CS555 Data Analysis and Visualization

Lecture 10

Kia Teymourian

One-Way Analysis of Variance and Regression

We can use the linear regression to conduct the same tests in the one-way ANOVA setting

The one-way ANOVA is the same as performing a regression where the explanatory variable in the model is a variable or variables that indicate group membership.

In order to represent **a one-way ANOVA model in the regression framework, construction of dummy variables is required.** For a particular categorical variable with k categories, k−1 dummy variables are needed.

Dummy variables are binary variables of the form:

$$\text{group}_2 = \begin{cases} 1, & \text{if observation is in group 2} \\ 0, & \text{otherwise} \end{cases}$$

$$\text{group}_3 = \begin{cases} 1, & \text{if observation is in group 3} \\ 0, & \text{otherwise} \end{cases}$$

$$\vdots$$

$$\text{group}_k = \begin{cases} 1, & \text{if observation is in group } k \\ 0, & \text{otherwise} \end{cases}$$

The category omitted (group 1 in this case) is referred to as the reference group.

A dummy variable for the reference group is not needed as those in this group can be identified as having values of 0 for all of the defined dummy variables.

One-Way Analysis of Variance and Regression

The table summarizes how dummy variables would be defined for a grouping variable with k categories.

If you knew the values of each of the dummy variables for a particular observation, you would then know what group the observation was a member of.

The following model can be used to compare means between groups in the regression framework:

$$y = \beta_0 + \sum_{i=2}^{k} \beta_{i-1} \text{group}_i + e$$

where

y is the response or dependent variable.

group_2, group_3,…, group_k are the dummy variables that represen membership in groups 2, 3, … and k, respectively.

β_0 is the intercept (the sample mean in the reference group [group 1 in this case]).

β_1 is the mean difference between group 2 and the reference group (group 1 in this case).

β_2 is the mean difference between group 3 and the reference group.

…

β_{k-1} is the mean difference between group k and the reference group.

e is the random error which we assume is normally distributed with a mean of 0 and a variance of σ^2.

Group	Dummy Variables			
	group_2	group_3	…	group_k
1	0	0		0
2	1	0		0
3	0	1		0
…				
k	0	0		1

One-Way Analysis of Variance and Regression

We can use this model to test the hypotheses of interest in the one-way ANOVA setting.

We can use this model to test the typical ANOVA null hypothesis
$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ (All underlying population means are equal)
$H_1: \mu_i \neq \mu_j$ for some i and j (At least two of the k underlying population means are different or not all of the underlying population means are the same/equal).

Mathematically, these hypotheses are equivalent to the global F-test hypotheses in regression where we test the null hypothesis that **all slope coefficients are equal to 0 ($H_0: \beta_1 = \beta_2 = ... = 0$)** versus the alternative that at least one of the slope coefficients is different from zero
($H_1: \beta_i \neq 0$ for at least one i).

Given that the **overall F-test for one-way ANOVA is equivalent to the global F-test for multiple linear regression** (**when dummy variables are used in the regression to represent group membership)**, the ANOVA table in each case is the same.

This also means that the sum of squares between and the sum of squares within from the **one-way ANOVA model** are equivalent to the **regression** sum of squares and the residual sum of squares, respectively.

Example – SBP by smoking status (continued)

Create dummy variables

```
> data$g0 <- ifelse(data$group=='currentHeavySmoker', 1, 0)
> data$g1 <- ifelse(data$group=='currentLightSmoker', 1, 0)
> data$g2 <- ifelse(data$group=='formerSmoker', 1, 0)
> data$g3 <- ifelse(data$group=='neverSmoker', 1, 0)
```

One-way ANOVA using lm() function

```
> m2 <- lm(data$SBP~data$g0+data$g1+data$g2, data=data)
> summary(m2)

> m3 <- lm(data$SBP~data$g1+data$g2+data$g3, data=data)
> summary(m3)

> m4 <- lm(data$SBP~data$g0+data$g2+data$g3, data=data)
> summary(m4)
```

The golf example

In the golf ball example, the one-way ANOVA table was as follows:

	SS (Sum of Squares)	df (degrees of freedom)	MS (Mean Square)	F
Between	2583.3335	2	1291.6668	20.67
Within	750	12	62.5	
Total	3333.3335			

We create dummy variables for 2 of the 3 brands, for example, for the Nike and Callaway brands, and ran a regression predicting distance from these two dummy variables.

The regression model in this case that is equivalent to the above one-way ANOVA model is given by $y = \beta_0 + \beta_{Nike} group_{Nike} + \beta_{Callaway} group_{Callaway} + e$

where $group_i$ is a dummy variable indicating whether or not the observation is of brand i.

The golf example: R commands

```
> golf$g0 <- ifelse(brand=='Callaway', 1, 0)
> golf$g1 <- ifelse(brand=='Nike', 1, 0)
> golf$g2 <- ifelse(brand=='Titleist', 1, 0)

> m <- lm(dist~g1+g2, data=golf)
> summary(m)
```

```
Call:
lm(formula = dist ~ g1 + g2, data = golf)

Residuals:
   Min     1Q Median     3Q    Max 
   -10     -5      0      5     10 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  285.000      3.536   80.61  < 2e-16 ***
g1           -25.000      5.000   -5.00 0.000309 ***
g2             5.000      5.000    1.00 0.337049    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.906 on 12 degrees of freedom
Multiple R-squared:  0.775,	Adjusted R-squared:  0.7375 
F-statistic: 20.67 on 2 and 12 DF,  p-value: 0.0001297
```

```
> anova(m)
```

```
Analysis of Variance Table

Response: dist
          Df Sum Sq Mean Sq F value   Pr(>F)    
g1         1 2520.8  2520.8  40.333 3.66e-05 ***
g2         1   62.5    62.5   1.000    0.337    
Residuals 12  750.0    62.5                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The golf example

Pairwise comparisons can also be performed in the regression framework. The most simple is the test of the underlying mean from group i to the reference group (group 1 as outlined above).

In the one-way ANOVA framework, the test of the null hypothesis $\mu_i=\mu_1$ would be accomplished through the use of the t statistic:

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{s_p^2(\frac{1}{n_i} + \frac{1}{n_j})}}$$

which follows a t-distribution with n−k degrees of freedom under H_0 (k here is the number of groups).

This test is the same as the t-test in the regression setting which tests the null hypothesis $\beta_i=0$ after controlling for the other independent variables in the model.

The t-statistic above is the same as the t-statistic for multiple regression: $t = \frac{\hat{\beta}_i}{SE_{\hat{\beta}_i}}$

which follows a t-distribution with n−k degrees of freedom under H H_0 (k here represents the number of groups and not the number of variables in the model as we had defined it in the regression context).

The golf example – Pairwise Comparisons

Though it is also possible to perform other **pairwise comparisons** in the regression setting (for example, to test if there is a difference between groups i and j where neither i nor j represent the reference group), it **requires testing of null hypotheses of the form $\beta_i=\beta_j$ which are not standard** output from the regression model in most statistical packages and thus requires additional steps in order to calculate the test statistic for such comparisons.

The same comparisons are relatively easy to make from the one-way ANOVA setting.

If one were interested in using regression to make these comparisons, the **easiest way to do it is to run more than one regression analysis and change the reference group each time to a different group**.

Obviously, conducting one-way ANOVAs gives this output standardly as such is often easier to implement in statistical software packages like R.

The golf example - Pairwise Comparisons vs. Regression Model

For the golf ball distance by brand example, the below table summarizes the results of the **pairwise comparisons**. The group means were 285, 260 and 290 for Callaway, Nike, and Titleist, respectively.

Comparison	Mean Difference	t	df	Unadjusted p-value
Titleist versus Callaway	5	1.0	12	0.34
Titleist versus Nike	30	6.0	12	0.000062
Callaway versus Nike	25	5.0	12	0.00031

The summary of the output of the **regression model** below are show in the table:

$$\hat{y} = \beta_0 + \beta_{\text{Nike}} \text{group}_{\text{Nike}} + \beta_{\text{Callaway}} \text{group}_{\text{Callaway}}$$

	Estimate	SE	t	$Pr(>\|t\|)$
Intercept	290	3.536	82.02	< 0.0001
Nike versus Titleist	-30	5	-6	0.000062
Callaway versus Titleist	-5	5	-1	0.34

One-Way Analysis of Covariance (ANCOVA)

One-Way Analysis of Covariance (ANCOVA) Intro.

Sometimes we have other variables or factors that have an effect on the dependent variables that we are analyzing. But these variables are not actually independent variables.

You want to control the effect of other variables in your analysis. **For example:** if you study the effect of different diet types on weight loss other factors like age or level of training can have an effect. You do not want to consider them as a separate independent variables, you just want to control their effect on your analysis portion.

We have multiple variables that have an effect on a single numeric outcome. This is where we use analysis of covariance.

The other variables are **variables** that are are called "**covariates**". These variables provide additional variations in the subject response value. **We want to control these variables. Covariates** are also named "Intervening" or "Confounding Variables".

Ancova is an extension of Anova and can be seen as a combination of Anova and Regression.

What ANCOVA does it simply takes out the effects of covariates out of anova analysis.

One-Way Analysis of Covariance (ANCOVA)

ANCOVA is a general linear model that is a **combination of ANOVA and regression** when you have some **categorical factors** and some **quantitative continuous variables**

The **variables** (on which to perform regression) are called “**covariates**”

Often these covariates are not necessarily of primary interest, but still their inclusion in the model **will help explain more of the response, and hence reduce the error variance**.

ANCOVA evaluates whether **population means of a response variable are equal across levels** of a categorical explanatory variable, while **statistically controlling** for the effects of other continuous variables that are not of primary interest, known as covariates.

Mathematically, **ANCOVA decomposes the variance in the response variable into variance explained by the covariates**, variance explained by the categorical explanatory variables, and residual variance.

One-Way Analysis of Covariance (ANCOVA)

The **one-way ANCOVA model** can be expressed as:

$$y = \beta_0 + \sum_{i=2}^{k} \beta_{i-1}\mathrm{group}_i + \sum_{i=1}^{j} \beta_{k+i-1} x_i + e$$

where
y is the response or dependent variable.
$group_2$, $group_3$,…, $group_k$ are the dummy variables that represent membership in groups 2, 3, … and k, respectively.
β_0 is the intercept (the sample mean in the reference group [group 1 in this case]).
β_1 is the **mean difference** between group 2 and the reference group (group 1 in this case).
β_2 is the **mean difference** between group 3 and the reference group.
…
β_{k-1} is the **mean difference** between group k and the reference group.
β_k is the expected change in y for each one unit change in x_1, across all groups and after adjusting for the other covariates x_2,x_3,…,x_j.
…
β_{k+j-1} is is the expected change in y for each one unit change in x_j, across all groups and after adjusting for the other covariates x_1, x_2,…,x_{j-1}.
e is the random error which we assume is normally distributed with a mean of 0 and a variance of σ^2.

One-Way ANCOVA

In one-way ANCOVA framework, we are mainly **interested in the overall model results as well as the results for the grouping factor** (after adjusting for other variables in the model).

We can test the overall null hypothesis
$\mathbf{H_0: \beta_1 = \beta_2 = \cdots = \beta_{k+j} = 0}$
(none of the variables in the model is predictive of the dependent variable y)

versus the alternative hypothesis

$\mathbf{H_1}$**: there is at least one** $\boldsymbol{\beta_i \neq 0}$ (at least one of the variables is predictive of the dependent variable y)

If the global null hypothesis is rejected, then we can test if the underlying population means are different across groups after controlling for the other variables in the model.
$\mathbf{H_0: \beta_1 = \beta_2 = \cdots = \beta_{k-1} = 0,\ \beta_k \neq 0,\ \ldots,\ \beta_{k+j} \neq 0}$
(all underlying population means are equal after controlling for $x_1,\ldots,x_j$)
versus the alternative hypothesis
$\mathbf{H_1}$**: At least two of the k underlying population means are different or not all** of the underlying population means are the same/equal after controlling for $x_1,\ldots,x_j$

One-Way ANCOVA

In the ANCOVA setting, if there is a difference between groups after adjusting for the other covariates in the model, we follow up with **a comparison for the adjusted group means** (instead of a comparison of the sample means).

The adjusted means are often called least-squares means or LS means.

The LS means represent the **mean value for each group** that is adjusted for the other **covariates included in the model.**
The LS mean in each group is calculated using the **least-squares regression equation** using the mean values of the covariates in the model.

The general procedure and interpretation of the **one-way ANCOVA model** is similar to the procedure introduced for the one-way ANOVA. The main **difference is that in the ANCOVA setting, the inferences made are based on comparisons made after adjusting for** or controlling for the covariates included in the model.

One-Way ANCOVA

Don’t use aov() function as it will not produce expected output

Need to use **Anova() function from package “car”** to get “Type III” sums of square

> \> **install.packages(‘car’)**
> \> **library(car)**

> \> **Anova**(lm(data$response~data$group **+ data$covariate**), type=3)

car: Companion to Applied Regression

Functions and Datasets to Accompany J. Fox and S. Weisberg, An R Companion to Applied Regression, Second Edition, Sage, 2011.

https://cran.r-project.org/web/packages/car/index.html

An example: One-Way ANCOVA

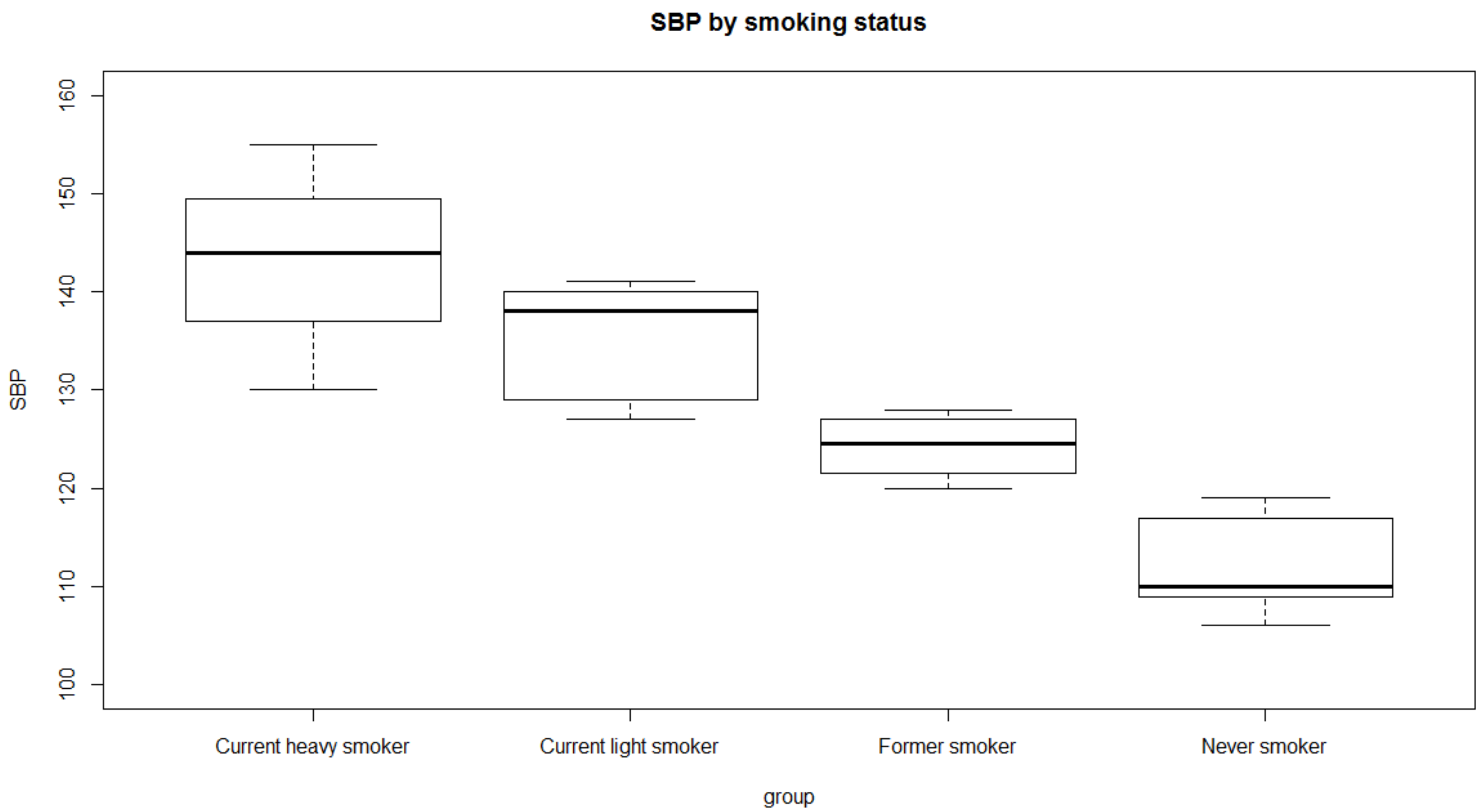
A random sample of current light smokers, current heavy smokers, former smokers, and those who have never smoked was taken to determine if mean systolic blood pressure (SBP) differs across smoking status categories.

Given that it is known that SBP increases with advancing age and given the fact that smoking preference are known to differ by age, we'd want to conduct an ANCOVA so that we can be sure that differences we see between smoking categories aren't due to differences in age.

It is preferable in this case to use the ANCOVA methodology so that we can adjust for age.

An example: One-Way ANCOVA

```
> data <- read.csv("smoker.csv")
> data
> aggregate(data$SBP, by=list(data$group), summary)
> boxplot(data$SBP~data$group, data=data, main="SBP by smoking status", xlab="group", ylab="SBP", ylim=c(100, 160))
```



An example: One-Way ANCOVA

First, ran a one-way ANOVA (without adjustment for age). The **global F-test showed that mean SBP differed by smoking category** (F=21.49 on 3 and 15 degrees of freedom, p <0.001).

```
> m<- aov(data$SBP~data$group, data=data)
> summary(m)

            Df Sum Sq Mean Sq F value  Pr(>F)
data$group   3 2786.2   928.7   21.49 1.1e-05 ***
Residuals   15  648.3    43.2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


> qf(.95, df1=3, df2=15)
[1] 3.287382
```

One-Way ANCOVA

Re-run ANOVA adjusting for Age
> library(car)

> **Anova(lm(data$SBP~data$group+data$age), type=3)**
Type should be 3. It defines the different types of sums of squares. Read more about it here

https://www.r-bloggers.com/anova-%E2%80%93-type-iiiiii-ss-explained/

Least square means
> install.packages('lsmeans') # one time installation only
> library(lsmeans)
> **options**(contrasts=c("contr.treatment", "contr.poly"))
> **lsmeans**(**lm(data$SBP~data$group+data$age)**, **pairwise~data$group**, adjust="none")

- **options**: Allow the user to set and examine a variety of global options which affect the way in which R computes and displays its results. The "constrasts" set in your R environment determine how categorical variables are handled in your models.

- In the **lsmeans** function, model specifies the model object that was previously fitted. Note the specialized formula where pairwise indicates that all pairwise comparisons should be conducted,

An example: One-Way ANCOVA

After adjusting for age using an ANCOVA model, the differences seen in the one-way ANOVA setting were attenuated and the F-test for the effect of smoking status was **no longer significant** (F=1.774 on 3 and 14 degrees of freedom, p =0.1982).

The least square means **(adjusted for age)** were 129.87, 144.01, 144.01, and 155.11 for the heavy, light, former and never smokers, respectively.

As such, **the differences that we saw in the one-way ANOVA model were due to age differences across the smoking groups** as **opposed to true differences in SBP attributable only to smoking status.**

One-Way ANCOVA

Generate least square means (covariate adjusted means) and comparisons

\> **library**(lsmeans)
\> options(contrasts=c("contr.treatment", "contr.poly"))
\> **lsmeans**(lm(data$response~data$group+data$**covariate**), **pairwise~data$group**, adjust=[method])

Note: method = "tukey", "scheffe", "sidak", "bonferroni", "dunnettx", "mvt", "none")

lsmeans: Least-Squares Means

Obtain least-squares means for many linear, generalized linear, and mixed models. Compute contrasts or linear functions of least-squares means, and comparisons of slopes. Plots and compact letter displays.

https://cran.r-project.org/web/packages/lsmeans/vignettes/using-lsmeans.pdf

Two-Way Analysis of Variance

Two-Way Analysis of Variance

Sometimes we are interested in comparing the mean response across groups when there is more than one factor to be considered.

When there are two factors (each with two or more levels) that we are interested in, we use the two-way ANOVA methodology instead of the one-way ANOVA methodology.

The goal of a two-way ANOVA is to look at the effects of each factor after controlling for the effects of the other factor.

Within the two-way ANOVA framework, we seek to test the following:

- **Whether or not the first factor impacts the mean outcome** after controlling for the second factor. Here, we test H_0: All underlying population means are equal across levels of the first factor, after controlling for the second factor.

- **Whether or not the second factor impacts the mean outcome** after controlling for the first factor. Here, we test H_0: All underlying population means are equal across levels of the second factor, after controlling for the first factor.

Two-Way Analysis of Variance

The alternative hypothesis in each case is that the underlying populations means are not equal across levels of the factor tested after controlling for the other.

In other words, at least two of the underlying group means are different across the factor tested, after controlling for the other factor.

Consider the following notation:
r = total number of levels of factor A.
c = total number of levels of factor B.
μ_{ij} = underlying population mean for those in the ith level of factor A and the jth level of factor B.

$\mu_{i.}$ = underlying population mean for those in the ith level of factor A across all *c* levels of factor B.

$\mu_{.j}$ = underlying population mean for those in the jth level of factor B across all *r* levels of factor A.

		Factor B				Total
		1	2	...	c	
Factor A	1	μ_{11}	μ_{12}		μ_{1c}	$\mu_{1.}$
	2	μ_{21}	μ_{22}		μ_{2c}	$\mu_{2.}$
	...					
	r	μ_{r1}	μ_{r2}		μ_{rc}	$\mu_{r.}$
Total		$\mu_{.1}$	$\mu_{.2}$		$\mu_{.c}$	

Two-Way Analysis of Variance

The global F-test is used to test the null hypothesis that there is no effect of either factor versus the alternative hypothesis that one of the factors has an effect (there is an effect of either factor A or factor B or both).

The test for Factor A is testing the null hypothesis H_0:$\mu_{1.}=\mu_{2.}=\ldots=\mu_{r.}$ against the alternative that at least two of the underlying means are different $\mu_{i.}\neq\mu_{j.}$ for some i and j.

The test for Factor B is testing the null hypothesis H_0:$\mu_{.1}=\mu_{.2}=\ldots=\mu_{.c}$ against the alternative that at least two of the underlying means are different $\mu_{.i}\neq\mu_{.j}$ for some i and j.

Generally, the procedure for testing follows the procedure for the one-way ANOVA.

Interaction

Besides **the global test and the tests for each factor**, **a test of interaction must be performed** to understand if there is a non-additive effect of the either of the factors on the response.

One must first conduct an interaction test in the two-way ANOVA setting before the “main effects” type model discussed above can be conducted.

The concept of interactions is very important in statistics and in performing any type of multivariable analysis (including **ANCOVA**, **multivariable regression, and two-way ANOVAs**).

Interactions are when

the effect of one factor is not constant over levels of another factor.

Interaction

There is **an effect of the drug**, but **no effect of gender**.

The presence and absence of an effect can be observed via the use of the **marginal means** (those in the total column/row). Those on drug B have a higher mean SBP (200) versus those on drug A (100). There does not appear to be an effect of gender (both males and females have an average SBP of 150).

		Drug		Total
		A	B	
Gender	F	100	200	150
	M	100	200	150
Total		100	200	

Interaction

There is **an effect of the drug** and **an effect of gender**.

The presence and absence of an effect can be observed via the use of the **marginal means** (those in the total column/row). Those on drug B have a higher mean SBP (350) versus those on Drug A (150). There also appears to be an effect of gender (males have a higher SBP than females [200 versus 300, respectively]).

Here, the effects of gender are consistent across each drug and the effect of the drug are consistent across levels of gender. We conclude that though males tend to have higher SBPs than women, those on drug A have smaller SBPs than those on drug B. An extension of this (assuming the right experimental conditions), is that regardless of whether someone is a male or female, they will have lower SBP if they take drug A. In this case, there is not an interaction present.

		Drug		Total
		A	B	
Gender	F	100	300	200
	M	200	400	300
Total		150	350	

Interaction

It appears as if **there is not an effect of drug or gender**, since all of the **marginal means** are equal.

On closer examination, it appears there is an effect of these factors, **but it varies based on the level of each factor.**
Interactions are when the effect of one factor is not constant over levels of another factor. There is not one simple way to describe the effect of each factor (since the effect depends on the value of the other factor).

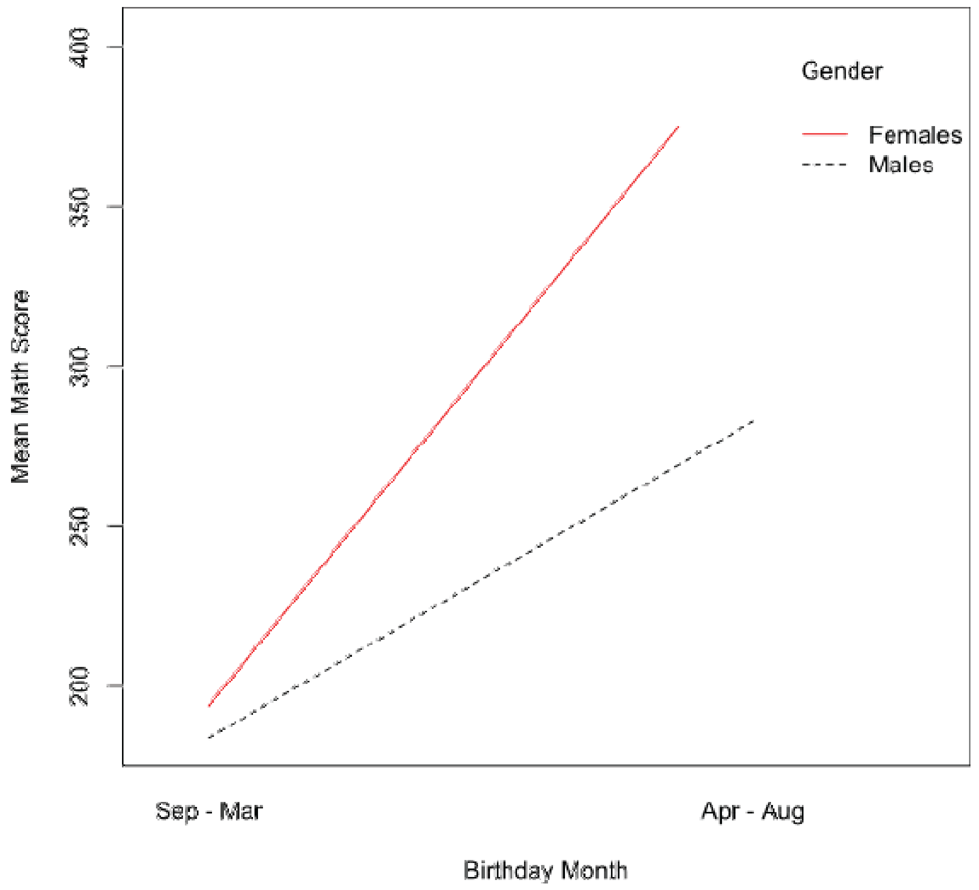
The effect of drug is not consistent. Males on drug A have higher SBPs than males on drug B. Females on drug B have higher SBPs than females on drug A. Likewise, the effect of gender is not consistent. Among those on drug A, males have higher SBPs than females. Among those on drug B, females have higher SBPs than males. Because the effect of the drug depends on gender (and equivalently, the effect of gender depends on the drug used), we say that these factors interact.

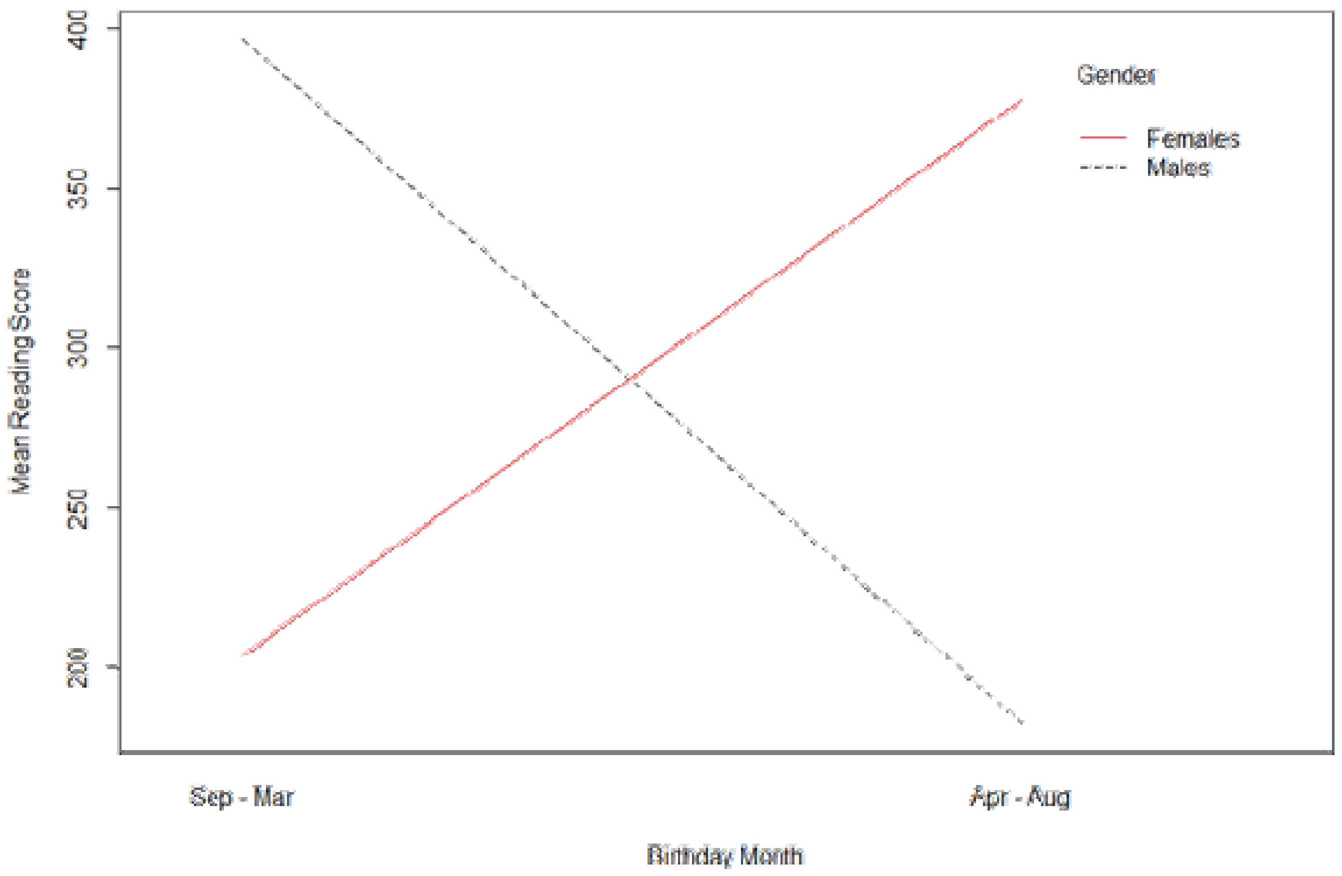
		Drug		Total
		A	B	
Gender	F	100	300	200
	M	300	100	200
Total		200	200	

Interaction plots

An educator is interested in the effect of **birth month and gender on average reading and math scores** for students in the fourth grade. A random sample is taken from each combination.

There is an interaction given that there is not one effect of birth month across both genders.





Interaction

If there is an interaction between two factors, then the conclusion that we'd get from the two-way ANOVA are incorrect at worst and misleading at best.

Before you perform a main effects two-way ANOVA, you must first check if an interaction is present.

You do this by including an interaction term in the two-way ANOVA model.
If the p-value of the interaction term **is significant**, the two-way ANOVA procedure is not appropriate. In this case, one must stratify the analysis by one of the factors and perform separate one-way ANOVAs for the second factor for each level of the first.

If the interaction p-value **is not significant**, then one can proceed with the typical two-way ANOVA procedure and assume that the effect of each factor is consistent over the other.

Given the importance of detecting interactions and the difference in interpretation of results in the presence of an interaction, alpha levels of 0.10 are often used to check if there is an interaction (as opposed to the more common 0.05 level of significance used in most other settings).

Two-Way Analysis of Variance

Use Anova() function for ANCOVA (Anova() with Capital ”A”)

> **Anova**([model], type=3)

First test interaction model
> model = **lm**(data$response ~ data$group1+ data$group2+ **data$group1 * data$group2**)

Visualize relationship using interaction.plot() function
> **interaction.plot**(data$group1, data$group2, data$response, col=1:2)

If **p-value for the interaction is not significant**, then **run regular two-way ANOVA**

> model = **lm**(data$response ~ data$group1+ data$group2)

An example: Two-Way Analysis of Variance

An exercise physiologist wants to examine whether stretching and wearing ankle weights affect the value of exercise on treadmills. To carry out her study, she recruits subjects who have roughly the same level of physical fitness, and divides them randomly into four groups:

- with or without ankle weighs
- with or without a stretching period before the exercise.

Using the amount of calories burned as the response, carry out a two-way ANOVA to determine whether stretching and wearing ankle weights have significant effects on exercise.

These are the variables in the data set:

Name	Type	Description
PreStretch	char	stretch group (Stretch, No stretch)
AnkleWeights	char	weights group (Weights, No weights)
Energy	num	calories burned
Speed	num	average speed (in meters per minute)
Oxygen	num	oxygen consumed (in liters)

An example: R commands

```
# Exercise example
exercise <- read.csv("exercise.csv")
exercise
attach(exercise)

#Test interactions
model <- lm(Energy~PreStretch+AnkleWeights+PreStretch*AnkleWeights, data=exercise)
summary(model)

Anova(model, type=3)
model1 <- lm(Speed~PreStretch+AnkleWeights+PreStretch*AnkleWeights, data=exercise)
summary(model1)

Anova(model1, type=3)
model2 <- lm(Oxygen~PreStretch+AnkleWeights+PreStretch*AnkleWeights, data=exercise)
summary(model2)
Anova(model2, type=3)
```

An example: R commands

```
# Generate interaction plots
interaction.plot(PreStretch, AnkleWeights, Energy, col=1:2)
interaction.plot(PreStretch, AnkleWeights, Speed, col=1:2)
interaction.plot(PreStretch, AnkleWeights, Oxygen, col=1:2)

#If interaction is significant, need to stratify (by more of the two factors)
stretch <- exercise[which(PreStretch=='Stretch'),]
nostretch <- exercise[which(PreStretch=='No stretch'),]
summary(aov(Energy~AnkleWeights, data=stretch))
summary(aov(Energy~AnkleWeights, data=nostretch))
summary(aov(Speed~AnkleWeights, data=stretch))
summary(aov(Speed~AnkleWeights, data=nostretch))
summary(aov(Oxygen~AnkleWeights, data=stretch))
summary(aov(Oxygen~AnkleWeights, data=nostretch))
```