	1.0	-1
2.0	7.0	1.0	3
0.5	0.5	6.5	2
8.0	7.0	6.5	11

ABC Effect
-0.50
0.50
0.50
-0.50
0.50
-0.50
-0.50
0.50

Mathematical Calculations

Effect plots help visualize the impact of each factor combination and identify which factors are most influential. However, a statistical hypotheses test is needed in order to determine if any of these effects are significant. Analysis of variance (ANOVA) consists of simultaneous hypothesis tests to determine if any of the effects are significant.

Note that saying “factor effects are zero” is equivalent to saying “the means for all levels of a factor are equal”. Thus, for each factor combination ANOVA tests the null hypothesis that the population means of each level are equal, versus them not all being equal.

Several calculations will be made for each main factor and interaction term:

Sum of Squares (SS) = sum of all the squared effects for each factor

Degrees of Freedom (df) = number of free units of information

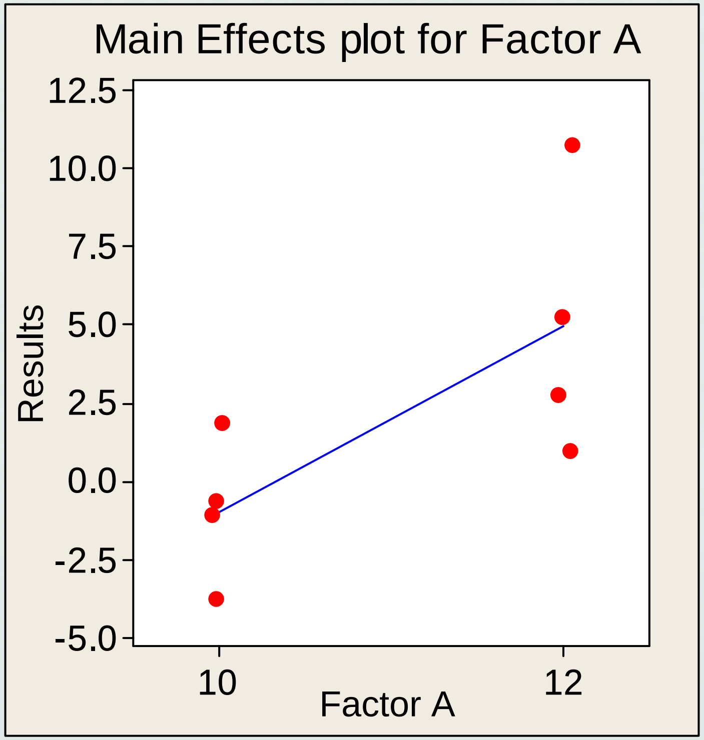
Mean Square (MS) = SS/df for each factor

Mean Square Error (MSE) = pooled variance of samples within each level

F-statistic = MS for each factor/MSE

Mathematical Calculations

The following main effects plot includes the actual data points. This plot illustrates both the between level variation and the within level variation. **Between-level variation** measures the spread of the level means (from -1 at the low level to 5 at the high level). The calculation for this variability is **Mean Square for factor A (MS_A).** **Withinlevel variation** measures the spread of points within each level. The calculation for this variability is **Mean Square Error (MSE).**

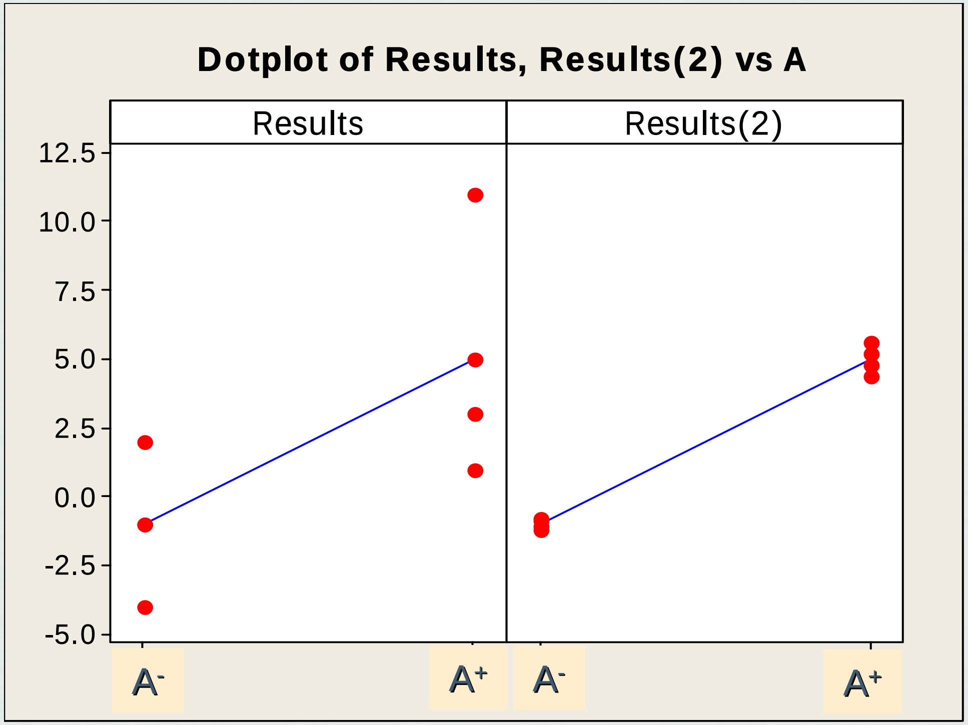


To determine if the difference between level means of factor A is significant, we compare the between-level variation of A (MS_A) to the within-level variation (MSE).

If the MS_A is much larger than MSE, it is reasonable to conclude that their truly is a difference between level means and the difference we observed in our sample runs was not simply due to random chance.

Mathematical Calculations

The first dotplot of the Results vs. factor A shows that the between level variation of A (from -1 to 5) *is not* significantly larger than the within level variation (the variation within the 4 points in A^- and the 4 points in A^+).



The second dotplot of Results(2) vs. factor A uses a hypothetical data set. The between-level variation is the same in both Results and Results(2). However the within-level variation is much larger for Results than Results(2). With Results(2) we are much more confident hat the effect of Factor A is not simply due to random chance.

Even though the averages are the same (and thus the MS_A are identical) in both data sets, Results(2) provides much stronger evidence that we can reject the null hypothesis and conclude that the effect of A^- is different than the effect of A^+ .

Mathematical Calculations

Sum of Squares (SS) is calculated by summing the squared factor effect for each run. The table below shows the calculations for the SS for factor A (written SS_A) = 72.0 = $4(-3)^2 + 4(3)^2$

The table below also shows SS_B = 40.5 and SS_C = 24.5

	A Effect	B Effect	C Effect
	-3	-2.25	***-1.75***
	3	-2.25	***-1.75***
	-3	2.25	***-1.75***
	3	2.25	***-1.75***
	-3	-2.25	1.75
	3	-2.25	1.75
	-3	2.25	1.75
	3	2.25	1.75
Sum	0	0	0

	A Effect Squared	B Effect Squared	C Effect Squared
	9	5.0625	**3.0625**
	9	5.0625	**3.0625**
	9	5.0625	**3.0625**
	9	5.0625	**3.0625**
	9	5.0625	3.0625
	9	5.0625	3.0625
	9	5.0625	3.0625
	9	5.0625	3.0625
SS	72.0	40.5	24.5

Mathematical Calculations

In the previous table, the effect of **A⁻** = -3 and effect of **A⁺** = 3 were calculated by subtracting the grand mean from the level averages.

The formula for calculating effects **A⁻** is: $(\bar{y}_1.. - \bar{\bar{y}})$ and **A⁺** is $(\bar{y}_2.. - \bar{\bar{y}})$

$\bar{y}_i..$ is the factor A average for level i.
In our example, i = 1 represents **A⁻** and $\bar{y}_1.. = -1$ also $\bar{y}_2.. = 5$

$\bar{\bar{y}}$ is the grand mean. In our example, $\bar{\bar{y}}$ = 2

To calculate SS_A , the effect is squared for each run and then summed. Note that there are n_1 = 4 runs for **A⁻** and the n_2 = 4 runs for **A⁺** .

$$4(\bar{y}_1.. - \bar{\bar{y}})^2 + 4(\bar{y}_2.. - \bar{\bar{y}})^2 = \quad 4(-1-2)^2 + 4(5-2)^2 = 72$$

Mathematical Calculations

The generalized formula for SS_A is:

$$SS_A = \sum_{i=1}^{I} n_i(\bar{y}_{i}.. - \bar{\bar{y}})^2 = n_1(\bar{y}_{1}.. - \bar{\bar{y}})^2 + n_2(\bar{y}_{2}.. - \bar{\bar{y}})^2$$

$$= 4(-1-2)^2 \; + 5(5-2)^2 \quad = \quad 72$$

Where:

I is the number of levels in factor A, in our example, I = 2

n_i is the number of samples in the i^{th} level of factor A, $n_1 = 4$ and $n_2 = 4$

$\bar{y}_{i}..$ is the factor A average for level i and $\bar{\bar{y}}$ is the grand mean

In the same manner, SS_B and SS_C are calculated by

$$SS_B = \sum_{j=1}^{J} n_j(\bar{y}._{j}. - \bar{\bar{y}})^2 \qquad SS_C = \sum_{k=1}^{K} n_k(\bar{y}.._{k} - \bar{\bar{y}})^2$$

Where:

J is the number of levels in factor B, K is the number of levels for C

n_j is the number of samples in the j^{th} level of factor B

n_k is the number of samples in the k^{th} level of factor B

$\bar{y}._{j}.$ is the factor B average for level j, $\bar{y}.._{k}$ is the factor C average for level k

Mathematical Calculations

Sum of Squares (SS) for interactions is also calculated by summing the squared factor effect for each run. The table below shows the calculations for $SS_{AB} = 4.5 = 2(.75)^2 + 2(-.75)^2 + 2(-.75)^2 + 2(.75)^2$

$$SS_{AB} = \sum_{j=1}^{J} \sum_{i=1}^{I} n_{ij} (ij^{th} \quad level \quad effect)^2 = \sum_{j=1}^{J} \sum_{i=1}^{I} n_{ij} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{\bar{y}})^2$$

$$= \sum_{j=1}^{J} \left(n_{1j} (\bar{y}_{1j\cdot} - \bar{y}_{1\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{\bar{y}})^2 + n_{2j} (\bar{y}_{2j\cdot} - \bar{y}_{2\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{\bar{y}})^2 \right) \qquad = 4.5$$

AB Effect	AC Effect	BC Effect
0.75	**0.25**	**0.50**
-0.75	**-0.25**	**0.50**
-0.75	**0.25**	**-0.50**
0.75	**-0.25**	**-0.50**
0.75	-0.25	-0.50
-0.75	0.25	-0.50
-0.75	-0.25	0.50
0.75	0.25	0.50

	AB Effect Squared	AC Effect Squared	BC Effect Squared
	.5625	.0625	.25
	.5625	.0625	.25
	.5625	.0625	.25
	.5625	.0625	.25
	.5625	.0625	.25
	.5625	.0625	.25
	.5625	.0625	.25
	.5625	.0625	.25
SS	4.5	0.5	2

n_{ij} is the number of samples in level ij. In our example each ij level has 2 samples.

$\bar{y}_{ij\cdot}$ is the mean of all AB factor runs at the i, j level.

On your own, calculate the SS_{AC} and SS_{BC}

Mathematical Calculations

Degrees of Freedom (df) = number of free units of information. In the example provided, there are 2 levels of factor A. Since we require that the effects sum to 0, knowing A^- automatically forces a known A^+ . If there are I levels for factor A, one level is fixed if we know the other I-1 levels. Thus, when there are I levels for a main factor of interest, there is I-1 free pieces of information.

A Effect	B Effect	C Effect
-3	-2.25	***-1.75***
3	-2.25	***-1.75***
-3	2.25	***-1.75***
3	2.25	***-1.75***
-3	-2.25	1.75
3	-2.25	1.75
-3	2.25	1.75
3	2.25	1.75

For a full factorial ANOVA, df for a main effect are the number of levels minus one:

$df_A = I - 1$

$df_B = J - 1$

$df_C = K - 1$

Mathematical Calculations

For the AB interaction term there are I*J effects that are calculated. Each effect is a piece of information. Restrictions in ANOVA require:

1) AB factor effects sum to 0. This requires 1 piece of information to be fixed.
2) The AB effects within A^- sum to 0. In our example, the AB effects restricted to A^- are (.75, -.75,.75,-.75). The same is true for the AB effect restricted to A^+. This requires 1 piece of information to be fixed in each level of A. Since 1 value is already fixed in restriction1), this requires I-1 pieces of information.
3) The AB effects within each level of B. This requires J-1 pieces of information.

A Effect	B Effect	AB Effect
-3	-2.25	0.75
3	-2.25	-0.75
-3	2.25	-0.75
3	2.25	0.75
-3	-2.25	0.75
3	-2.25	-0.75
-3	2.25	-0.75
3	2.25	0.75

Thus, general rules for a factorial ANOVA:

$$df_{AB} = IJ - [(I-1) + (J-1) + 1] = (I-1)(J-1)$$

$$df_{BC} = (J-1)(K-1)$$

$$df_{AC} = (I-1)(K-1)$$

$$df_{ABC} = (I-1)(J-1)(K-1)$$

Note the relationship between the calculation of df_{AB} and the calculation of the AB interaction effect.

$$df_{AB} = \#\text{ of effects} - [\text{pieces of information already accounted for}]$$

$$= \#\text{ of effects} - [df_A + df_B + 1]$$

Mathematical Calculations

Mean Squares (MS) = SS/df for each factor. MS is a measure of variability for each factor. MS_A is a measure of the spread of the Factor A level means. This is sometimes called **between** level variability.

$$MS_A = \frac{SS_A}{df_A} = \frac{\sum_{i=1}^{I} n_i (\bar{y}_{i\cdot\cdot} - \bar{\bar{y}})^2}{I-1}$$

Notice how much the MS_A equation looks like the overall variance equation:

$$Overall \quad Variance = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N-1}$$

N = overall number of samples. This is the variance formula taught in all introductory statistics courses.

Mean Square Error (MSE) = SS_E/df_E

MSE is also a measure of variability, however MSE measures the pooled variability **within** each level. While many texts give specific formulas for Sum of Squares Error (SS_E) and degrees of freedom Error (df_E), they can be most easily calculated by subtracting all other SS from the Total SS =(N-1)(Overall variance).

Mathematical Calculations

F-statistic = MS for each factor/MSE. The F-statistic is a ratio of the between variability over the within variability.

If the true population mean of **A^-** equals true population mean of **A^+**, then we would expect the variation between levels in our sample runs to be equivalent to the variation within levels. Thus we would expect the F-statistic would be close to 1.

If the F-statistic is large, it seems unlikely that the population means of each level of factor A are truly equal.

Mathematical theory proves that if the appropriate assumptions hold, the F-statistic follows an F distribution with df_A (if testing factor A) and df_E degrees of freedom.

The **p-value** is looked up in an F table and gives the likelihood of observing an F statistic at least this extreme (at least this large) assuming that the true population factor has equal level means. Thus, when the p-value is small (i.e. less than 0.05 or 0.1) the effect size of that factor is statistically significant.

Mathematical Calculations

These calculations are summarized in an ANOVA table. Each row in the ANOVA table tests the null hypothesis that the population means of each factor level are equal, versus them not all being equal.

Source	DF	SS	MS	F
A	I-1	$\sum_{i=1}^{I} n_i(\bar{y}_{i\cdot\cdot} - \bar{\bar{y}})^2$	SSA/dfA	MSA/MSE
B	J-1	$\sum_{j=1}^{J} n_j(\bar{y}_{\cdot j\cdot} - \bar{\bar{y}})^2$	SSB/dfB	MSB/MSE
C	K-1	$\sum_{k=1}^{K} n_k(\bar{y}_{\cdot\cdot k} - \bar{\bar{y}})^2$	SSC/dfC	MSC/MSE
AB	(I-1)(J-1)	$\sum_{j=1}^{J} \sum_{i=1}^{I} n_{ij}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{\bar{y}})^2$	SSAB/dfAB	MSAB/MSE
AC	(I-1)(K-1)	$\sum_{k=1}^{K} \sum_{i=1}^{I} n_{ik}(\bar{y}_{i\cdot k} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot k} - \bar{\bar{y}})^2$	SSAC/dfAC	MSAC/MSE
BC	(J-1)(K-1)	$\sum_{k=1}^{K} \sum_{j=1}^{J} n_{jk}(\bar{y}_{\cdot jk} - \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot k} - \bar{\bar{y}})^2$	SSBC/dfBC	MSBC/MSE
Error	subtraction	subtraction	SSE/dfE	
Total	N-1	(N-1)(Overall Variance)		

Even though the SS calculations look complex, remember they can always be found by simply summing the column of squared effect sizes.

Mathematical Calculations

For the bottle filling example, we calculate the following results.

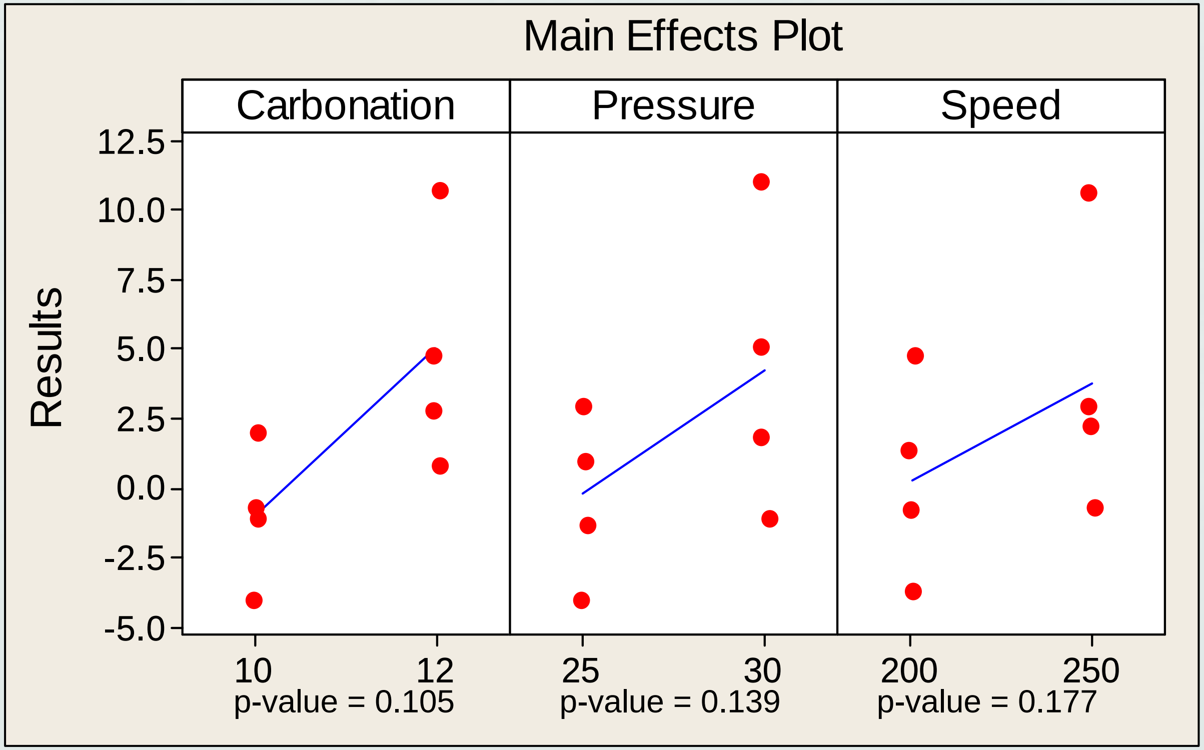
Source	DF	SS	MS	F	p-value
A	1	72.0	72.0	36.00	0.105
B	1	40.5	40.5	20.25	0.139
C	1	24.5	24.5	12.25	0.177
A*B	1	4.5	4.5	2.25	0.374
A*C	1	0.5	0.5	0.25	0.705
B*C	1	2.0	2.0	1.00	0.500
Error	1	2.0	2.0		
Total	7	146.0			

Each row in the ANOVA table represents a null hypothesis that the means of each factor level are equal. Each row shows an F statistic and a p-value corresponding to each hypothesis. When the p-value is small (i.e. less than 0.05 or 0.1) reject the null hypothesis and conclude that the levels of the corresponding factor are significantly different (i.e. conclude that the effect sizes of that factor are significantly large).

Mathematical Calculations

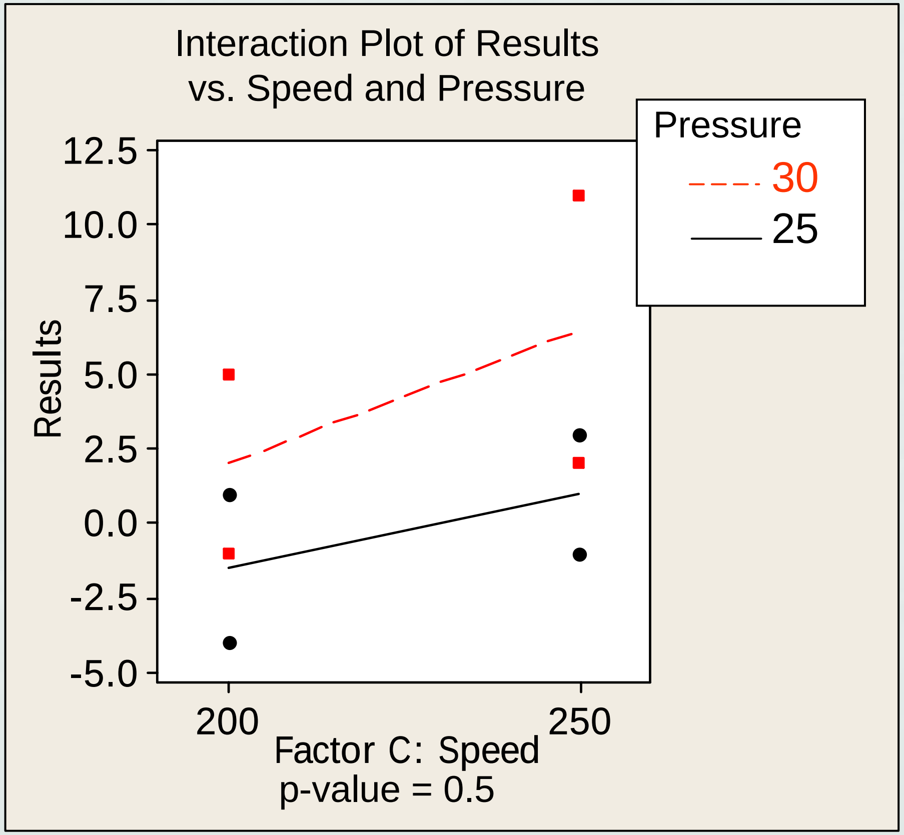
Viewing the effect plots with the appropriate p-values clearly shows that while factor A (Carbonation) had the largest effect sizes in our sample of 8 runs, effect sizes this large would occur in 10.5% of our samples even if factor A truly has no effect on the Results.

Effect sizes as large as were observed for factor C (Speed) would occur in 17.7% of samples of 8 runs even if there truly was no difference between the mean Results of speed run at 200 bpm and at 250 bpm.



Mathematical Calculations

The BC interaction plot with the appropriate p-value shows that the lack of parallelism between the lines is relatively small based on the sampling variability. The p-value also shows that we would expect a lack of parallelism at least this large in 50% of our samples even if no interaction existed between factor B (pressure) and factor C (speed).



Fisher Assumptions

In order for the p-values to be accurate, the F statistics that are calculated in ANOVA are expected to follow the F distribution. While we will not discuss the derivation of the F distribution, it is valuable to understand the six Fisher assumptions that are used in the derivation. If any experimental data does not follow these assumptions, then ANOVA give incorrect p-values.

1) The unknown true population means (and effect sizes) of every treatment are constant.

2) The additivity assumption: each observed sample consists of a true population mean for a particular level combination *plus* sampling error.

3) Sampling errors are normally distributed and 4) Sampling errors are independent. Several **residual plots** should be made to validate these assumptions every time ANOVA is used.

5) Every level combination has equivalent variability among its samples. ANOVA may not be reliable if the **standard deviation** within any level is more than twice as large as the standard deviation of any other level.

6) Sampling errors have a mean of 0, thus the average of the samples within a particular level should be close to the true level mean.

Advanced Designs: Fixed Vs. Random Effects

Fixed factors: the levels tested represent all levels of interest

Random factors: the levels tested represent a random sample of an entire set of possible levels of interest.

EXAMPLE 3: A statistics class wanted to test if the speed at which a game is played (factor A: slow, medium, or fast speed) effects memory. They created an on-line game and measured results which were the number of sequences that could be remembered.

If four friends wanted to test who had the best memory. They each play all 3 speed levels in random orders. A total of 12 games were played. Since each student effect represents a specific level that is of interest, student should be considered a fixed effect.

If four students were randomly selected from the class and each student played each of the three speed levels. A total of 12 games were played. How one student compared to another is of no real interest. The effect of any particular student has no meaning, but the student-to-student variability should be modeled in the ANOVA. Student should be considered a random effect.

Advanced Designs
Crossed Vs. Nested Effects

Factors A and B are **crossed** if every level of A can occur in every level of B. Factor B is **nested** in factor A if levels of B only have meaning within specific levels of A.

EXAMPLE 3 (continued):

If 12 students from the class were assigned to one of the three speed levels (4 within each speed level), students would be considered nested within speed. The effect of any student has no meaning unless you also consider which speed they were assigned. There are 12 games played and MSE would measure student to student variability. Since students were randomly assigned to specific speeds, the student speed interaction has no meaning in this experiment.

If four friends wanted to test who had the best memory they could each play all 3 speed levels. There would be a total of 12 games played. Speed would be factor A in the ANOVA with 2 df. Students would be factor B with 3 df. Since the student effect and the speed effect are both of interest these factors would be crossed. In addition the AB interaction would be of interest.