Two-Way ANOVA

Example: A student wanted to know if accuracy depends on the distance from the dart board and which hand darts are thrown with so she ran a small designed experiment. The variable "***Accuracy***" is the number of inches the dart is from the bull's-eye, so large values have less accuracy. The data is on StatCrunch, called "***Darts***".

To start, she would be interested in if there is a "***Hand***" effect (should be) and if there is a "***Distance***" effect (should be).

Make side-by-side boxplots for each treatment: (no need to draw these in the notes)

Statcrunch (there does look to be effects)

The ANOVA model would look like this:

General: $y_{ijk} = \mu + \tau_j + \gamma_k + \varepsilon_{ijk}$

Better: $y_{ijk} = \mu + \text{Hand Effect}_j + \text{Distance Effect}_k + \text{Error}_{ijk}$

The null hypotheses are that the treatment effects are zero.

There are two sets of hypotheses:

$H_0: \tau_1 = \tau_2$

H_A: (Hands Differ)

$H_0: \gamma_1 = \gamma_2 = \gamma_3$

H_A: (Distances Differ)

When we run our two-way ANOVA, the error term holds the variability that is Left Over after the effects of both variables have been removed.

Two-Way ANOVA actually improves the experiment and its Analysis

Example: Run the two-way ANOVA on StatCrunch and make the basic conclusions.

For Now, check ☑ "Fit Additive Model"

To test for a hand effect,

$df = 2 - 1 = 1$

F-Stat = 44.42

P-Value < 0.0001

Reject Ho, Conclude mean accuracy differs by hand.

To test for a Distance effect,

$df = 3 - 1 = 2$, F-Stat = 28.56

P-Value < 0.0001, Reject Ho,

Conclude there is a difference in mean accuracy by distance.

Give an estimate of the standard deviation we'd expect if we made repeated throws at any treatment condition:

$\sqrt{MSE} \approx$ SD at any treatment

$= \sqrt{0.8936}$

≈ 0.9453 inches

Assumptions and Conditions

- Make side-by-side boxplots and look for outliers and check that each group has relatively equal spread.
- If outliers can be corrected, do it. If they are true values, consider removing them for the analysis and note that you did so.
- ANOVA fails on the "safe side" because outliers will increase MSE, making it harder to reject the null for any treatment. This means we will tend to make more type II errors, which is generally considered safer. As a result, removing outliers from the analysis isn't a horrible idea.
- Additivity is an assumption that we can just add the effects of two treatments together. Recall our model:

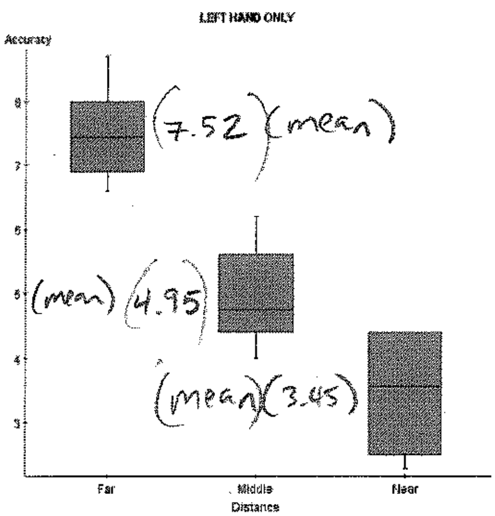
$$y_{ijk} = \mu + \tau_j + \gamma_k + \varepsilon_{ijk}$$

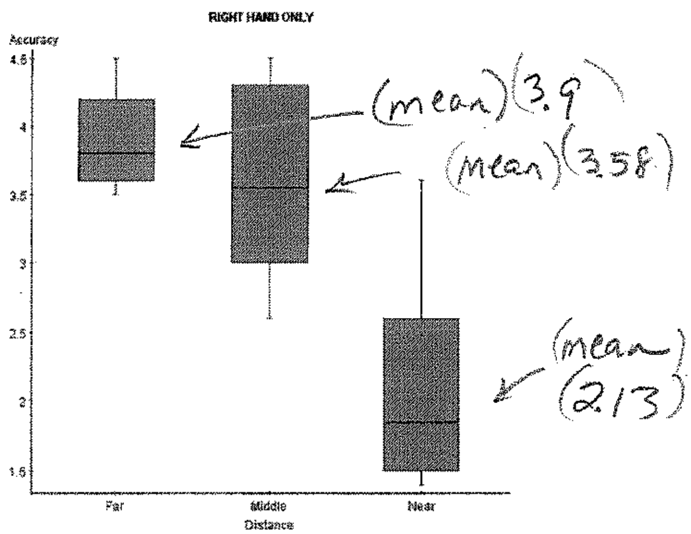
Additivity means effects of one treatment don't influence the other treatment

We can check if the additivity assumption is a good one (it might not be).

- For our darts example, we can check additivity as follows using this idea: Changing hands must make the same difference in "***Accuracy***" no matter what distance you throw from.

Make side-by-side boxplots for "Accuracy" by "Distance" where hand is left and then right.

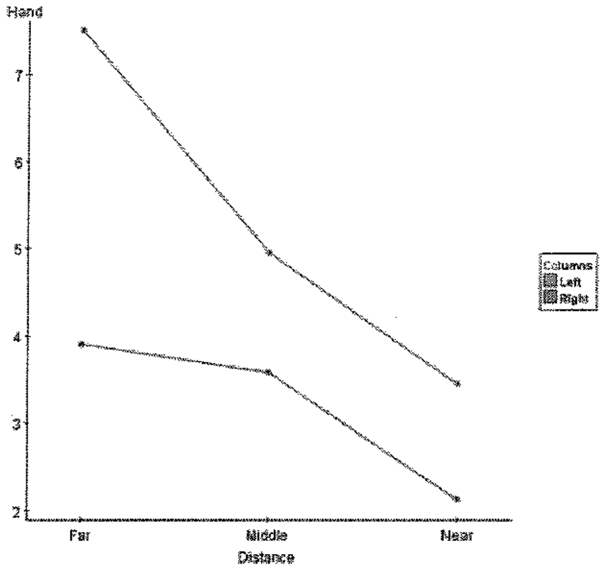


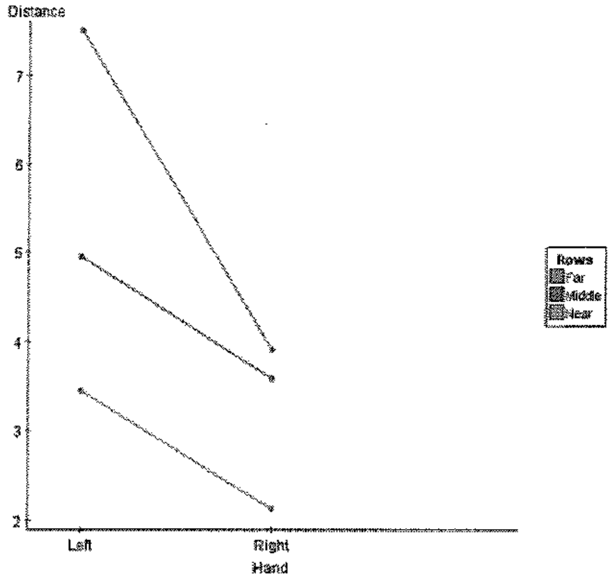


Distance affects the left hand more than the right!

Model might not ~~be~~ additive.

- The changes in accuracy due to distance for the left hand look larger, so we have what's called an interaction effect.
- Check interaction plots (checkbox on the Two-Way ANOVA menus).





- Interaction plots graph the mean of the observations at each level of one factor broken down by the levels of the other factor.
- In the left plot, we have the average accuracy plotted by distance for each hand. Since the lines are not Parallel, it is giving us some indication that the additivity assumption is not met.
- In the right plot, we have the average accuracy plotted by hand for each distance. Again, the lines are not parallel, indicating the model is not additive.
- We can add an interaction term to the model if we believe the effects of one factor change for different levels of another factor.
- On StatCrunch, we uncheck the "Fit additive model" checkbox on the Two-Way ANOVA menus.
- For the darts example, the interaction effect means:

Distance affects left and right hands differently.

- To check the equal variance assumption, we should plot the model's residuals against the model's predicted values. Currently, StatCrunch will not do this. If you see (could see) the plot thickening, re-express your data values. (take logs, square roots, reciprocals, etc...)
- Finally, the underlying error terms must follow a normal model. You would make a QQ plot or a histogram. Currently, StatCrunch will not do this either.

Testing for an Interaction

- We should test for an interaction of our treatments. We would add a new term to the model, so now it looks like this:

$y_{ijk} = \mu + \tau_j + \gamma_k + \omega_{jk} + \varepsilon_{ijk}$ where ω_{jk} is the interaction of level j of factor 1 with level k of factor 2.

- We will see a new line on the ANOVA table. The interaction term for the darts example is:

Interaction = Distance × Hand with P-Value of 0.0012
df = 2 df = 1

- The degrees of freedom for the interaction term is:

df Dist × df Hand = $2 \times 1 = 2$

- The degrees of freedom are deducted from the Error error term.
- Important: When a significant interaction effect is present, it's difficult to interpret the main effects of each factor.
- **Example:** How much does distance matter? It depends on which hand you use.
- **Example:** How much does switching from left to right hand matter? It depends on your distance.
- If a significant interaction is present, always include the interaction plots as part of the analysis. It is the best way to display what the data is telling us.

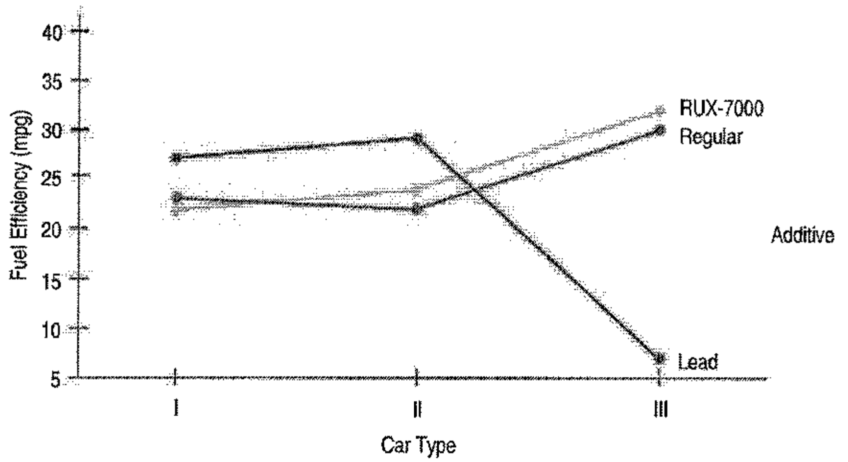
A Few Comments on Interaction Effects

- There is always a chance that your two factors will interact. Check with an interaction plot and use the results of the F-test to determine if the model should include this term.
- If the interaction plot looks parallel and the F-test has a high P-value, remove the interaction term and proceed with the two-way ANOVA without the interaction.
- If the interaction term is significant and the lines on your interaction plot do not cross, interpret your results as follows:

First, interpret the interaction term in the context of the problem. Second, if the lines in the plot do not cross, determine if the main effects individually can be interpretted.

- If the interaction term is significant and the lines on the interaction plot cross, you need to be extra careful in interpreting the meaning of your analysis. It's possible the main effects are completely meaningless.

Example: Three fuel additives were added to three types of vehicles to see any improvements to fuel efficiency. Interpret the interaction plot.



Fuel with lead is great for two vehicles but a disaster for the third.

But, on average lead lowers mpg. The main effect is significant but meaningless.

Example: At supermarkets, experiments are typically designed to determine effective sales strategies. At one supermarket, the two factors were "Price" level (regular, reduced, or cost) and "Display" level (normal display space, normal plus an end cap, and twice the normal space).

Each treatment was applied three times to a particular product, and the response variable was unit sales for the week. Fire up the "Supermarket Marketing" dataset.

a. Make side-by-side boxplots for each factor.
b. Run a two-way ANOVA with an interaction term and make the interaction plots. Run all the appropriate tests and make all the appropriate conclusions.

a.) By "Display", normal plus looks to have ~~best~~ effect.
Variance might not be equal!!
No outliers.
By Price, clearly cost is best, reduced in 2nd, regular in third.
Variance not equal again....

b.) 2-WAY ANOVA.

1. Interaction term is significant
$F = 258.07$, P-Value < 0.0001.
Interaction plot → not perfectly parallel.

- At regular prices, normal plus or double displays don't make a big impact on Sales.

For reduced pricing, Normal plus is the best display.

Also, we can see that sales go up nomatter what display is used, as long as we keep reducing prices.

By itself, pricing is significant

$\overline{Y}_{COST} = 1972.22$

$\overline{Y}_{REDUCED} = 1535.44$

$\overline{Y}_{REGULAR} = 1144.11$

By itself, Display is significant

$\overline{Y}_{NORMAL + ENDCAP} = 1874.56$

$\overline{Y}_{TWICE} = 1512.11$

$\overline{Y}_{NORMAL} = 1265.11$

Best Combo.

$\overline{Y}_{COST,\ ENCAP}$

$= 2510.$

Multiple Regression

Example: Fire up the "**Chicken Sandwiches**" dataset on StatCrunch. In Math 127, we might have used "***Fat***" grams to predict "***Calories***".

The simple linear model is: $\widehat{\text{Calories}} = 210.77 + 12.37(\text{fat})$

The scatterplot of x vs. y looks: Linear & Positive & Strong

The residual plot of x vs. residuals looks: Random & Equal Spread

R-sq = 82.5%, which seems "pretty good", but can we do better?

There are many other variables in the dataset that could be used to help in our prediction of how many calories a sandwich might have. Let's start by introducing "***Serving Size***" to the model as well.

- Multiple regression means multiple explanatory (x) variables for a single response variable (y).
- We still use the principle of least squares to determine the best fitting model, but the calculations are so complicated and tiresome that we rely solely on technology (a theme in Math 128, to be sure).
- R^2 will give the fraction of variability in our y variable that is accounted for by this multiple regression model. Higher is better, but we want a simple model. The more explanatory variables we add, the higher R^2, but keep things simple and interpretable.
- The multiple regression model looks like this:

$$\widehat{\text{Calories}} = 52.04 + 0.99(\text{Serving Size } g) + 9.65(\text{Fat } g)$$

$$R^2 = 93.36\%$$

- The standard deviation of the residuals for the model is s = $\sqrt{MSE} = 39.91$ on StatCrunch. This number gives us an indication of how precisely we can predict calories "***Calories***" based on "***Fat***" and "***Serving Size***".

- Interpreting the multiple regression coefficients takes a bit of savvy and dose of critical thinking:

The coefficient on "***Serving Size***" is 0.99, and it was significant with a *t*-ratio of 8.37 and a *P*-value <0.0001.

- This means that for a regression model already with "***Fat***" included, "***Serving Size***" provides additional predictive value.

- How can we put this into words? "*For sandwiches with a certain fat content, say 15 grams…*

each additional gram of sandwich adds 0.99 calories, on average.

The coefficient on "***Fat***" is 9.65, and it was significant with a *t*-ratio of 15.22 and a *P*-value <0.0001.

- This means that for a regression model already with "***Serving Size***" included, "***Fat***" provides additional predictive value.

- How can we put this into words? "*For sandwiches that are, say 200 grams…*

each additional fat gram adds 9.65 calories, on average

- Multiple regression coefficients on your x-variables measure the average conditional relationship

The Multiple Regression Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

* There could be more than two explanatory variables

Assumptions and Conditions (These Go In Order)

1. Check scatterplots for each *x* variable against the *y* variable. Every one of them must look linear. No curves / patterns.

   Do this for "**Chicken Sandwiches**". Weak linear is OK

   Serving Size v Calories: linear

   Fat v Calories: linear

2. Run the multiple regression on StatCrunch and checkbox "Save Residuals" and "Save Predicted values." Make a scatterplot (*x* = predicted) with these two added columns and hope to see no pattern and hope to see equal spread.

   Do this for "**Chicken Sandwiches**".

   No Pattern & Equal Spread ✓✓

   One glaring outlier → Hardies "Low Carb Charbroiled Club" 132.45 calories less than predicted.

   - Consider removing and seeing effect.

3. Think about how the data were collected – was randomization suitable? Are the data an unbiased subset of a larger population? Any outliers that can be justifiably removed?

If the data were collected in time-series order, plot the residuals against time to check for a pattern that might suggest the data values are not independent of each other (no need for chicken sandwiches).

4*. If any conditions are not met, the main suggestion is to re-express your data values to correct the problems. You can try taking logs, square roots, squares, reciprocals, or some other mathematical function of your data. Just remember Parsimony. Simpler is better.

5. With all conditions met, test the main hypothesis:

$H_0: \beta_1 = \beta_2 = \beta_K = 0$

H_A: At least one coefficient differs from 0.

F test, see Statcrunch output

In words, this means "Is my regression model better than just using the Mean of y to predict for any x value?"

Test the Main Hypothesis for "**Chicken Sandwiches**".

ANOVA Table:

Model: df = 2 (two predictors)

F-Stat = 295.16

P-Value ≤ 0.0001

Reject H_0 and conclude the model is useful.

6. If the main hypothesis test has rejected the null, we can investigate any individual coefficients, check the **partial regression plots**.

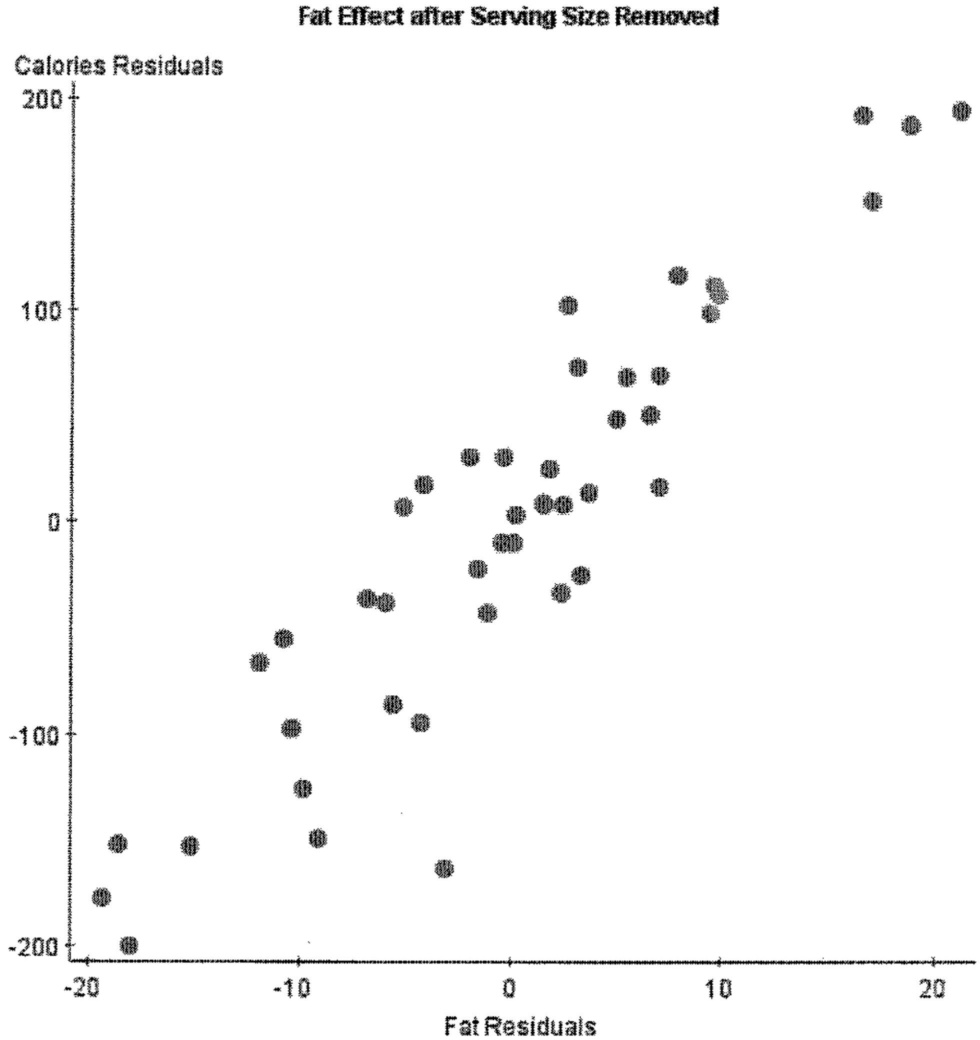
Partial regression plots remove the effects of the other explanatory variables so that you can investigate one particular explanatory variable. A bit complicated to produce, you can handle it.

Say we want to investigate the "***Fat***" effective on "***Calories***", but after "***Serving Size***" has been accounted for. We need to let StatCrunch calculate two sets of residuals:

y-axis: Residuals of "***Calories***" on "***Serving Size***".

x-axis: Residuals of "***Serving Size***" on "***Fat***".

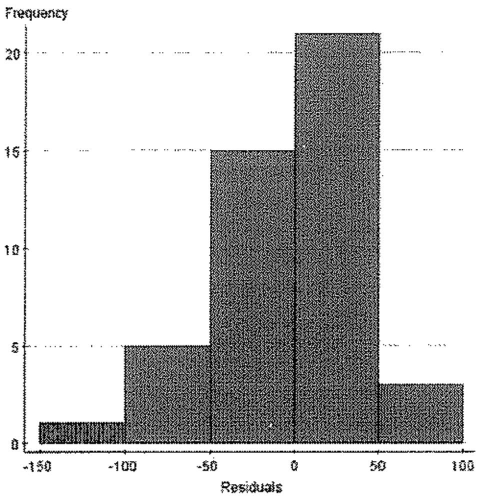
Make this partial regression plot for "**Chicken Sandwiches**":

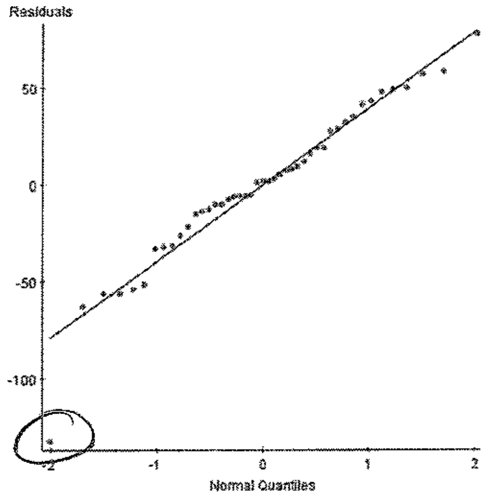


This tells us that Fat is still meaningful even after Serving Size is accounted for

7. To run any inference on individual coefficients, check the residuals for Normality. On StatCrunch, checkbox to save residuals and then make the appropriate graphs.

Do this for "**Chicken Sandwiches**":





Skewed left, one low outlier (low-carb Hardee's). $n = 45$ so we're only mildly concerned.

8. Now we can test if the individual regression coefficients are significant (the main test must have rejected the null to do this).

Is "***Serving Size***" a meaningful explanatory variable for a model that already has "***Fat***" in it? We test and conclude:

$H_0: \beta_{\text{serving size}} = 0$ vs. $H_A: \beta_{\text{serving size}} \neq 0$

P-Value < 0.001, Reject H_0, conclude it is a worthwhile predictor

Is "***Fat***" a meaningful explanatory variable for a model that already has "***Serving Size***" in it? We test and conclude:

$H_0: \beta_{\text{FAT}} = 0$ vs. $H_A: \beta_{\text{FAT}} \neq 0$

P-Value < 0.001, Reject H_0, conclude it ~~too~~ is a worthwhile predictor.

9. If all conditions are met, we can do inference on predictions.

Model: $\widehat{\text{Calories}} = 52.04 + 0.99(\text{Serving Size}) + 9.65(\text{Fat})$

Easy part: Predict the calories for a 250 gram sandwich with 25 fat grams.

$\widehat{\text{Calories}} = 52.04 + 0.99(250) + 9.65(25) = 540.79$

StatCrunch part: If we want to calculate a confidence interval for the mean y value for given x-values, we must choose exact x-values found in the dataset. Same goes for calculating prediction intervals for the next y-value for given x-values. Formulas exist, but we won't cover them. Checkboxes on StatCrunch.

Example: The McChicken is 147 grams of sandwich with 16 grams of fat (row 1). Give the interpretations of the confidence interval and prediction interval for calories.

For all sandwiches SS = 147 g
Fat = 16 g

Calories: (333.24, 370.96) 95% conf. mean calories in here.

Next Sandwich at SS = 147 g
Fat = 16 g

(269.37, 434.83) Calories.