

IST772 Chapter Notes Template: After Completing Please Submit as a PDF.

Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Wanyue Xiao

Date: October 21, 2020

Chapter Number: #9

Title of Chapter: **Interactions in ANOVA and Regression**

## Notes for Each Major Section

**Main Effect:** which is also called simple effect, is an effect of one independent variable on the dependent variable.

- Once there are at least two main effects occurred in a model, interaction might exist.

**Interaction:** occurs when the effect of one variable depends on the value of another variable.

- One can draw an interaction plot for visualization. If the lines are not parallel, that will be a hint that an interaction effect **might exist**. One needs more evidence or information to see if the interaction is statistically significant.

**One-way ANOVA:** only has a single factor with at least two levels, suiting for single factor testing.

**Factorial ANOVA:** an analysis of variance test with more than one factor.

1. Find the mean for each of the groups.
2. Find the overall mean (the mean of the groups combined).
3. Find the Within Group Variation.
4. Find the Between Group Variation
5. Find the F statistic: the ratio of Between Group Variation to Within Group Variation.

**eta-squared effect size value:** the proportion of variance in the dependent variable accounted for by the independent variable.

**Statistical power:** one should research and find out the minimum amount of data required for detecting interaction effect. Otherwise, the result will be invalid.

**Interaction in Multiple Regression:**

One could analyze the slopes of different regression lines to see if the result is significant or not.

The interpretation is that “the coefficient of the interaction term is statistically different from 0 with a t-value of X and a p-value less than 0.05”.

What’s more, one can create a scatter plot between residuals and independent variable to further check the nonlinearity issue. An ideal regression model should have a plot whose dispersion of points above and below the 0-center-line would be “random and uniform at all levels of the independent variable”, according to the textbook.

- **Heteroscedasticity:** the variance of residuals will vary based on the different levels of the independent variable.
- **Homoscedasticity:** the variance of residuals will stay stable (or about the same) across all levels.

**Centering Issue:** The interaction may be highly correlated with the other independent variables.

To solve such an issue, one can eliminate a majority of the spurious correlation by centering those independent variables around a mean of 0. One thing worth noting is that the coefficients and significance tests for main effect might change.

**delta-R-squared:** represents the quantity changed in R-squared. One can use the `modelCompare()` command (from the `LmSupport` package). Normally, the larger the value is, the stronger the interaction effect.

IST772 Chapter Notes Template: After Completing Please Submit as a PDF.

Originality Assertion: By submitting this file you affirm that this writing is your own.

## Degrees of freedom for ANOVA interactions

According to the example listed in the textbook. Originally, the dataset has 40 records. One is taken for grand mean calculation. Then, 2 df were taken since 1 df for a factor is 1 less than the number of levels in the factor. In the example, it has two factors while each factor has two level. Therefore, 1 is taken for dietary source while another 1 is taken for the type of protein. Up to now, only  $40 - 1 - 2 = 37$  were left. The between-groups df is 2.

However, the df for interaction equals to the product of the df for the main effects. Specifically, it is  $1 * 1 = 1$ . Finally,  $37 - 1 = 36$  were left, which is the residuals or the within-groups degrees of freedom.

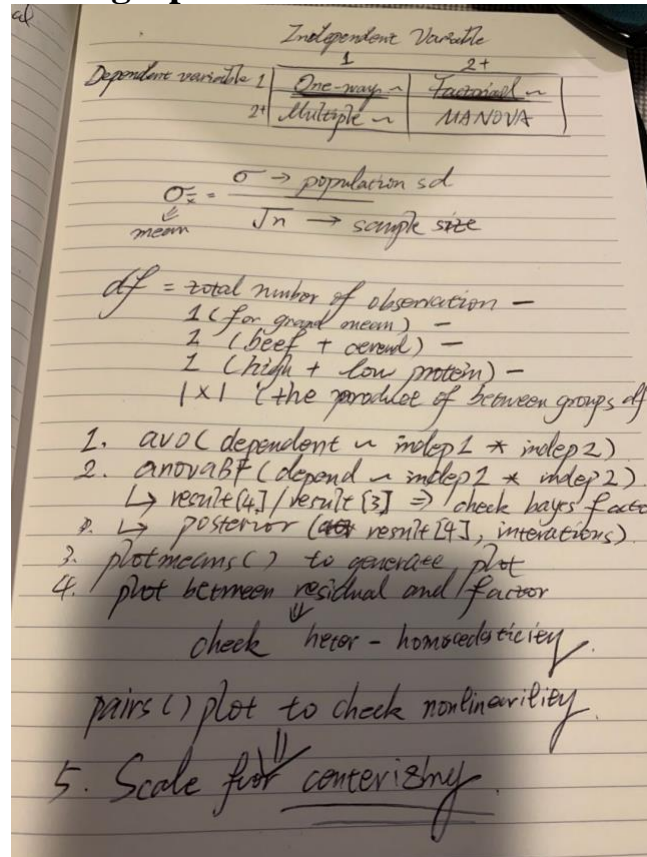
## Calculation of Standard Error

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Formula:

The value of standard value of the mean equals to the value of population standard deviation divided by the value of square root of the sample size. Such a formula indicates that the value of standard error becomes smaller when the sample size increases.

## Photographs of Your Handwritten Notes

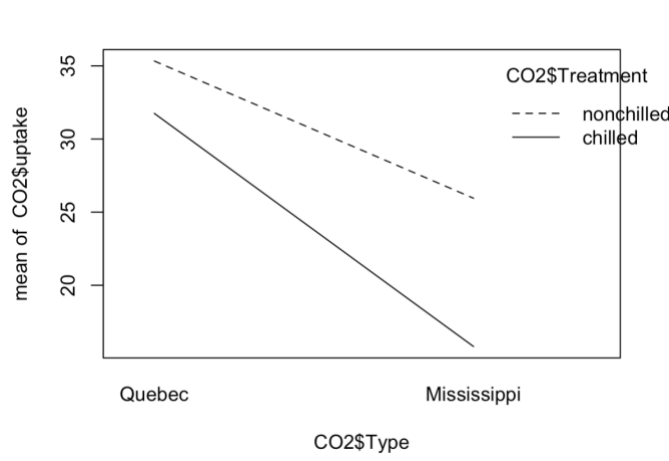


## Exercise Review

For this chapter, we don't have any exercise requirement. I will choose some question to answer.

1.  
 The dependent variable is the uptake (numeric). The factorial independent variables are plant, type, treatment, corresponding to 12, 2, and 2 levels respectively. One possible two-way interaction is that the type's and treatment's effect on uptake.

2.



Since the two lines are not parallel, there is a possibility that a interaction might exist.

3.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Type	1	3366	3366	52.509	2.38e-10	***
Treatment	1	988	988	15.416	0.000182	***
Type:Treatment	1	226	226	3.522	0.064213	.
Residuals	80	5128	64			

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

For Type and Treatment, the p-value is far lower than 0.05, suggesting that either of them is significant. The interaction, however, has a p-value which is higher than 0.05, indicating a failure to reject the null hypotheses which the effect of interaction between Type and Treatment on uptake will equal to 0 (make no difference).

4.

```
> anovaBF(uptake ~ Type * Treatment, data=newCO2)
|=====| 100%
Bayes factor analysis
-----
[1] Type : 2366422 ±0%
[2] Treatment : 11.62652 ±0%
[3] Type + Treatment : 248419553 ±0.84%
[4] Type + Treatment + Type:Treatment : 286576836 ±2.26%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

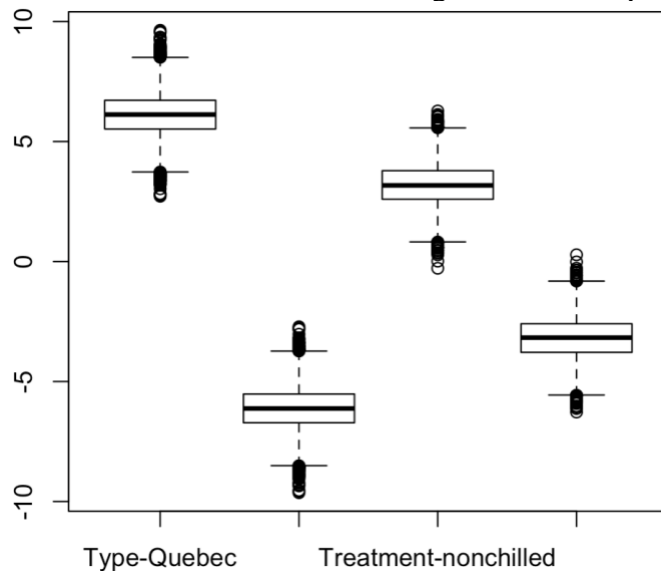
IST772 Chapter Notes Template: After Completing Please Submit as a PDF.  
Originality Assertion: By submitting this file you affirm that this writing is your own.

```
> result[4]/result[3]
Bayes factor analysis
-----
[1] Type + Treatment + Type:Treatment : 1.228648 ±1.69%

Against denominator:
  uptake ~ Type + Treatment
---
Bayes factor type: BFlinearModel, JZS
```

The result has odds of 1.2 to 1 in favor of the alternative hypothesis. According to the rules of the thumb, it barely worth mentioning.

After 10000 interactions, we can generate a box plot which displays the main effects.



This plot confirms that there is a credible difference between the two levels of type and treatment since the two boxplots of each factors don't overlap.

## R Code Fragment and Explanation

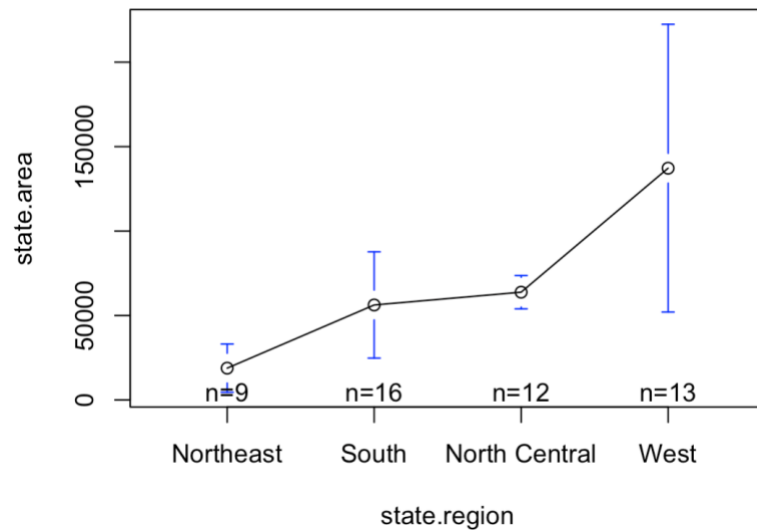
**plotmeans()** can be used to plot group means and confidence intervals.

**plotmeans(formula, data, ...)**

- **formula** symbolic expression specifying the outcome (continuous) and grouping variable (factor).
- **data** optional data frame containing the variables in the model.
- **subset** an optional vector specifying a subset of observations to be used in the fitting process.

Example:

```
plotmeans(state.area ~ state.region)
```



Each mean has an upper and a lower marker. The distance between the middle point to one of the markers indicates the standard error. Namely, the area between those two markers can be perceived as confidence interval around the mean.

### Question for Class

1. Is it possible that the result of those measurement will be contradictory? For example, the result of `anovaBF` function indicates a weak evidence in favor of the interaction effect while others indicate that the interaction is statistically significant.