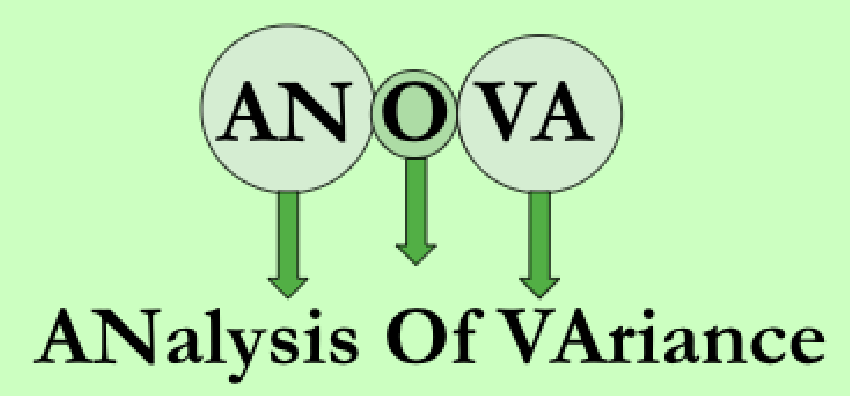


Introduction to ANOVA

10/24/2024 (Week 9)

Jingjing Yang, PhD
Associate Professor of Human Genetics
Jingjing.yang@emory.edu

Outline

- ANOVA
- One-way ANOVA
- Connection with Regression
- Two-way ANOVA

Regression or ANOVA/t-tests?

- Regression emphasizes overall weight of an independent variable predictively
    - Prediction
    - Test

- ANOVA/t-tests emphasize "statistical significance" after experiment
    - Test mean differences of a continuous variable between two groups: Two sample t-test
    - Test differences of a continuous variable among multiple groups: **An**alysis **o**f **Va**riance (ANOVA)

Categorical Variables

Study how a continuous variable would change according to different factors/levels of categorical variables (e.g., group variable)

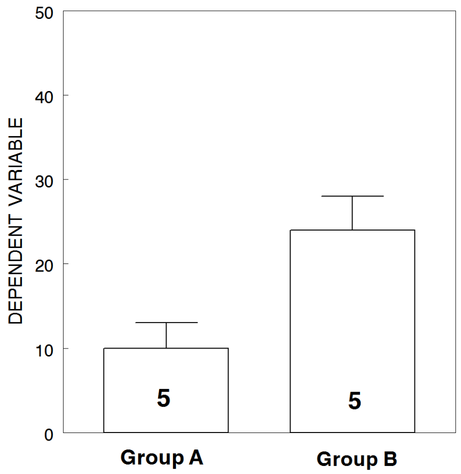
Especially when the categorical variable has factors/levels >2

Factor data type in R if often used for categorical variables

For example, sex variable in the *abalones* dataset contains values for M, F, and I

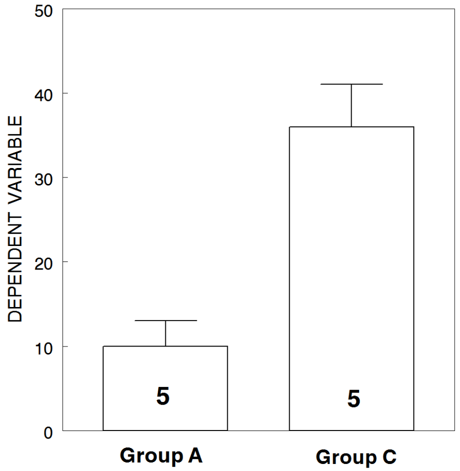
Let A, B & C be 3 levels of one factor: do any differ from the others?

If we do multiple pair-wise Two-sample **t-test**



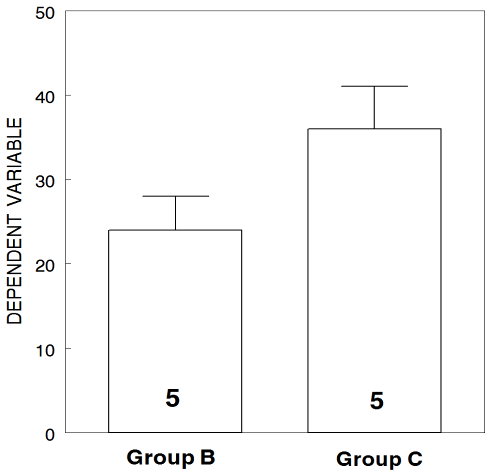
H_O: $\mu_A = \mu_B$

α=5%



H_O: $\mu_A = \mu_C$

α=5%



H_O: $\mu_B = \mu_C$

α=5%

Why Multiple Testing Matters

- In general, if we perform m hypothesis tests, what is the probability of at least 1 false positive?

P(Making an error) = a

P(Not making an error) = 1 - a

P(Not making an error in m tests) = $(1 - a)^m$

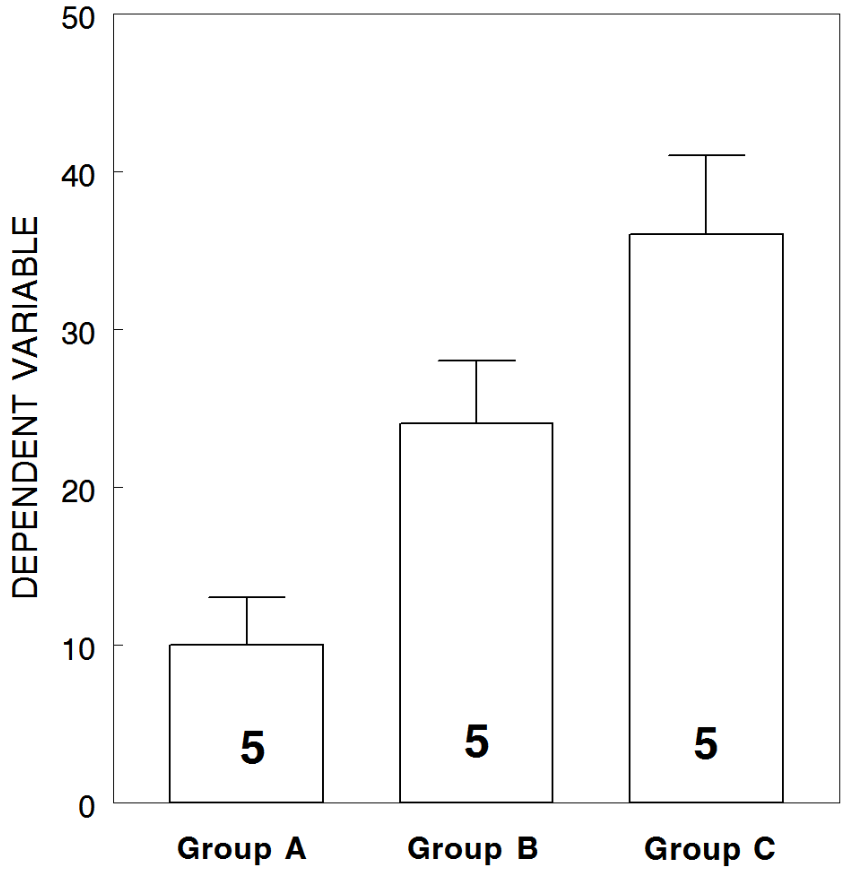
P(Making at least 1 error in m tests) = $1 - (1 - a)^m$

3 Hypothesis tests

Family-wise type1 error: 14.2% = 1 – (1 – 0.05)^3

(Week 14 Lecture about Multiple Testing)

One-way ANOVA



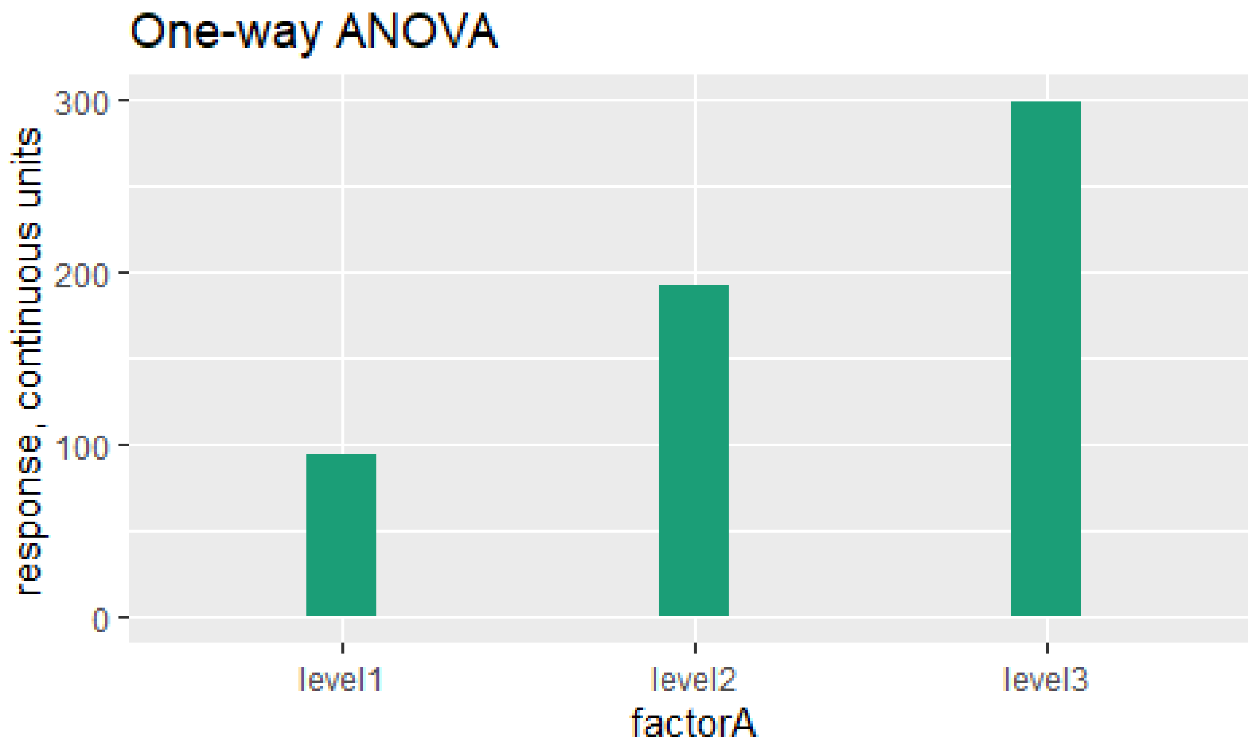
$$H_0: \mu_A = \mu_B = \mu_C$$

H_a: **At least one of the means is different**

1 Hypothesis test

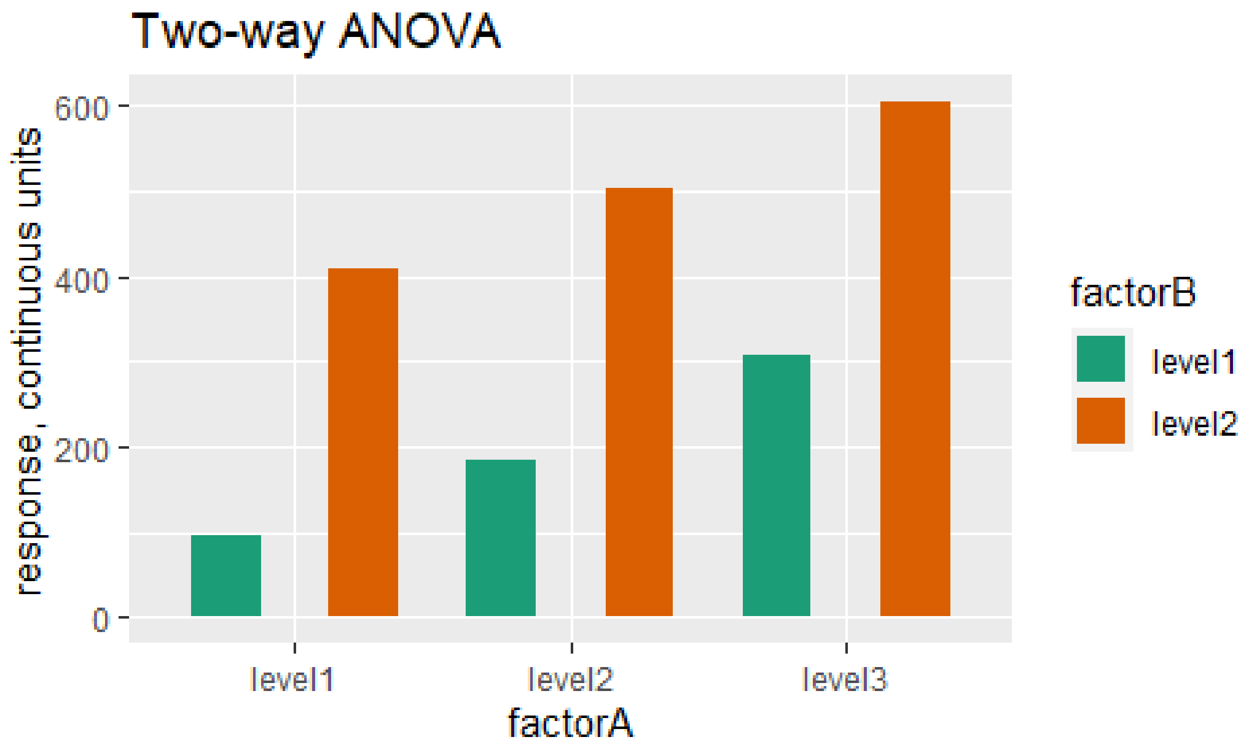
Family-wise type1 error : 5%

1 question for one factor



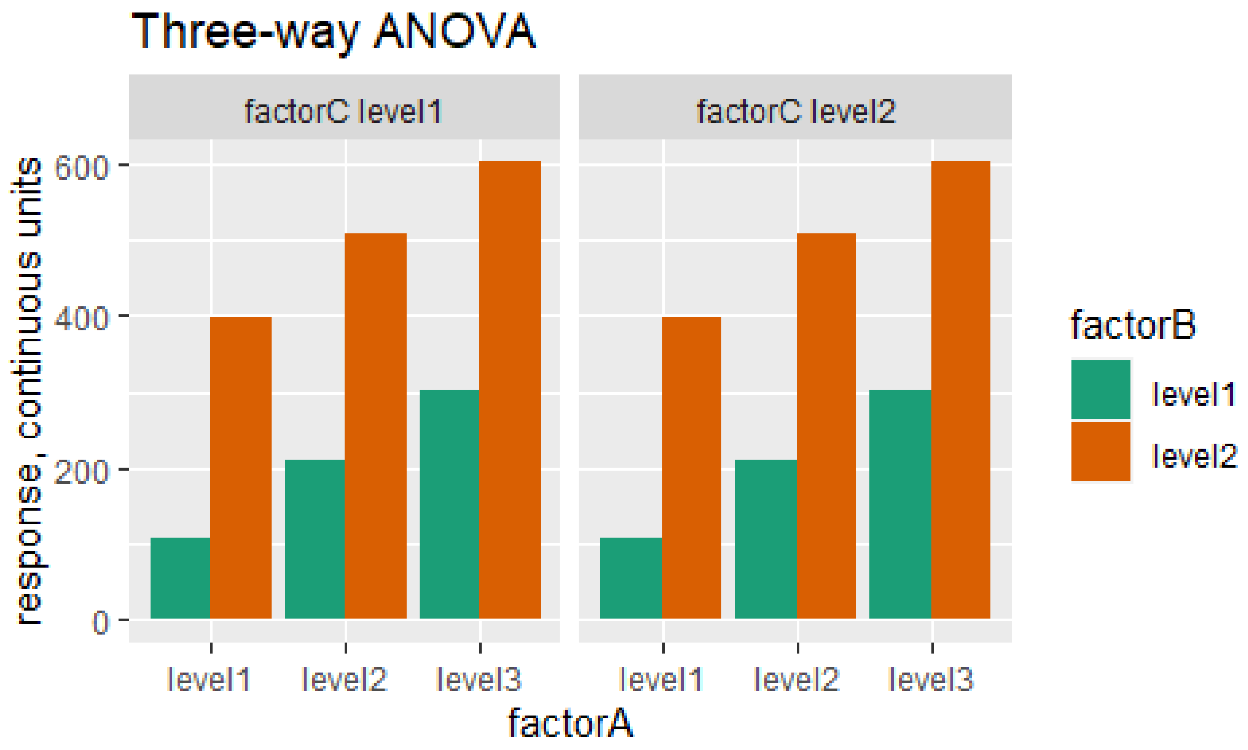
- Is factorA associated with a response?

3 questions for two factors



- Is factorA associated with a response?
- Is factorB associated with a response?
- Is the interaction between factorA and factorB associated with a response?

7 questions for three factors



- Is factorA associated with a response?
- Is factorB associated with a response?
- Is factorC associated with a response?

- Is the interaction between factorA and factorB associated with a response?
- Is the interaction between factorA and factorC associated with a response?
- Is the interaction between factorB and factorC associated with a response?

- Is the interaction among factorA, factor B, and factorC associated with a response?

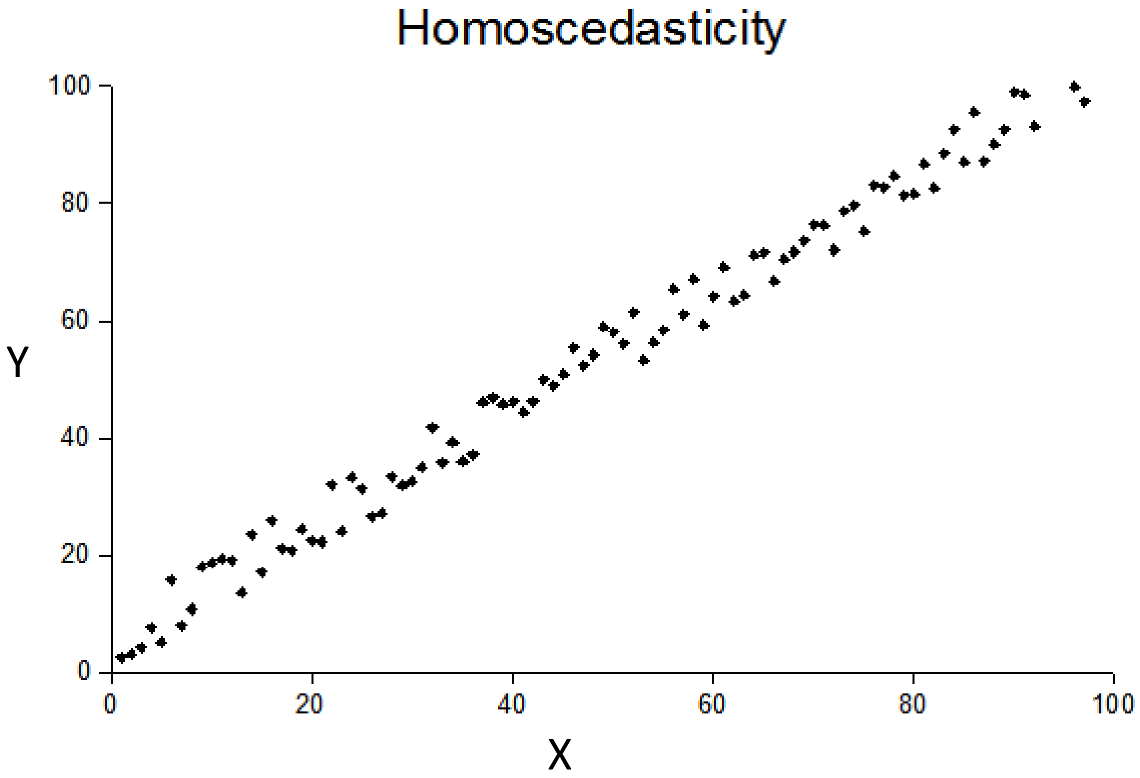
One-way ANOVA

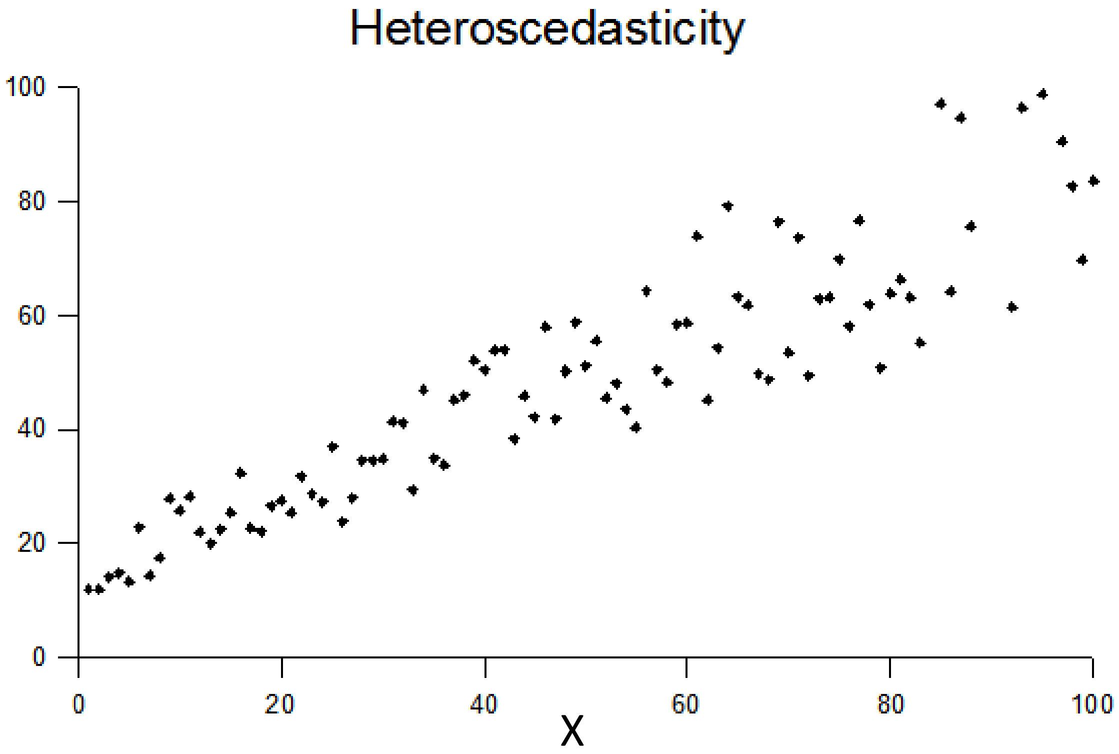
One-way ANOVA

- **Hypothesis**
  - $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$
  - H_a: At least one of the means is different

- **Assumptions**
  1. Sample independence
  2. Normality of the continuous variable per group
  3. Homogeneity of variances across all groups (aka, Homoscedasticity): assuming the residuals all have the same variance

Homoscedasticity vs. Heteroscedasticity





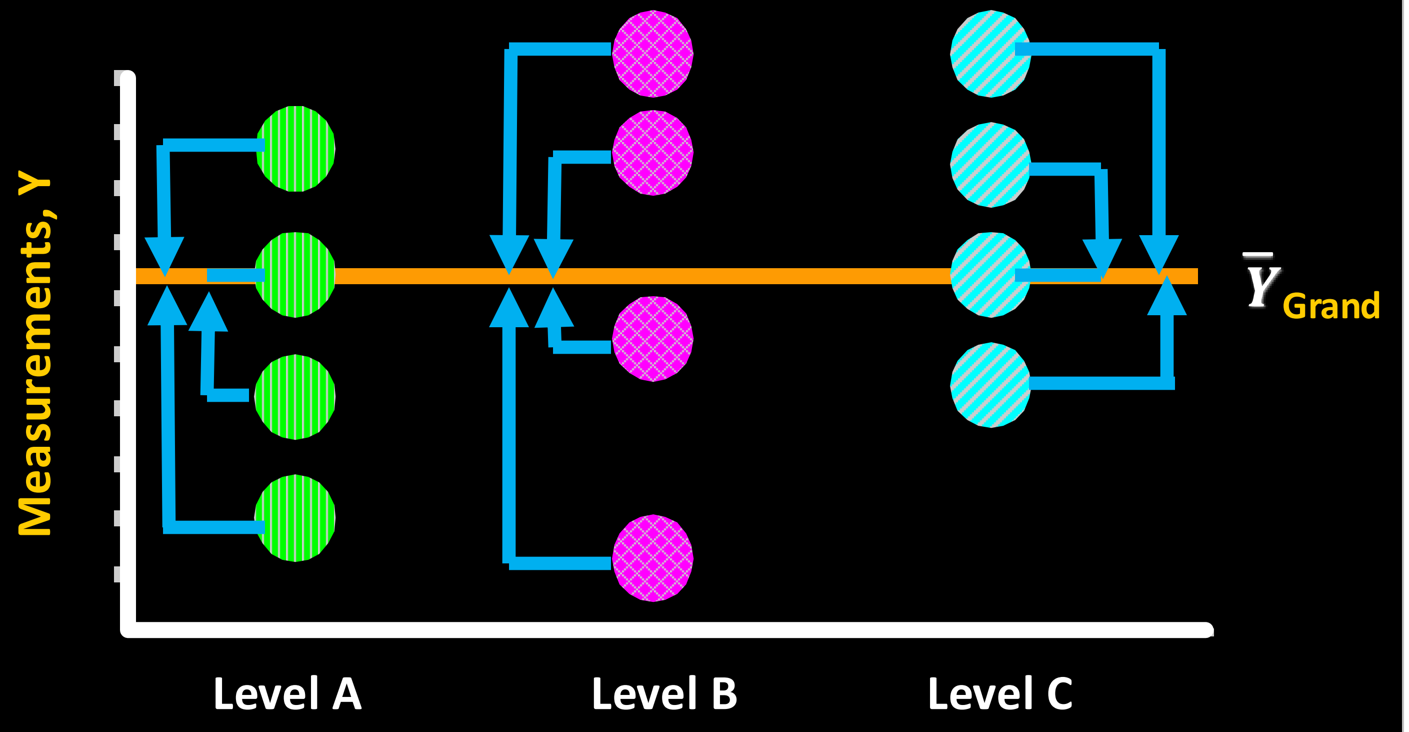
Rational of ANOVA

- Partition total data variation into two sources
  - Between levels/groups (explained by the model)
  - Within levels/groups (remain in the residuals)

- If $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ is true, the standardized variances (**between groups vs. within groups**) are equal to one another

- Ratio of between-group variance and within-group variance follows a F distribution under H_0
  - F test statistic

Grand mean and total deviation



$$\bar{Y}_{grand} = \frac{\sum y_{ij}}{N}$$

↑ $deviate = y_{ij} - \bar{Y}_{grand}$

$\sum (y_{ij} - \bar{Y}_{grand})^2$ = *Sum of Squares Total*
=*SST*

$$s^2 = variance = MS_{total} = \frac{SST}{df_{total}}$$

df_{total}=N-1

$$s = \sqrt{\frac{SST}{df_{total}}}$$

Partitioning Total Variation

- Variation is simply Average Squared Deviations from the Mean

$$\boldsymbol{SST} = SST_{group} + SSE_{residual}$$

$$\sum_{\boldsymbol{j=1}}^{\boldsymbol{K}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n_j}} \left(\boldsymbol{y_{ij}} - \boldsymbol{\bar{Y}}\right)^{\boldsymbol{2}} = \sum_{j=1}^{K} n_j \left(\bar{y}_j - \bar{Y}\right)^2 + \sum_{j=1}^{K} \sum_{i=1}^{n_j} \left(y_{ij} - \bar{y}_j\right)^2$$

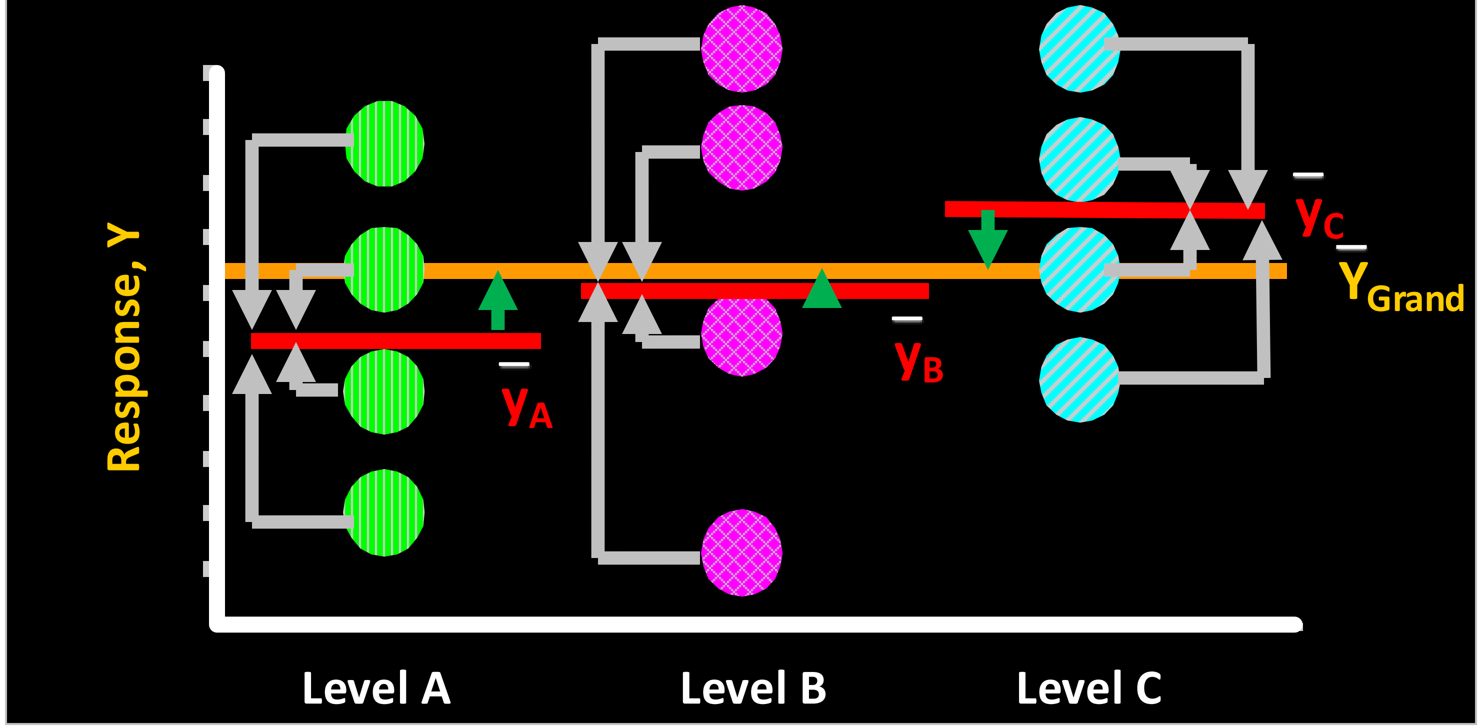
Sum of squared deviations from the grand mean across all N observations

Sum of squared deviations for each group mean from the grand mean

Sum of squared deviations for all observations from each group mean across all K groups

$$N = n_1 + n_2 + \cdots + n_K$$

Group Means: explains model & residual deviation



$$\bar{Y}_{grand} = \frac{\sum y_{ij}}{N}$$

$$group\ means\ \bar{y}_j = \frac{\sum y_i}{n_j}$$

⬆ $group\ deviate = \bar{y}_j - \bar{Y}_{grand}$

⇧ $residual\ deviate = y_{ij} - \bar{y}_j$

$\sum n(\text{⬆})^2$ = *Sum of Squares group*

$$MS_{group} = \frac{SST_{group}}{df_{group}}$$

$\sum(\text{⇧})^2$ = *Error Sum of Squares residual*

$$MSE_{residual} = \frac{SSE_{residual}}{df_{residual}}$$

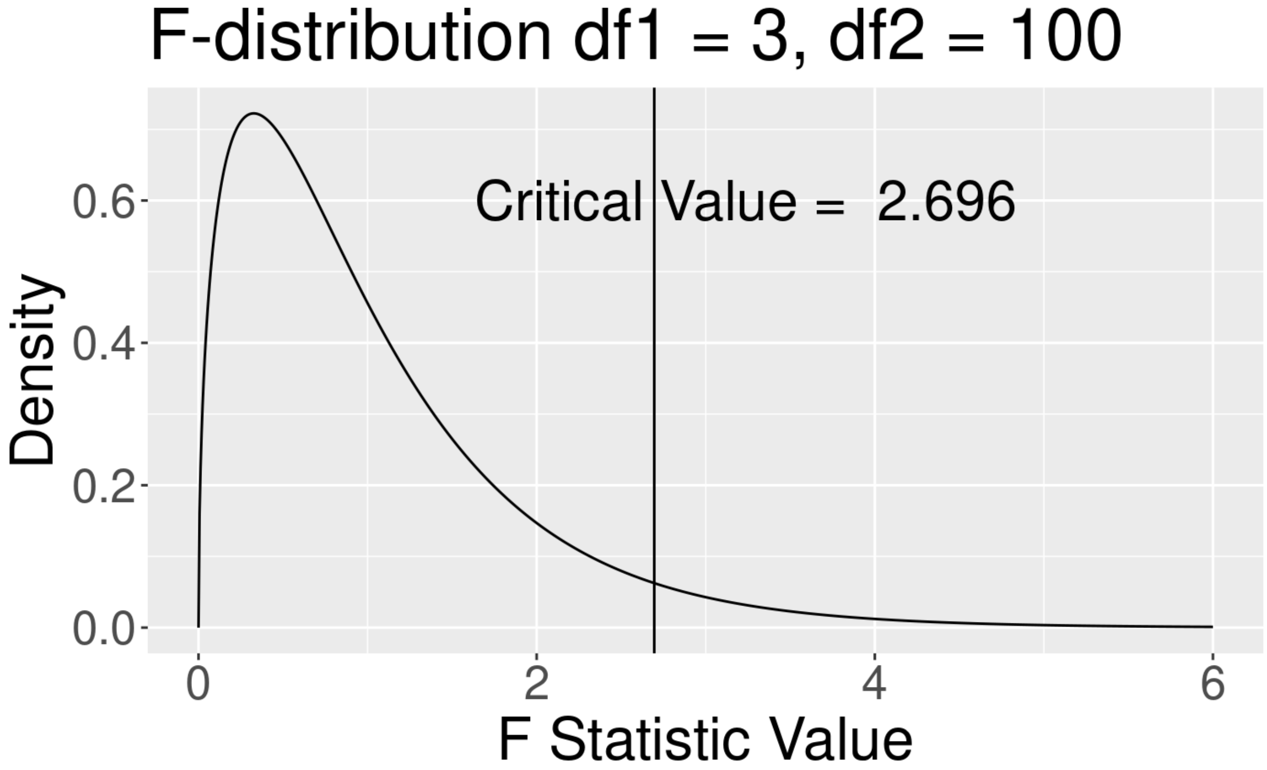
$$F_{K\text{-}1,\ N\text{-}K} \sim \frac{\sum(\text{⬆})^2/df_g = MS_{group}}{\sum(\text{⇧})^2/df_r = MSE_{residual}}$$

$df_g = K\text{-}1, \quad df_r = N\text{-}K$

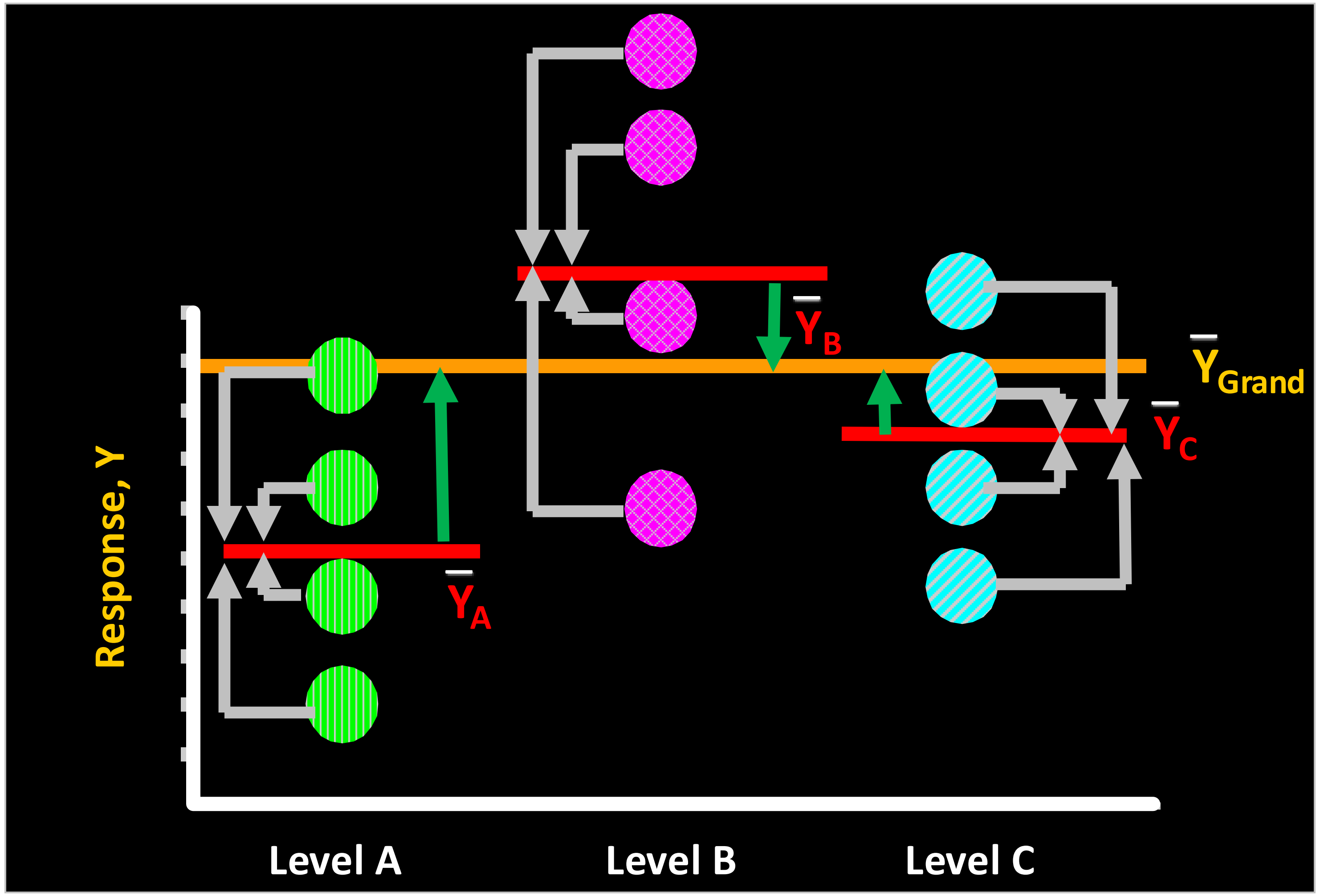
ANOVA F-test

$$F_{df1,df2} = \frac{model\ or\ group\ variance}{residual\ variance} = \frac{MS_{dfg,\,effect}}{MSE_{dfr}}$$

Follows an F-distribution under the NULL hypothesis.



- Null distributions of F(3, 100)
- One-sided critical values of F(3, 100)
  qf(0.95, df1 = 3, df2 = 100)
  2.696
- If our test F is as or more extreme than the critical value, we reject the null hypothesis.



When the factor is associated with the response:

$$F_{K,\,(N-(K+1))} \sim \frac{\sum(\uparrow)^2/df_g = MS_{group} \quad ⬆⬆⬆}{\sum(\uparrow)^2/df_r = MSE_{residual} \quad ▬}$$

is expected to be greater than the corresponding critical value.

Statistical Analysis of an ANOVA design is usually a two-step process

- Step 1: F Test of the **omnibus** null

$$H_o : \sigma^2_{\text{model}} \leq \sigma^2_{\text{residual}}$$

- Step 2: Multiple *post-hoc* comparisons of group means

$$H_o : \mu_A \leq \mu_B \leq \mu_C \ldots, \leq \mu_k$$

ANOVA Table

Source of Variation	df	Sum of Squares	MS	F
Group	k-1	SST_G	$\frac{SST_G}{k-1}$	$\frac{SST_G}{k-1} \Big/ \frac{SST_E}{N-k}$
Error	N-k	SST_E	$\frac{SST_E}{N-k}$	
Total	N-1	SST		

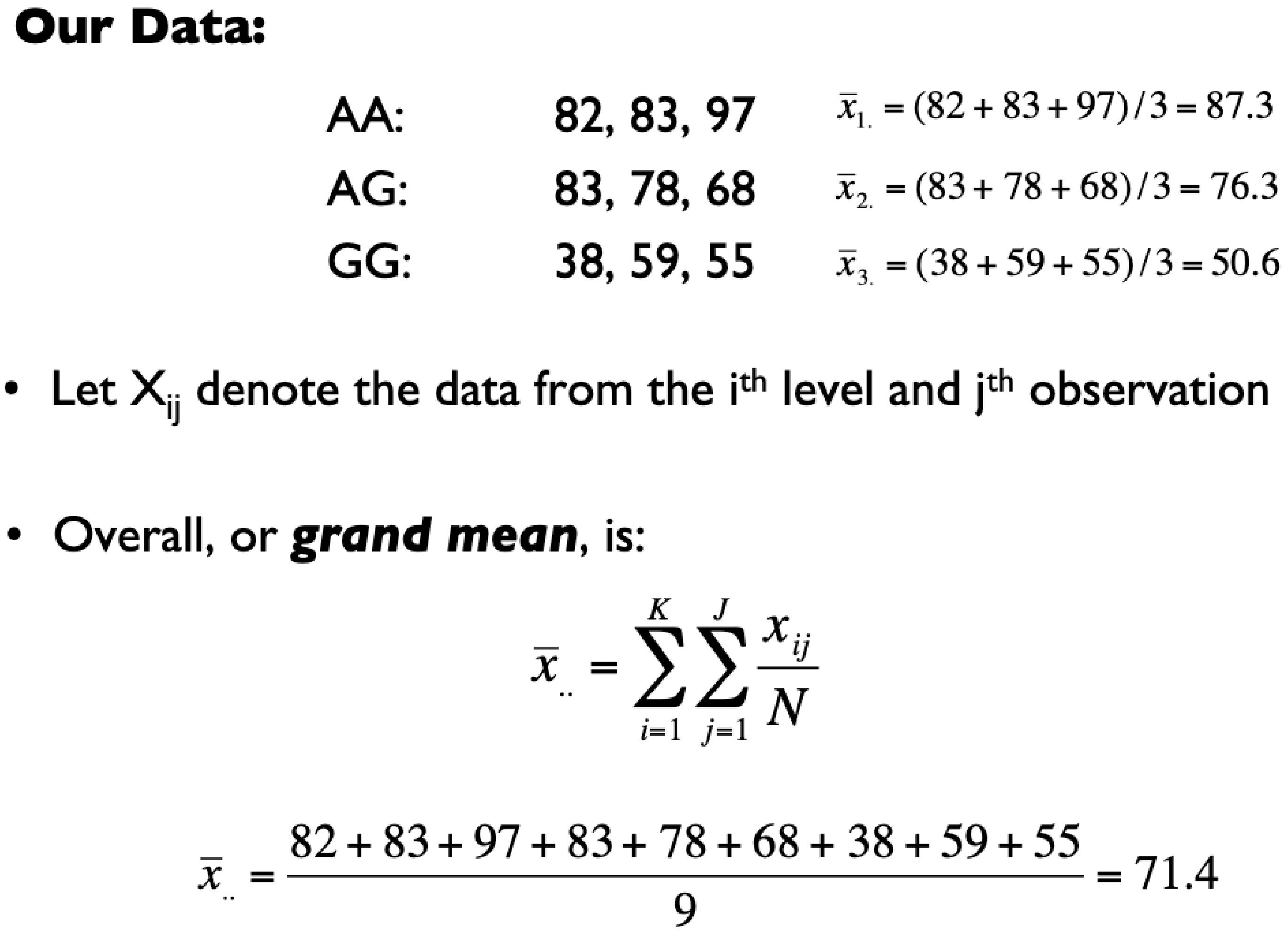
$SST_G = SST_{group}$

$SST_E = SSE_{residual}$

$\eta^2 = \frac{SST_G}{SST_{Total}}$, "ges" generalized eta square in results by ezANOVA()

Equivalent to regression $R^2 = \frac{SSR_{Model}}{SST_{Total}} = 1 - \frac{SSE_{residual}}{SST_{Total}}$

Example dataset: a quantitative trait X was measured, and a single SNP was genotyped



Our Data:

AA: **82, 83, 97** $\bar{x}_{1.} = (82+83+97)/3 = 87.3$

AG: **83, 78, 68** $\bar{x}_{2.} = (83+78+68)/3 = 76.3$

GG: **38, 59, 55** $\bar{x}_{3.} = (38+59+55)/3 = 50.6$

- Let X_{ij} denote the data from the ith level and jth observation

- Overall, or ***grand mean***, is:

$$\bar{x}_{..} = \sum_{i=1}^{K} \sum_{j=1}^{J} \frac{x_{ij}}{N}$$

$$\bar{x}_{..} = \frac{82+83+97+83+78+68+38+59+55}{9} = 71.4$$

(X is the continuous response variable Y in the previous slide)

Partitioning Total Variation

- $SST_G = SST_{group}$
- $SST_E = SSE_{residual}$

$$SST = SST_G + SST_E$$

$$\sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij}-\bar{x}_{..})^2$$

$$\downarrow$$

$$(82-71.4)^2+(83-71.4)^2+(97-71.4)^2+$$
$$(83-71.4)^2+(78-71.4)^2+(68-71.4)^2+$$
$$(38-71.4)^2+(59-71.4)^2+(55-71.4)^2=$$

2630.2

$$\sum_{i=1}^{K} n_i \bullet (\bar{x}_{i.}-\bar{x}_{..})^2$$

$$\downarrow$$

$$3\bullet(87.3-71.4)^2+$$
$$3\bullet(76.3-71.4)^2+$$
$$3\bullet(50.6-71.4)^2=$$

2124.2

$$\sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij}-\bar{x}_{i.})^2$$

$$\downarrow$$

$$(82-87.3)^2+(83-87.3)^2+(97-87.3)^2+$$
$$(83-76.3)^2+(78-76.3)^2+(68-76.3)^2+$$
$$(38-50.6)^2+(59-50.6)^2+(55-50.6)^2=$$

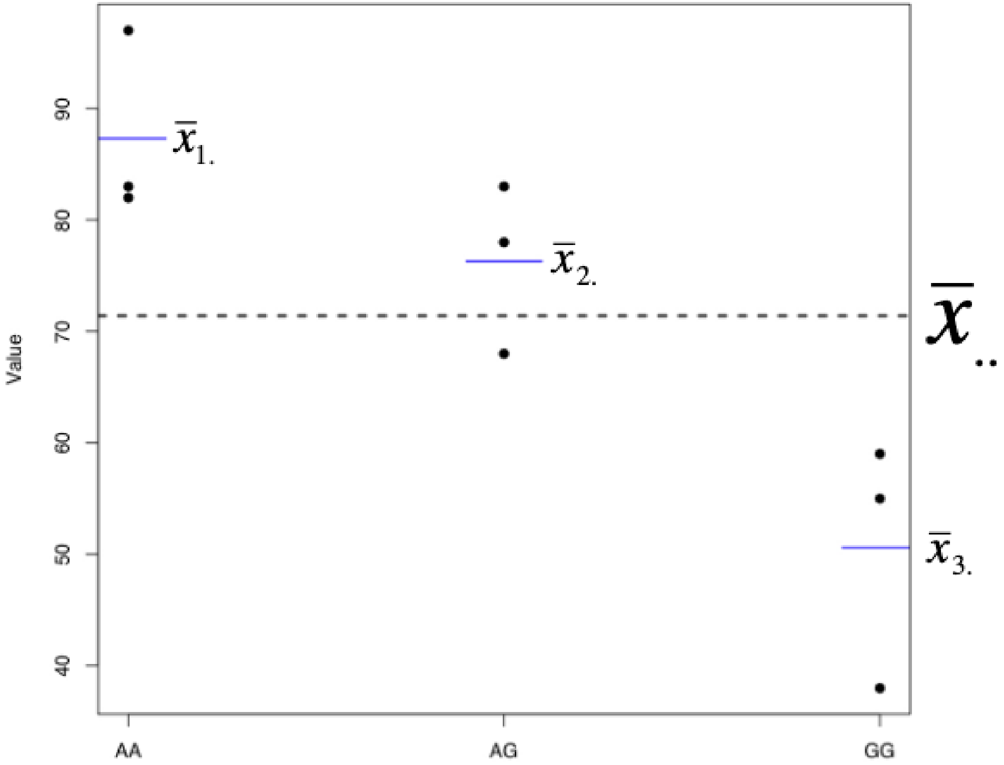
506

Partitioning Total Variation

$$\text{SST} = \text{SST}_G + \text{SST}_E$$

$$\sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{..})^2 \qquad \sum_{i=1}^{K} n_i \bullet (\bar{x}_{i.} - \bar{x}_{..})^2 \qquad \sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{i.})^2$$

- $\text{SST}_G = \text{SST}_{group}$
- $\text{SST}_E = \text{SSE}_{residual}$



Calculating Mean Squares

- To make the sum of squares comparable, we divide each one by their associated degrees of freedom
    - SST_G = k - 1 (3 - 1 = 2)
    - SST_E = N - k (9 - 3 = 6)
    - SST_T = N - 1 (9 - 1 = 8)

- MST_G = 2124.2 / 2 = 1062.1

- MST_E = 506 / 6 = 84.3

$MST_G = MS_{group}$

$MST_E = MSE_{residual}$

Almost There… Calculating F Statistic

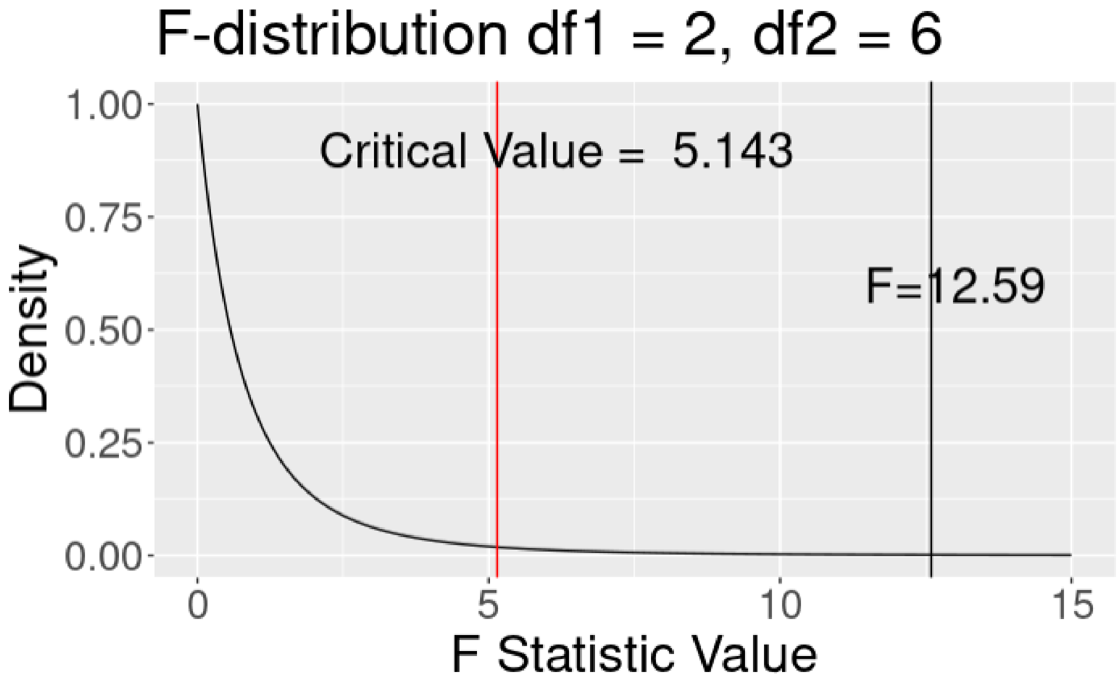
- The test statistic is the ratio of group and error mean squares

$$F = \frac{MST_G}{MST_E} = \frac{1062.2}{84.3} = 12.59$$

- If H_0 is true MST_G and MST_E are equal

- Critical value for rejection region is $F_{\alpha,\ k-1,\ N-k}$

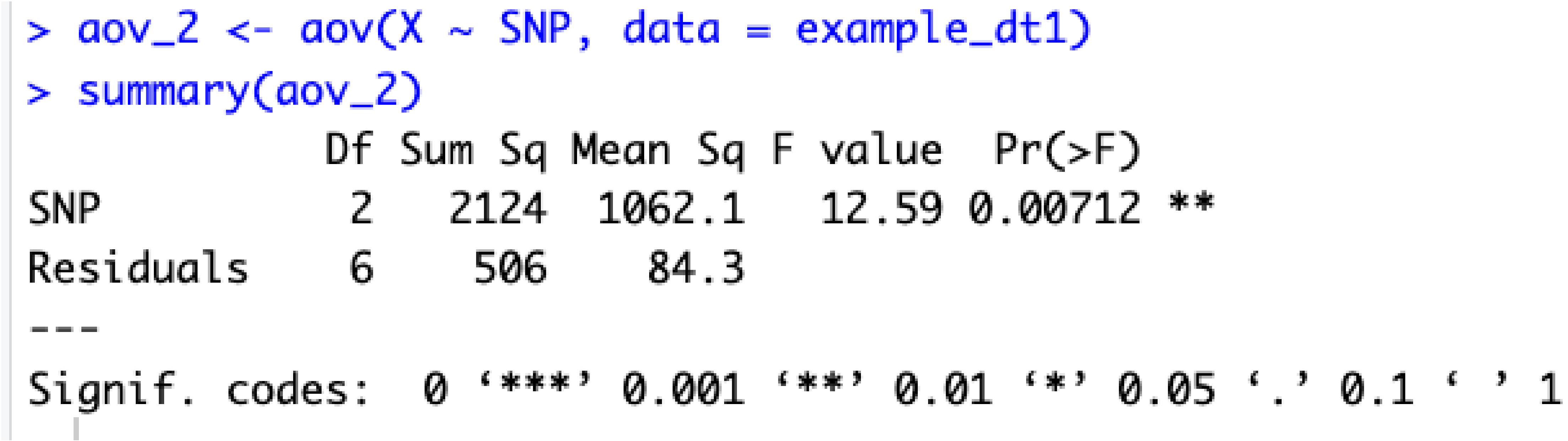
- If we define $\alpha = 0.05$, then $F_{0.05,\ 2,\ 6} = 5.14$



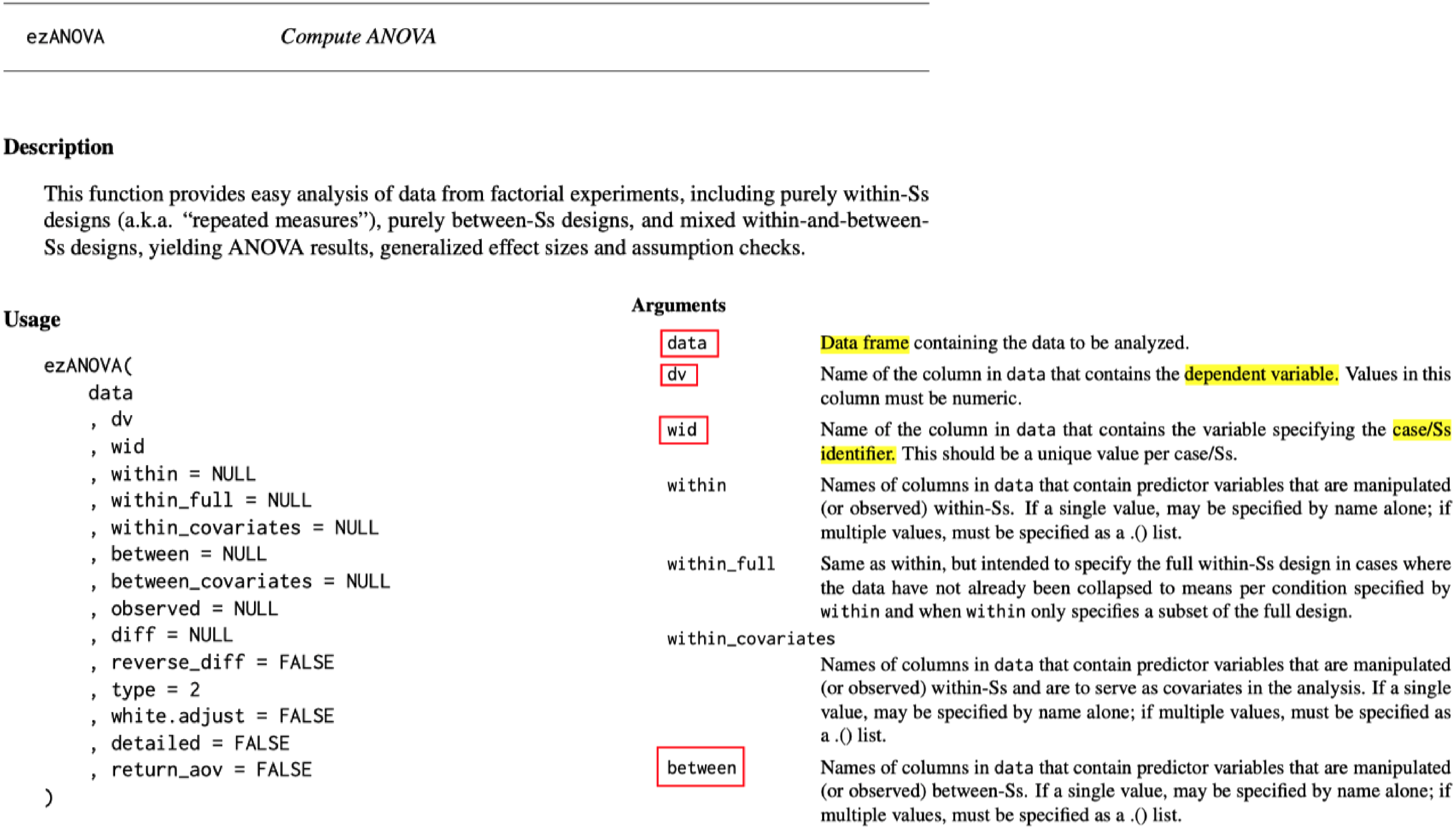
How to do ANOVA analysis in R?

- Base R function : **aov()**
- R function: **ezANOVA()** from R library “ez”

One-way ANOVA by aov() with Completely Randomized Samples



```
> aov_2 <- aov(X ~ SNP, data = example_dt1)
> summary(aov_2)
            Df Sum Sq Mean Sq F value  Pr(>F)   
SNP          2   2124  1062.1   12.59 0.00712 **
Residuals    6    506    84.3                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



`ezANOVA` *Compute ANOVA*

Description

This function provides easy analysis of data from factorial experiments, including purely within-Ss designs (a.k.a. "repeated measures"), purely between-Ss designs, and mixed within-and-between-Ss designs, yielding ANOVA results, generalized effect sizes and assumption checks.

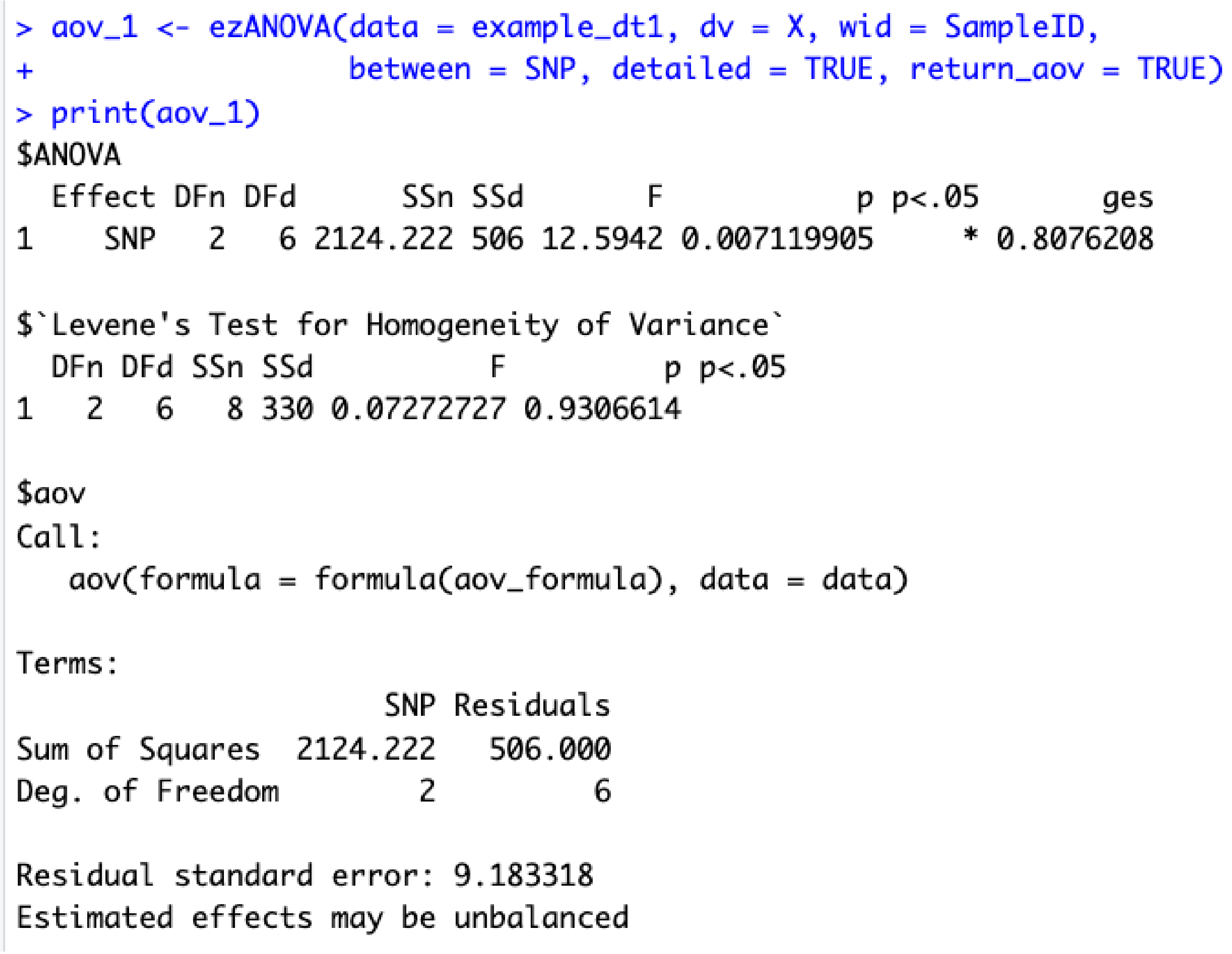
Usage

```
ezANOVA(
    data
    , dv
    , wid
    , within = NULL
    , within_full = NULL
    , within_covariates = NULL
    , between = NULL
    , between_covariates = NULL
    , observed = NULL
    , diff = NULL
    , reverse_diff = FALSE
    , type = 2
    , white.adjust = FALSE
    , detailed = FALSE
    , return_aov = FALSE
)
```

Arguments

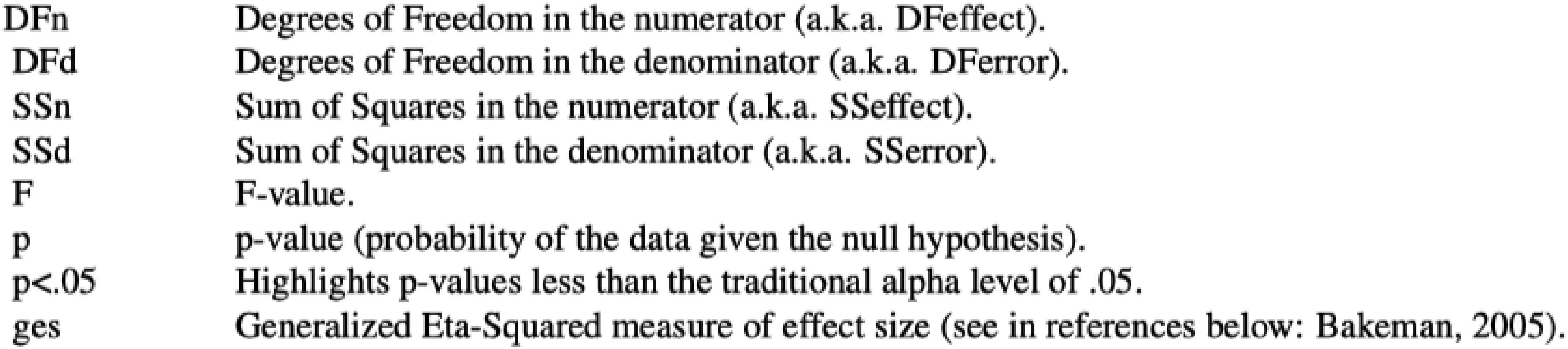
`data`	Data frame containing the data to be analyzed.
`dv`	Name of the column in `data` that contains the dependent variable. Values in this column must be numeric.
`wid`	Name of the column in `data` that contains the variable specifying the case/Ss identifier. This should be a unique value per case/Ss.
`within`	Names of columns in `data` that contain predictor variables that are manipulated (or observed) within-Ss. If a single value, may be specified by name alone; if multiple values, must be specified as a .() list.
`within_full`	Same as within, but intended to specify the full within-Ss design in cases where the data have not already been collapsed to means per condition specified by `within` and when `within` only specifies a subset of the full design.
`within_covariates`	Names of columns in `data` that contain predictor variables that are manipulated (or observed) within-Ss and are to serve as covariates in the analysis. If a single value, may be specified by name alone; if multiple values, must be specified as a .() list.
`between`	Names of columns in `data` that contain predictor variables that are manipulated (or observed) between-Ss. If a single value, may be specified by name alone; if multiple values, must be specified as a .() list.

One-way ANOVA by ezANOVA() with Completely Randomized Samples



```
> aov_1 <- ezANOVA(data = example_dt1, dv = X, wid = SampleID,
+                   between = SNP, detailed = TRUE, return_aov = TRUE)
> print(aov_1)
$ANOVA
  Effect DFn DFd      SSn SSd       F           p p<.05       ges
1    SNP   2   6 2124.222 506 12.5942 0.007119905     * 0.8076208

$`Levene's Test for Homogeneity of Variance`
  DFn DFd SSn SSd          F         p p<.05
1   2   6   8 330 0.07272727 0.9306614

$aov
Call:
   aov(formula = formula(aov_formula), data = data)

Terms:
                     SNP Residuals
Sum of Squares  2124.222   506.000
Deg. of Freedom        2         6

Residual standard error: 9.183318
Estimated effects may be unbalanced
```

Output Variables by ezANOVA():



DFn	Degrees of Freedom in the numerator (a.k.a. DFeffect).
DFd	Degrees of Freedom in the denominator (a.k.a. DFerror).
SSn	Sum of Squares in the numerator (a.k.a. SSeffect).
SSd	Sum of Squares in the denominator (a.k.a. SSerror).
F	F-value.
p	p-value (probability of the data given the null hypothesis).
p<.05	Highlights p-values less than the traditional alpha level of .05.
ges	Generalized Eta-Squared measure of effect size (see in references below: Bakeman, 2005).

Connection with Linear Regression

Predicted and Residual Values

- ***Predicted***, or fitted, values are values of y predicted by the least-squares regression line obtained by plugging in $x_1, x_2, \ldots, x_n$ into the estimated regression line

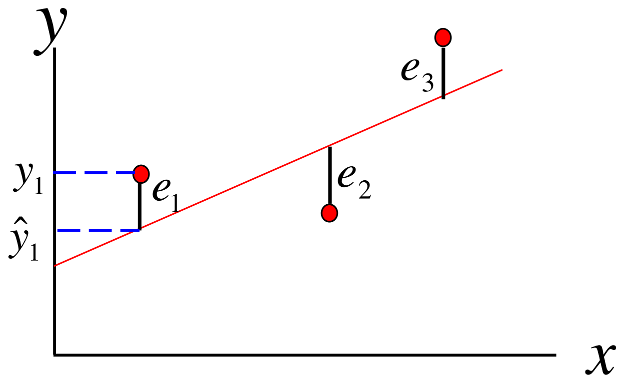
$$\hat{y}_1 = \hat{\beta}_0 - \hat{\beta}_1 x_1$$

$$\hat{y}_2 = \hat{\beta}_0 - \hat{\beta}_1 x_2$$

- ***Residuals*** are the deviations of observed and predicted values

$$e_1 = y_1 - \hat{y}_1$$

$$e_2 = y_2 - \hat{y}_2$$



Residuals Are Useful!

- They allow us to calculate the error sum of squares (SSE):

$$SSE = \sum_{i=1}^{n} (e_i)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- Which in turn allows us to estimate σ^2:

$$\hat{\sigma}^2 = \frac{SSE}{n-2}$$

n is Sample Size

Degree of freedom n-2 for having 1 predictor.

Will be n-(k+1) for having k predictors in the linear regression model.

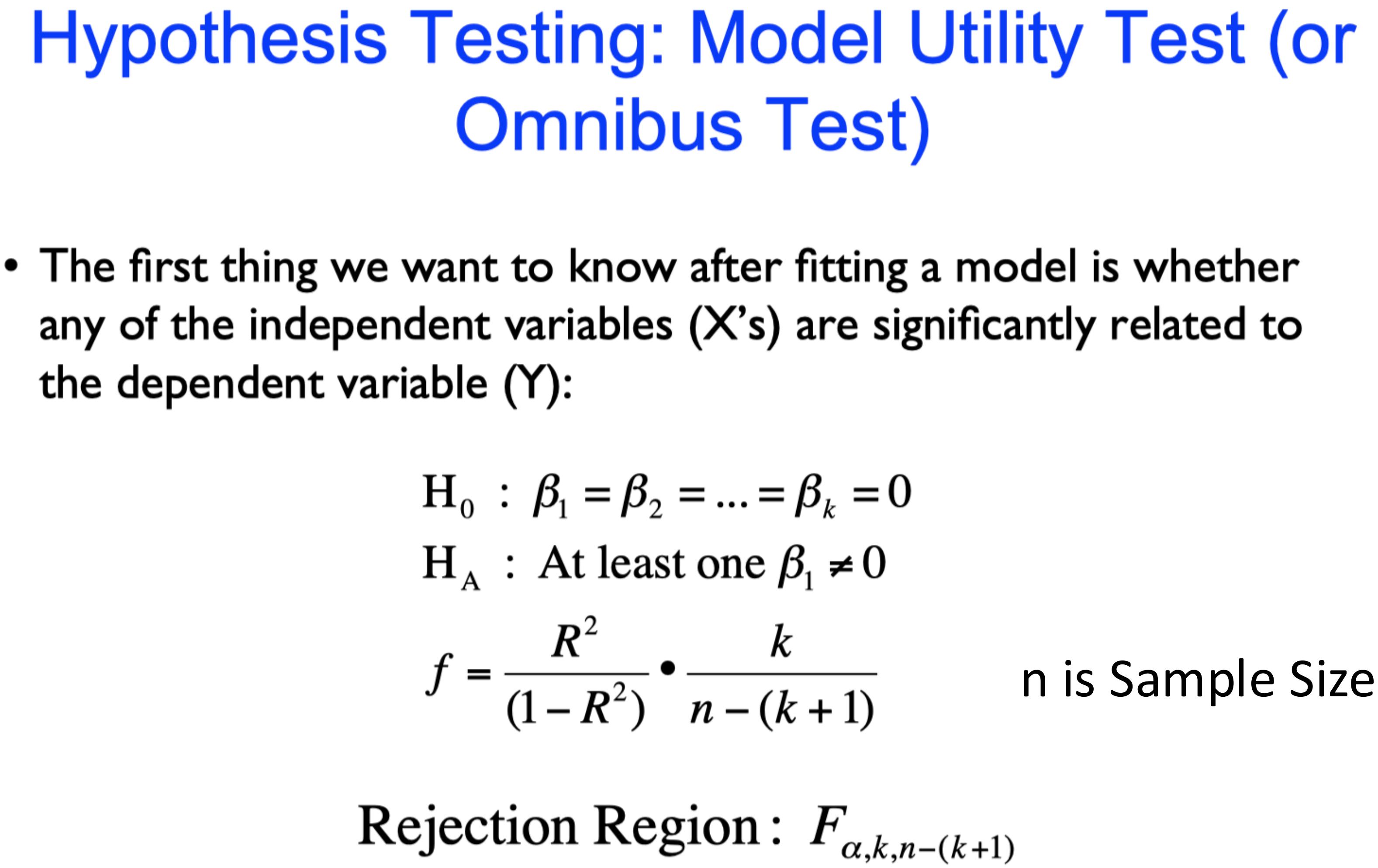
- As well as an important statistic referred to as the coefficient of determination:

$$r^2 = 1 - \frac{SSE}{SST} \qquad SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Aka. Regression R^2

Multivariate Linear Regression

- Linear regression model to two or more independent variables (i.e., multiple predictors)
- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$



Hypothesis Testing: Model Utility Test (or Omnibus Test)

- The first thing we want to know after fitting a model is whether any of the independent variables (X's) are significantly related to the dependent variable (Y):

$$\mathrm{H}_0 \;:\; \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

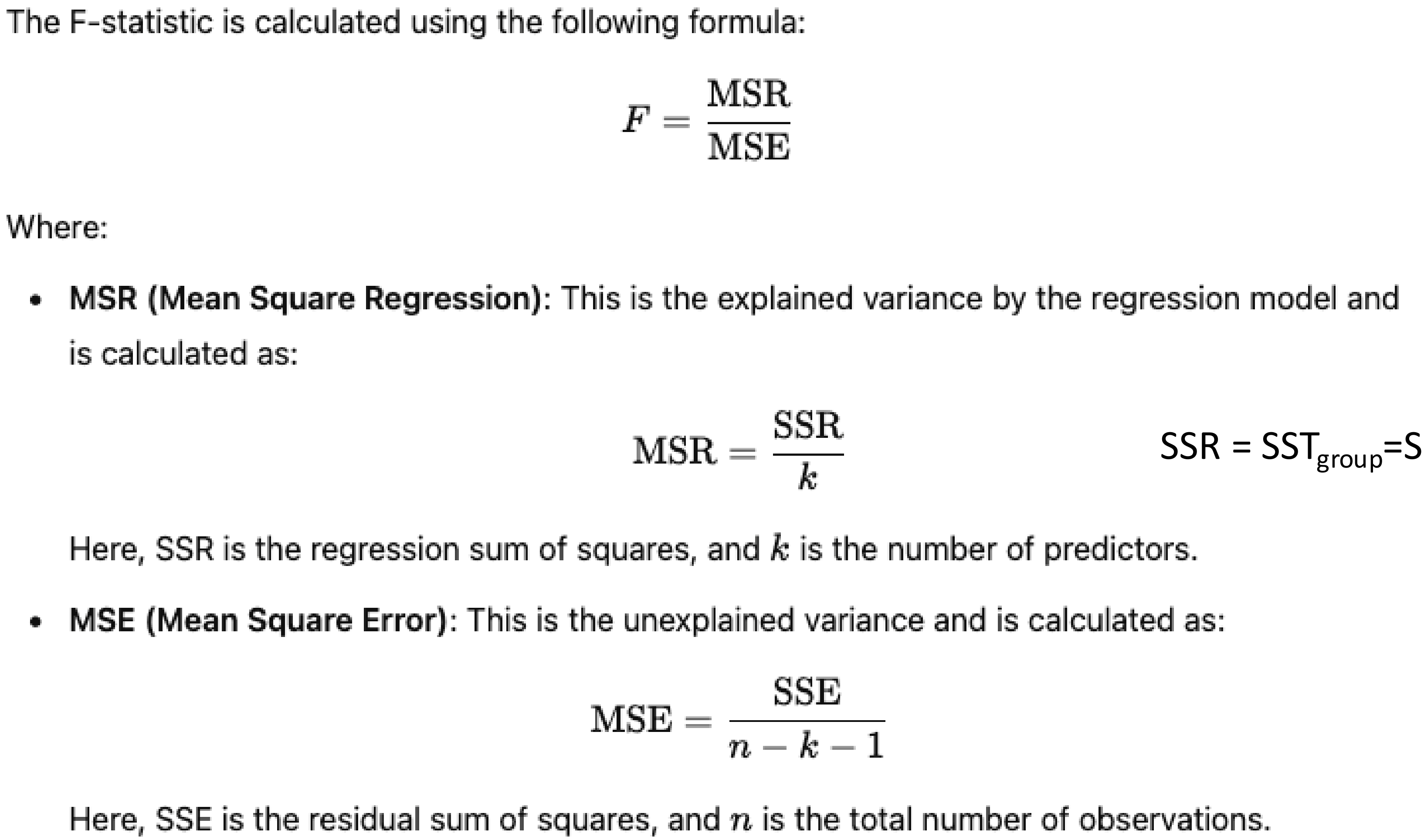
$$\mathrm{H}_\mathrm{A} \;:\; \text{At least one } \beta_1 \neq 0$$

$$f = \frac{R^2}{(1-R^2)} \bullet \frac{k}{n-(k+1)}$$

n is Sample Size

$$\text{Rejection Region}: \; F_{\alpha,k,n-(k+1)}$$

F Statistic in Regression Analysis



The F-statistic is calculated using the following formula:

$$F = \frac{\mathrm{MSR}}{\mathrm{MSE}}$$

Where:

- **MSR (Mean Square Regression)**: This is the explained variance by the regression model and is calculated as:

$$\mathrm{MSR} = \frac{\mathrm{SSR}}{k}$$

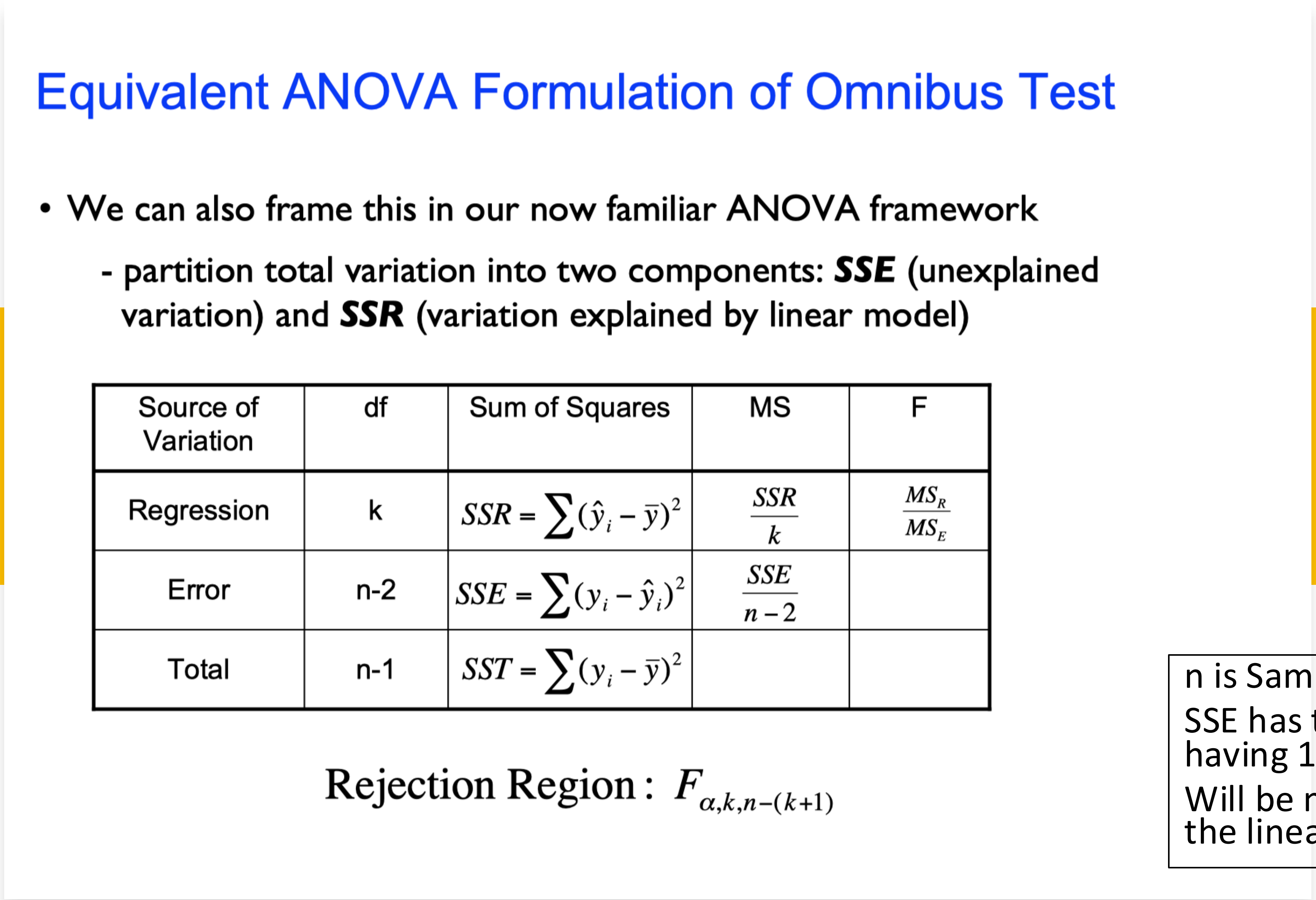
  Here, SSR is the regression sum of squares, and k is the number of predictors.

- **MSE (Mean Square Error)**: This is the unexplained variance and is calculated as:

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{n - k - 1}$$

  Here, SSE is the residual sum of squares, and n is the total number of observations.

$\mathrm{SSR} = \mathrm{SST}_{\mathrm{group}} = \mathrm{SST}_{\mathrm{model}}$



Equivalent ANOVA Formulation of Omnibus Test

- We can also frame this in our now familiar ANOVA framework
  - partition total variation into two components: ***SSE*** (unexplained variation) and ***SSR*** (variation explained by linear model)

Source of Variation	df	Sum of Squares	MS	F
Regression	k	$SSR = \sum(\hat{y}_i - \bar{y})^2$	$\frac{SSR}{k}$	$\frac{MS_R}{MS_E}$
Error	n-2	$SSE = \sum(y_i - \hat{y}_i)^2$	$\frac{SSE}{n-2}$	
Total	n-1	$SST = \sum(y_i - \bar{y})^2$		

$$\text{Rejection Region}: F_{\alpha,k,n-(k+1)}$$

$SSR = SST_{group} = SST_{model}$

n is Sample Size
SSE has the degree of freedom n-2 for having 1 predictor.
Will be n-(k+1) for having k predictors in the linear regression model.

F Test For Subsets of Independent Variables

- A powerful tool in multiple regression analyses is the ability to compare two models

- For instance say we want to compare:

$$\text{Full Model}: \; y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

$$\text{Reduced Model}: \; y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

- Again, another example of ANOVA:

SSE_R = error sum of squares for reduced model with l predictors

SSE_F = error sum of squares for full model with k predictors

$$f = \frac{(SSE_R - SSE_F)/(k-l)}{SSE_F/([n-(k+1)]}$$

n is Sample Size

Example of Model Comparison

- We have a quantitative trait and want to test the effects at two markers, M1 and M2.

Full Model: Trait = Mean + M1 + M2 + (M1*M2) + error

Reduced Model: Trait = Mean + M1 + M2 + error

$$f = \frac{(SSE_R - SSE_F)/(3-2)}{SSE_F/([100-(3+1)]} = \frac{(SSE_R - SSE_F)}{SSE_F/96}$$

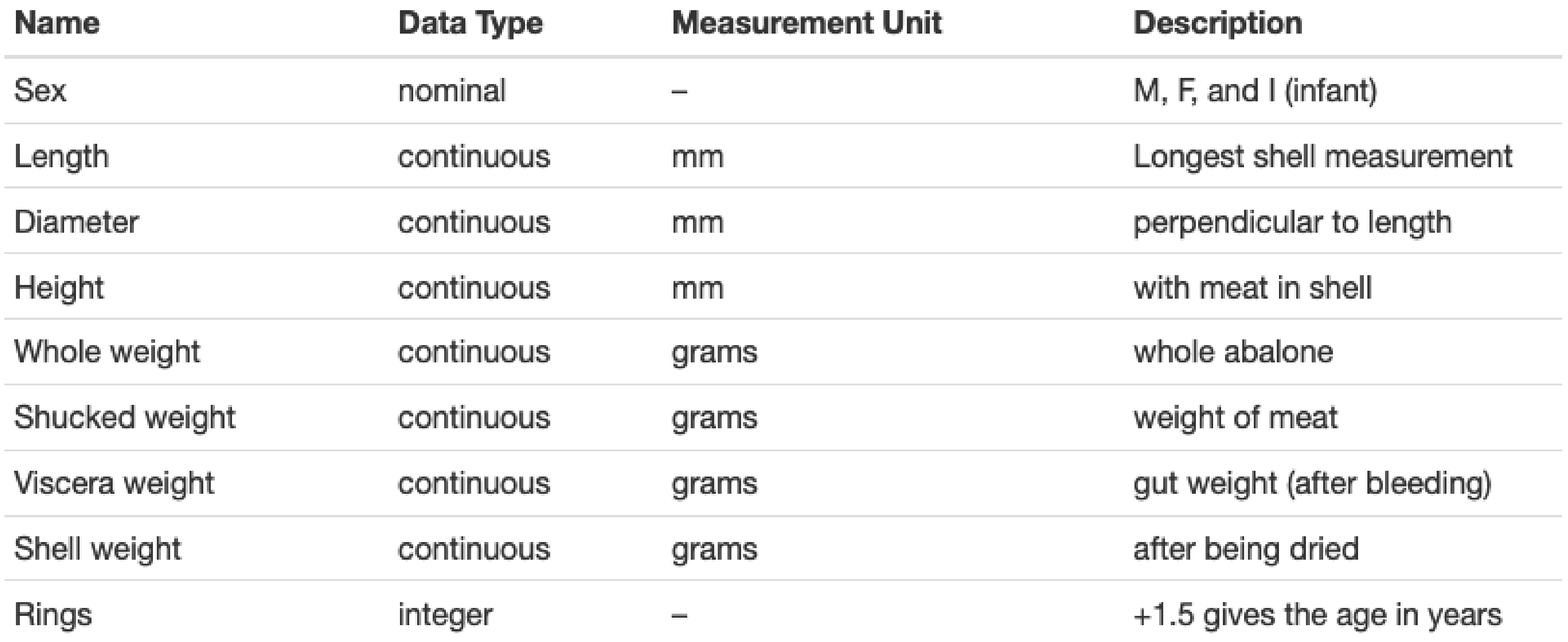
$$\text{Rejection Region}: \; F_{a,\,1,\,96}$$

Model 1. Rings/Age ~ factor(sex) + length + diameter + height + wholeWeight + shuckedWeight + visceraWeight + shellWeight + diameter * height

vs.

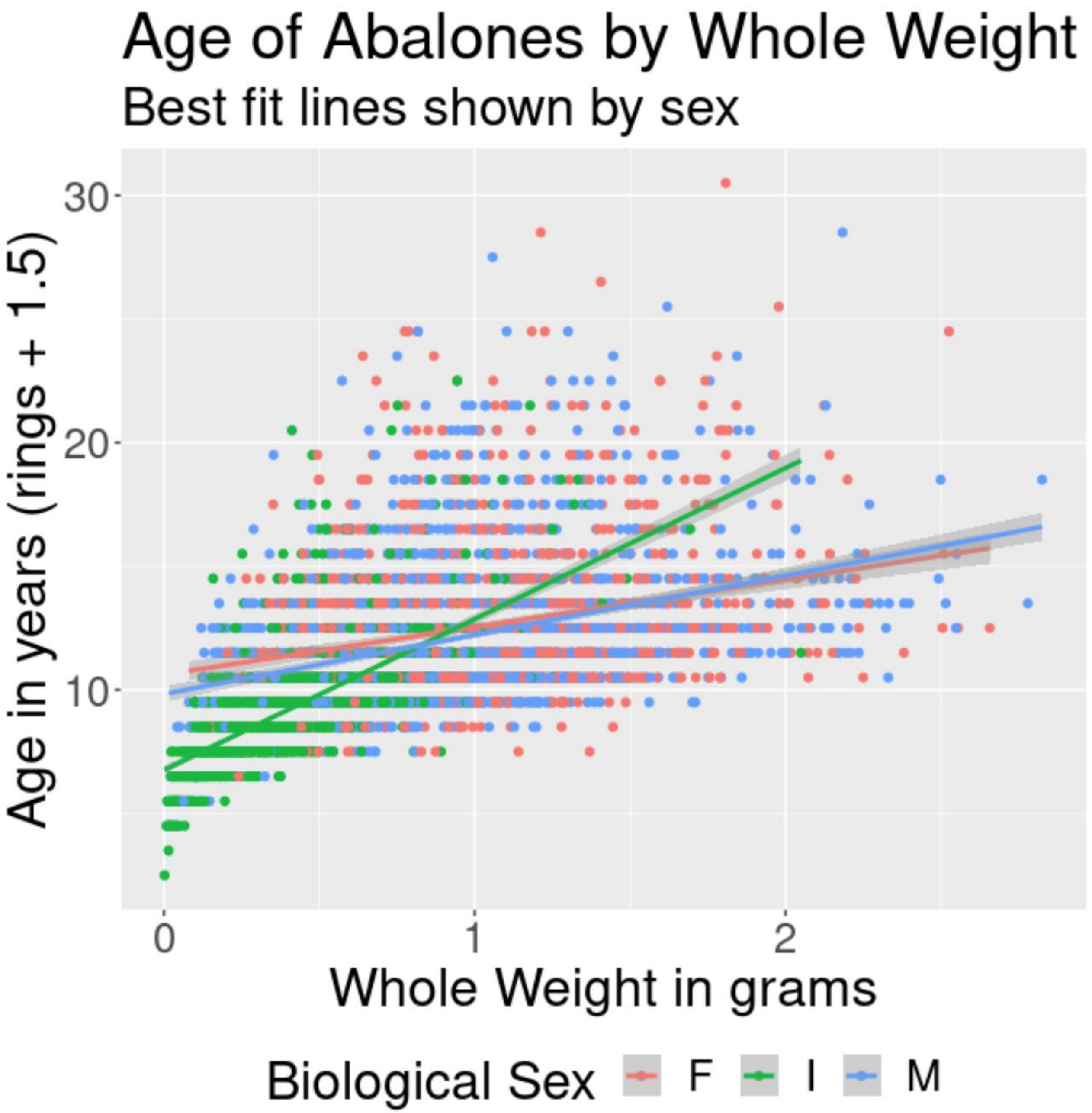
Model 2. Rings/Age ~ factor(sex) + length + diameter + height + wholeWeight + shuckedWeight + visceraWeight + shellWeight

Abalones Dataset

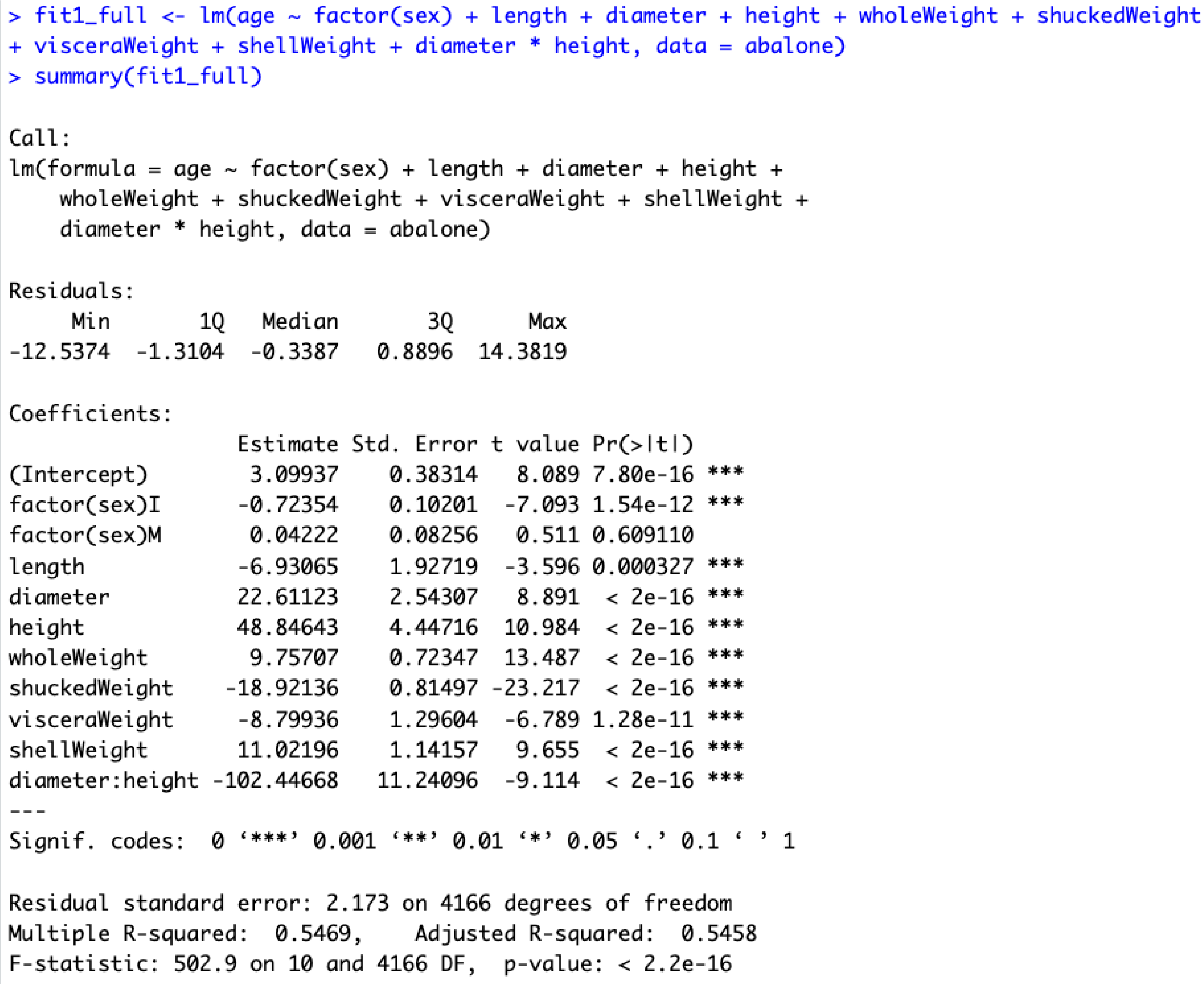


Name	Data Type	Measurement Unit	Description
Sex	nominal	–	M, F, and I (infant)
Length	continuous	mm	Longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)
Shell weight	continuous	grams	after being dried
Rings	integer	–	+1.5 gives the age in years

Relationship between Abalone age/rings and Whole Weight

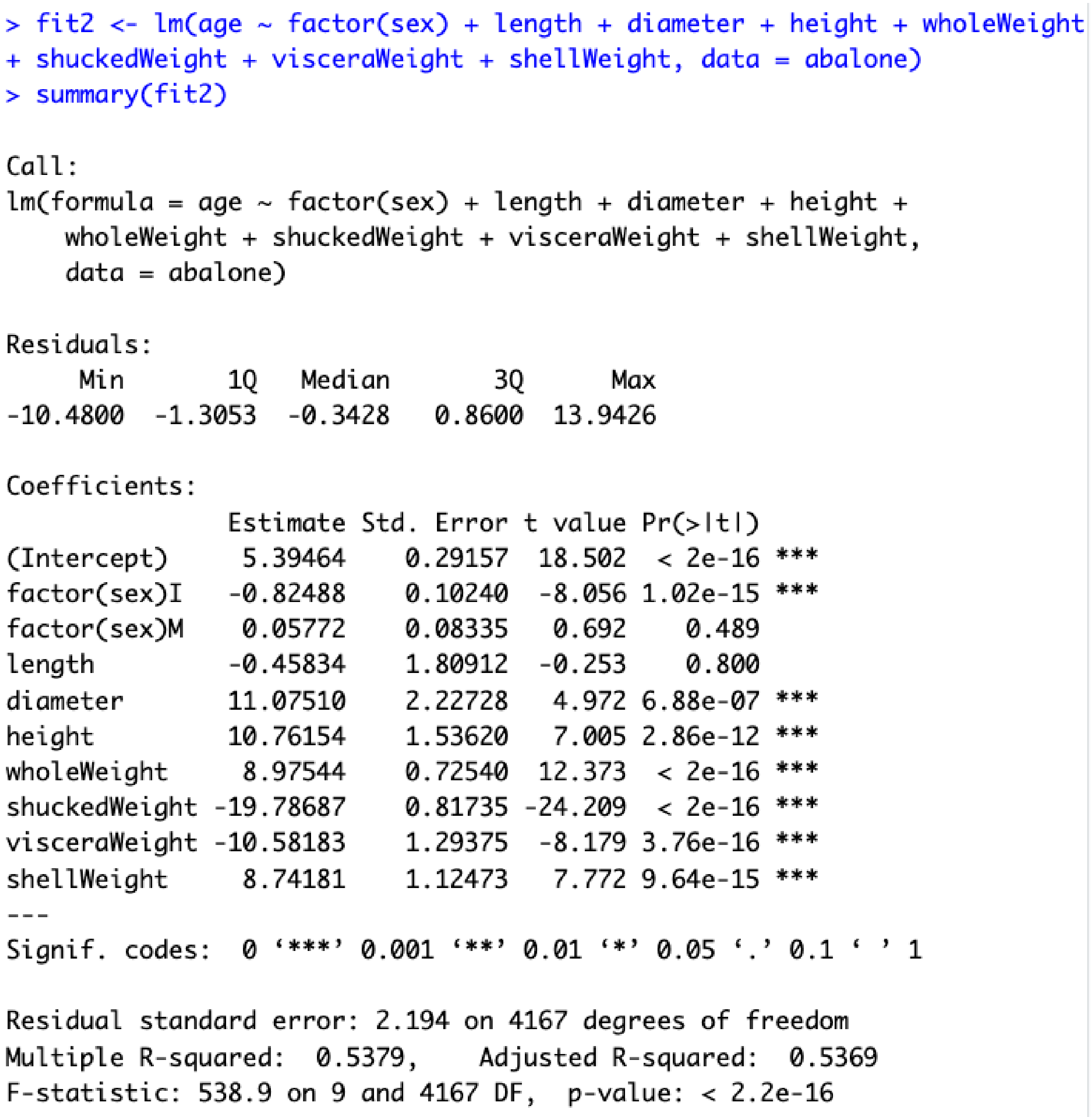


Fit the full model: Model 1



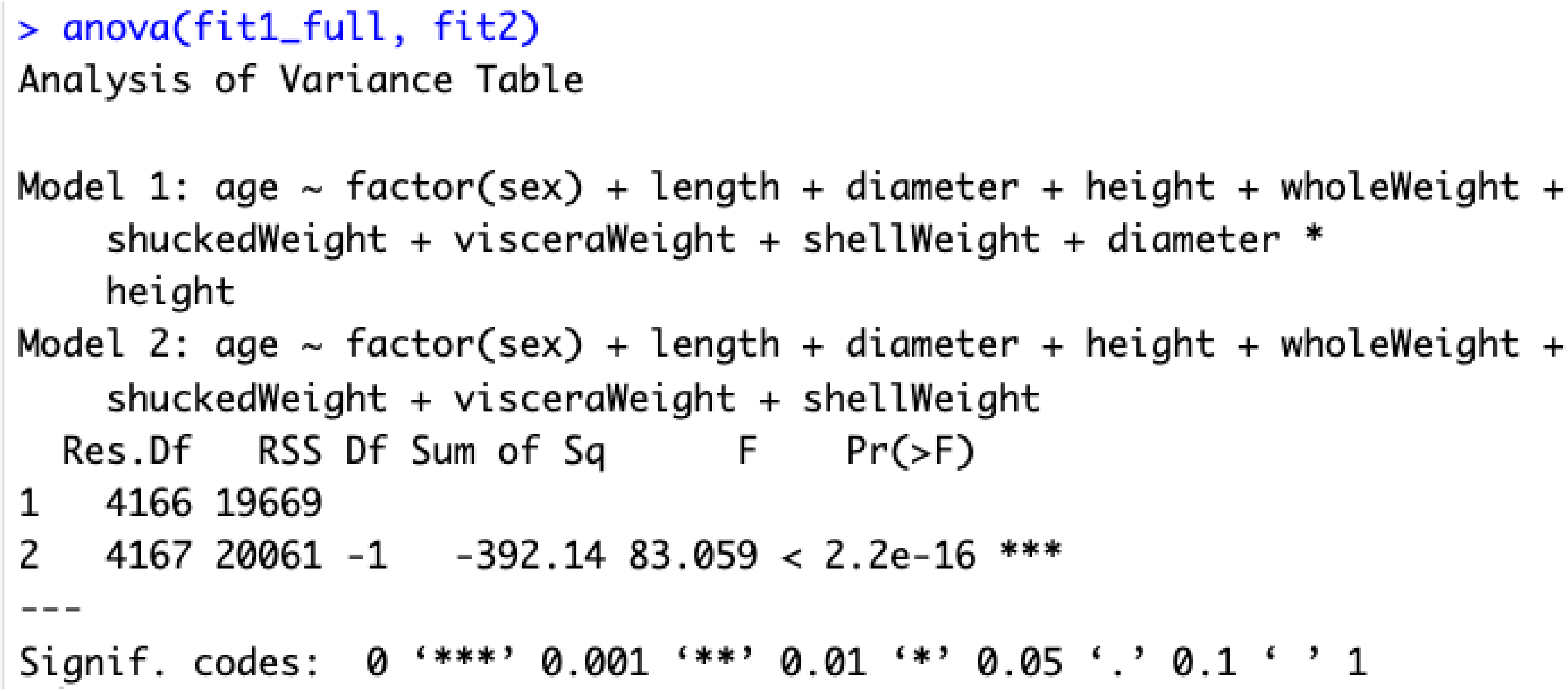
```
> fit1_full <- lm(age ~ factor(sex) + length + diameter + height + wholeWeight + shuckedWeight
+ visceraWeight + shellWeight + diameter * height, data = abalone)
> summary(fit1_full)

Call:
lm(formula = age ~ factor(sex) + length + diameter + height + 
    wholeWeight + shuckedWeight + visceraWeight + shellWeight + 
    diameter * height, data = abalone)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.5374  -1.3104  -0.3387   0.8896  14.3819 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        3.09937    0.38314   8.089 7.80e-16 ***
factor(sex)I      -0.72354    0.10201  -7.093 1.54e-12 ***
factor(sex)M       0.04222    0.08256   0.511 0.609110    
length            -6.93065    1.92719  -3.596 0.000327 ***
diameter          22.61123    2.54307   8.891  < 2e-16 ***
height            48.84643    4.44716  10.984  < 2e-16 ***
wholeWeight        9.75707    0.72347  13.487  < 2e-16 ***
shuckedWeight    -18.92136    0.81497 -23.217  < 2e-16 ***
visceraWeight     -8.79936    1.29604  -6.789 1.28e-11 ***
shellWeight       11.02196    1.14157   9.655  < 2e-16 ***
diameter:height -102.44668   11.24096  -9.114  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.173 on 4166 degrees of freedom
Multiple R-squared:  0.5469,    Adjusted R-squared:  0.5458 
F-statistic: 502.9 on 10 and 4166 DF,  p-value: < 2.2e-16
```

Fit the subset model: Model 2



```
> fit2 <- lm(age ~ factor(sex) + length + diameter + height + wholeWeight
+ shuckedWeight + visceraWeight + shellWeight, data = abalone)
> summary(fit2)

Call:
lm(formula = age ~ factor(sex) + length + diameter + height + 
    wholeWeight + shuckedWeight + visceraWeight + shellWeight, 
    data = abalone)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.4800  -1.3053  -0.3428   0.8600  13.9426 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     5.39464    0.29157  18.502  < 2e-16 ***
factor(sex)I   -0.82488    0.10240  -8.056 1.02e-15 ***
factor(sex)M    0.05772    0.08335   0.692    0.489    
length         -0.45834    1.80912  -0.253    0.800    
diameter       11.07510    2.22728   4.972 6.88e-07 ***
height         10.76154    1.53620   7.005 2.86e-12 ***
wholeWeight     8.97544    0.72540  12.373  < 2e-16 ***
shuckedWeight -19.78687    0.81735 -24.209  < 2e-16 ***
visceraWeight -10.58183    1.29375  -8.179 3.76e-16 ***
shellWeight     8.74181    1.12473   7.772 9.64e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.194 on 4167 degrees of freedom
Multiple R-squared:  0.5379,	Adjusted R-squared:  0.5369 
F-statistic: 538.9 on 9 and 4167 DF,  p-value: < 2.2e-16
```

Model comparison by anova()



```
> anova(fit1_full, fit2)
Analysis of Variance Table

Model 1: age ~ factor(sex) + length + diameter + height + wholeWeight +
    shuckedWeight + visceraWeight + shellWeight + diameter *
    height
Model 2: age ~ factor(sex) + length + diameter + height + wholeWeight +
    shuckedWeight + visceraWeight + shellWeight
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1   4166 19669
2   4167 20061 -1   -392.14 83.059 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Conclusion: stay with the full model

In-Class Exercise 1 : One-way ANOVA

In-class Participation Credit

Complete Task 3 in Exercise 1. Show your work for credits.

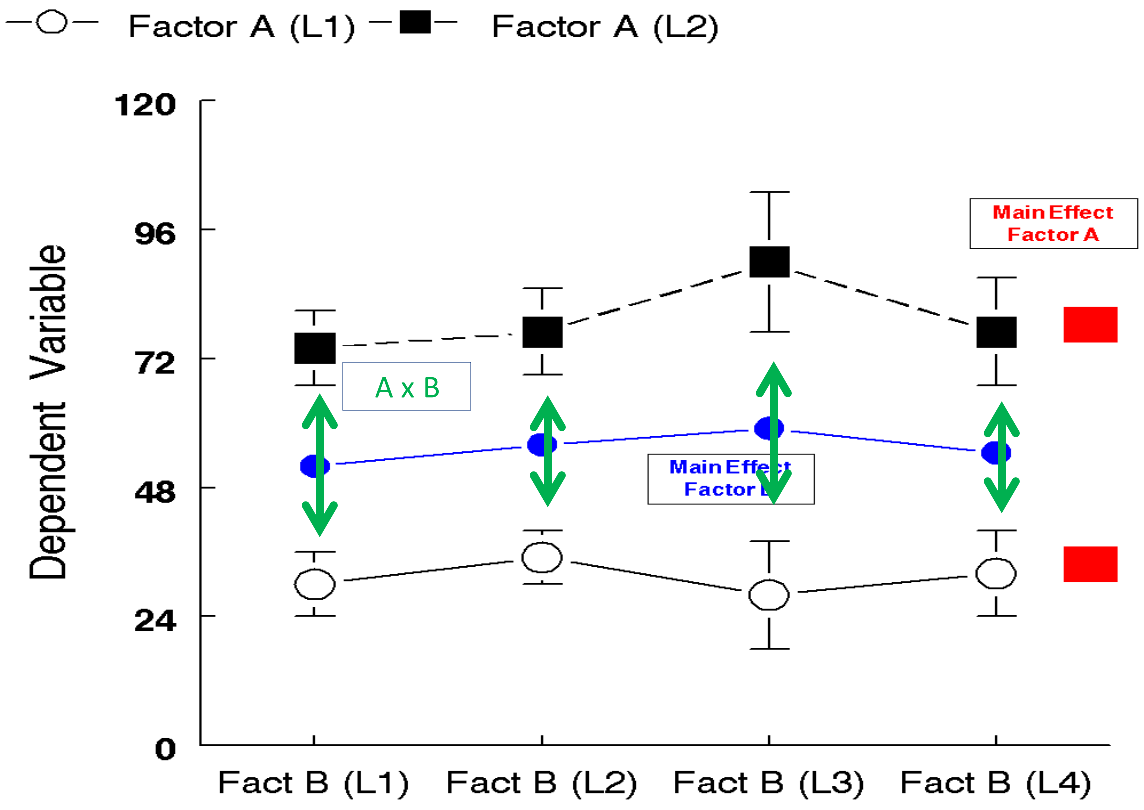
Two-way ANOVA

Two-way ANOVA

- Participating the total variation with respect to two-way factors/groups:
    - $SST = \mathbf{SST_{model}} + SSE_{residual}$
    - $\mathbf{SST_{model}}$= $\mathrm{SST_{factorA}}$ + $\mathrm{SST_{factorB}}$ +$\mathrm{SST_{AxB\ interaction}}$

Two-way ANOVA's Have Three Models

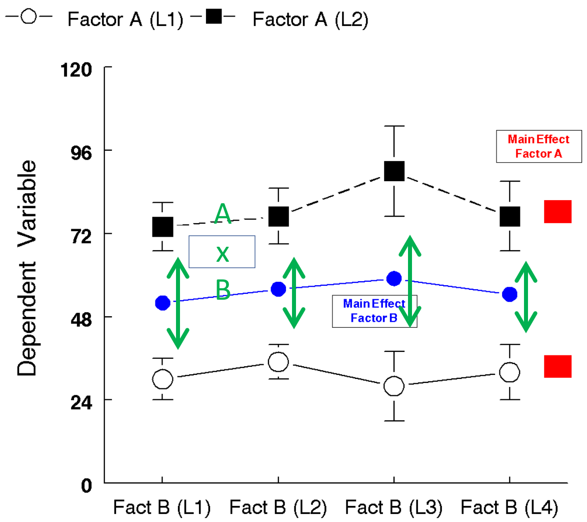
- Main Effect of Factor B (blue symbols, means)
- Main Effect of Factor A (red symbols, means)
- A x B Interaction (green lines, means of differences)



Two-way ANOVA

THREE Null Hypotheses

- Hypothesis for Factor A = ***Main Effect of Factor A***
- Hypothesis for Factor B = ***Main Effect of Factor B***
- Hypothesis for Interaction between Factor A and Factor B = ***Factor A x Factor B Interact***



Two-way ANOVA Table with independent/random samples

Source	Sum of Squares	df	Mean Square	F-Ratio
Main Effect-A	SST_A	df_A	MS_A	$MS_A / MSE_{residual}$
Main Effect-B	SST_B	df_B	MS_B	$MS_B / MSE_{residual}$
A X B Interact	SST_{AXB}	df_{AXB}	MS_{AXB}	$MS_{AXB} / MSE_{residual}$
Residual (Error)	$SSE_{residual}$	$df_{residual}$	$MSE_{residual}$	
Total	SST_{Total}	df_{Total}	MS_{Total}	

- F-statistic for Factor A:

$$F_A = \frac{MSA}{MSE}$$

where $MSA = \frac{SSA}{DFA}$ and $MSE = \frac{SSE}{DFE}$.

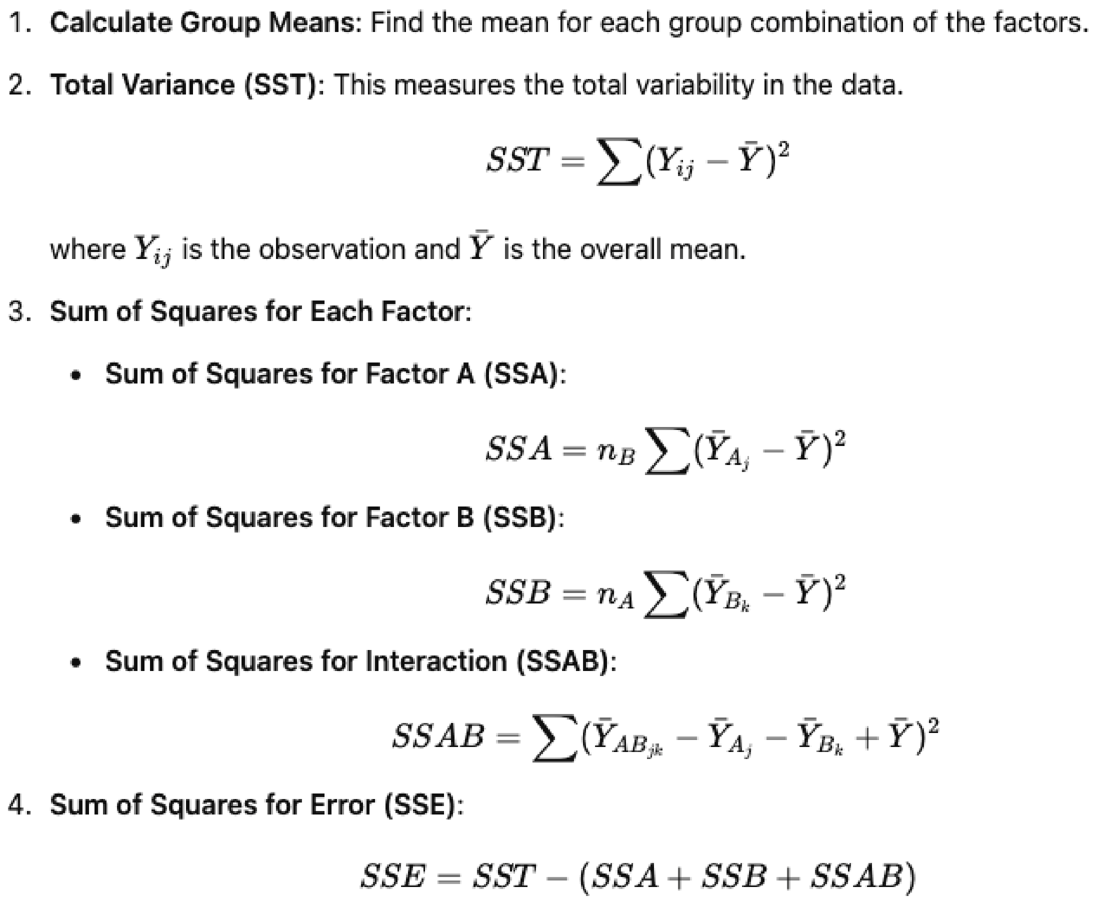
- F-statistic for Factor B:

$$F_B = \frac{MSB}{MSE}$$

- F-statistic for Interaction:

$$F_{AB} = \frac{MSAB}{MSE}$$

F-Statistic Calculation



1. **Calculate Group Means:** Find the mean for each group combination of the factors.
2. **Total Variance (SST):** This measures the total variability in the data.

$$SST = \sum (Y_{ij} - \bar{Y})^2$$

   where Y_{ij} is the observation and $\bar{Y}$ is the overall mean.

3. **Sum of Squares for Each Factor:**
   - **Sum of Squares for Factor A (SSA):**

$$SSA = n_B \sum (\bar{Y}_{A_j} - \bar{Y})^2$$

   - **Sum of Squares for Factor B (SSB):**

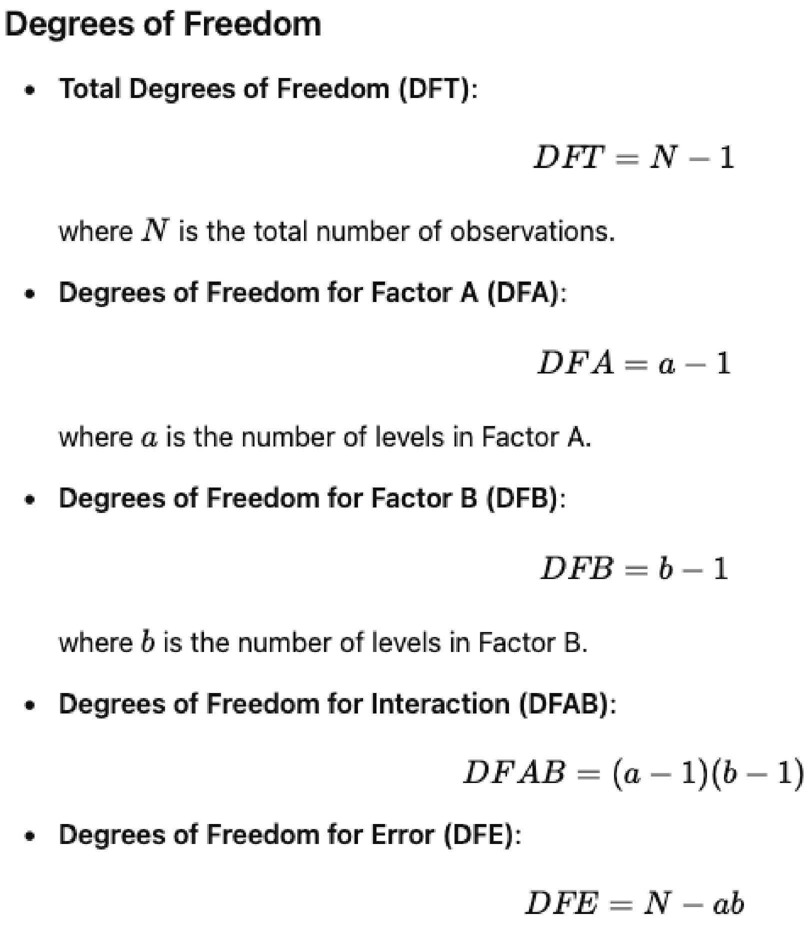
$$SSB = n_A \sum (\bar{Y}_{B_k} - \bar{Y})^2$$

   - **Sum of Squares for Interaction (SSAB):**

$$SSAB = \sum (\bar{Y}_{AB_{jk}} - \bar{Y}_{A_j} - \bar{Y}_{B_k} + \bar{Y})^2$$

4. **Sum of Squares for Error (SSE):**

$$SSE = SST - (SSA + SSB + SSAB)$$



Degrees of Freedom

- **Total Degrees of Freedom (DFT):**

$$DFT = N - 1$$

  where N is the total number of observations.

- **Degrees of Freedom for Factor A (DFA):**

$$DFA = a - 1$$

  where a is the number of levels in Factor A.

- **Degrees of Freedom for Factor B (DFB):**

$$DFB = b - 1$$

  where b is the number of levels in Factor B.

- **Degrees of Freedom for Interaction (DFAB):**

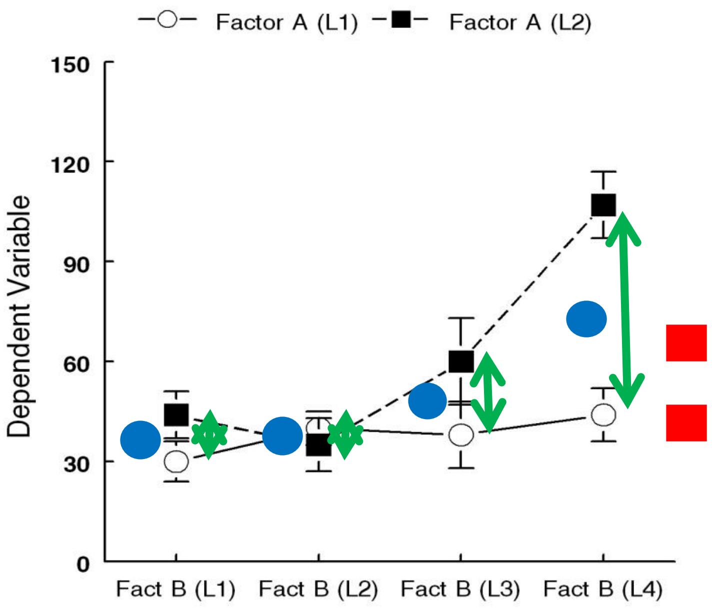
$$DFAB = (a - 1)(b - 1)$$

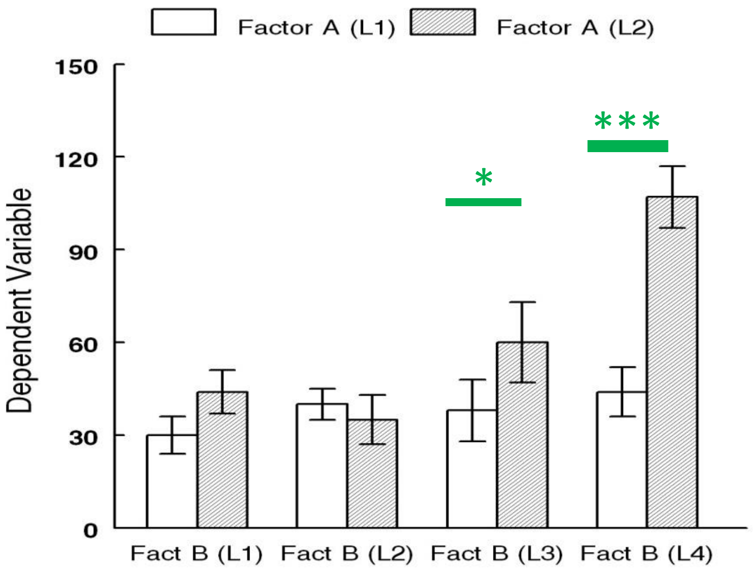
- **Degrees of Freedom for Error (DFE):**

$$DFE = N - ab$$

Two-way ANOVA: 3 omnibus null hypotheses

- Main effect Factor A: $H_0\colon \sigma_A^2 \le \sigma_r^2$
- Main effect Factor B: $H_0\colon \sigma_B^2 \le \sigma_r^2$
- Interaction AxB: $H_0\colon \sigma_{A\times B}^2 \le \sigma_r^2$





Main effects usually detected when interaction occurs

Main effect of A?

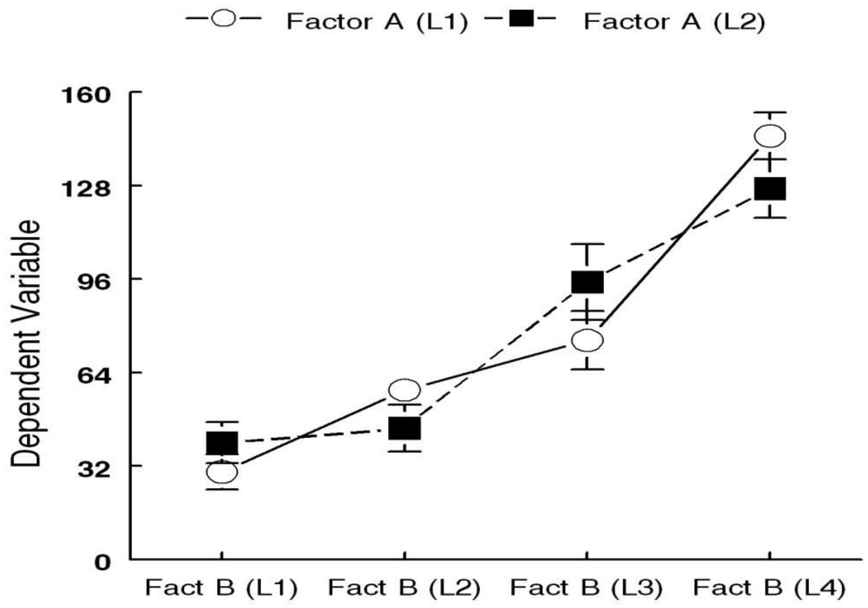
yes

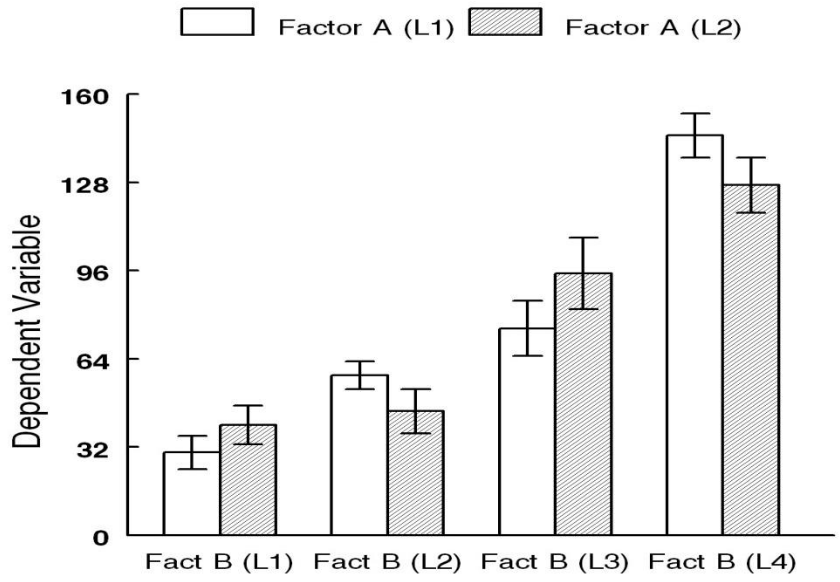
Main effect of B?

yes

AXB Interaction?

yes





Main effect of A?

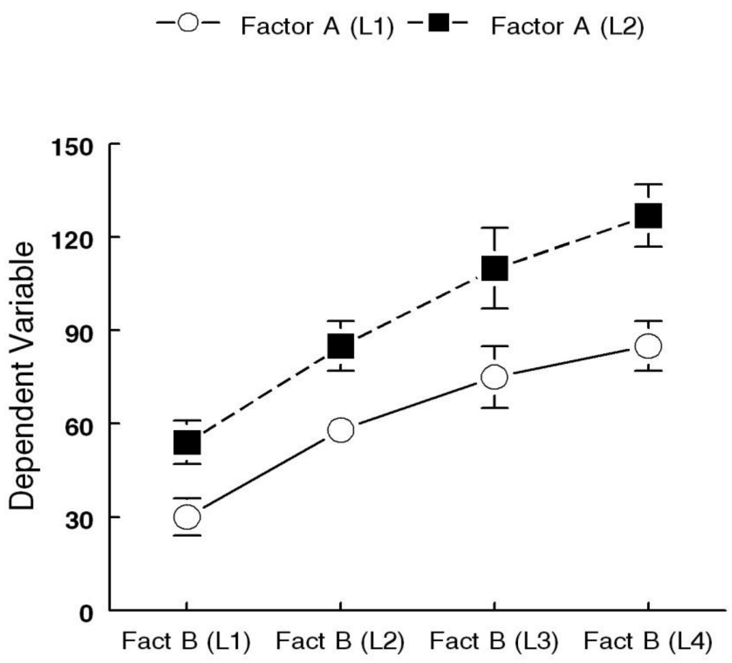
none

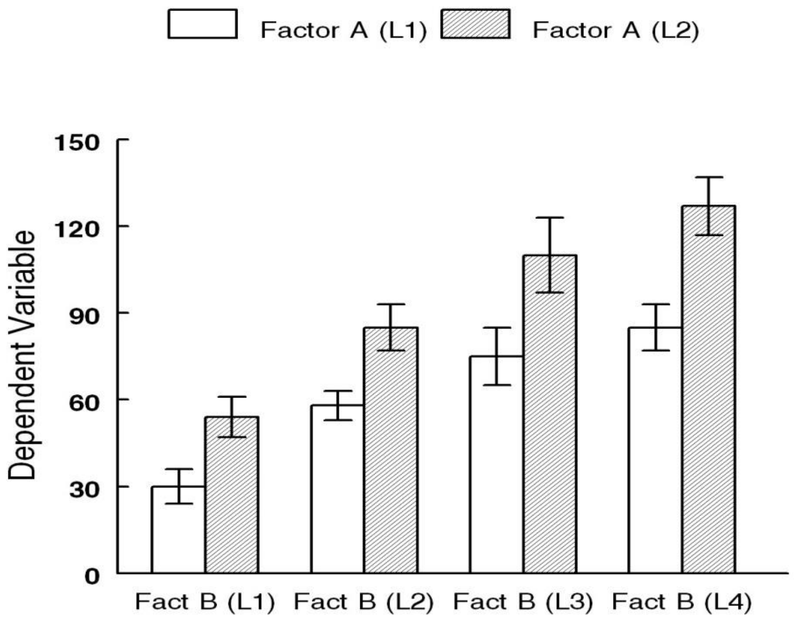
Main effect of B?

yes

AXB Interaction?

none





Main effect of A?

yes

Main effect of B?

yes

AXB Interaction?

none

ANOVA: Completely Randomized vs. Related Measures

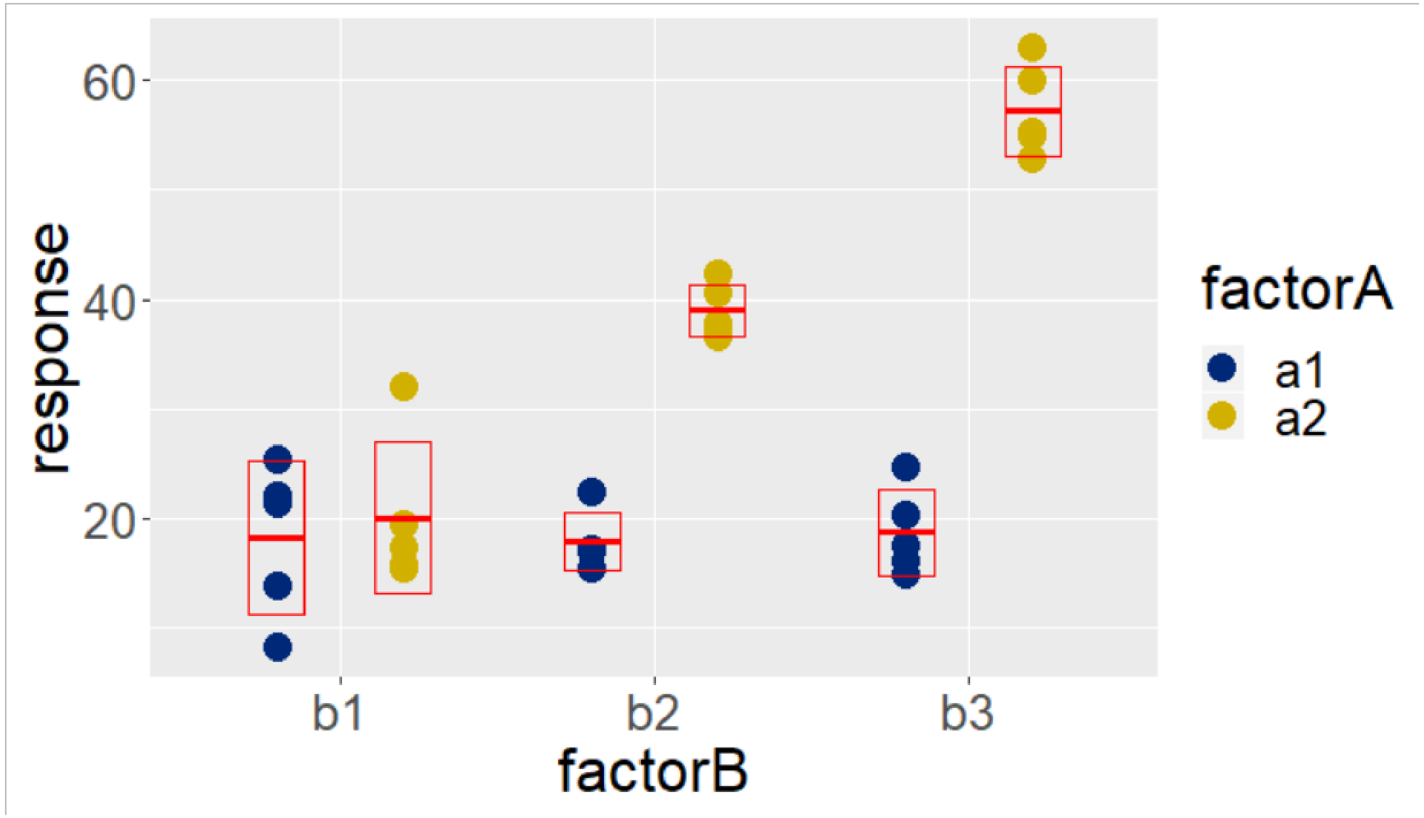
Example related measures:

- before-after
- identical twins
- isogenic littermates
- split tissue
- cell culture
- one cell
- one extract

Question: Are any measurements intrinsically-related?

Yes? Repeated measurements of the same sample, or measurements of closely related samples.

Two-way ANOVA: Completely Randomized on factorA and factorB



N=30

Groups are independent

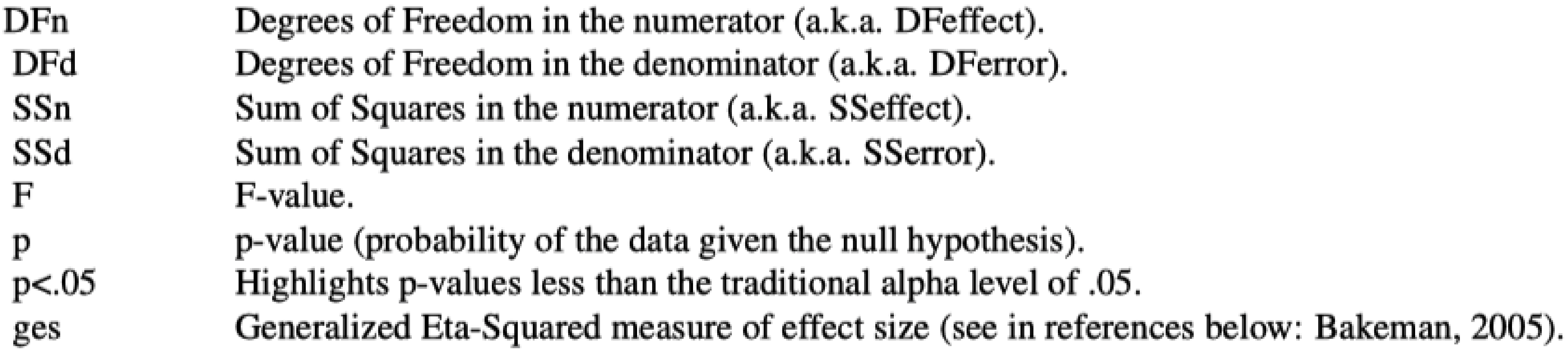
Two-way ANOVA: Completely Randomized on factorA and factorB

```
out <- ezANOVA(
  …
 between= c(factorB,
            factorA),
…
)
```

between: “levels of this factor vary between samples”

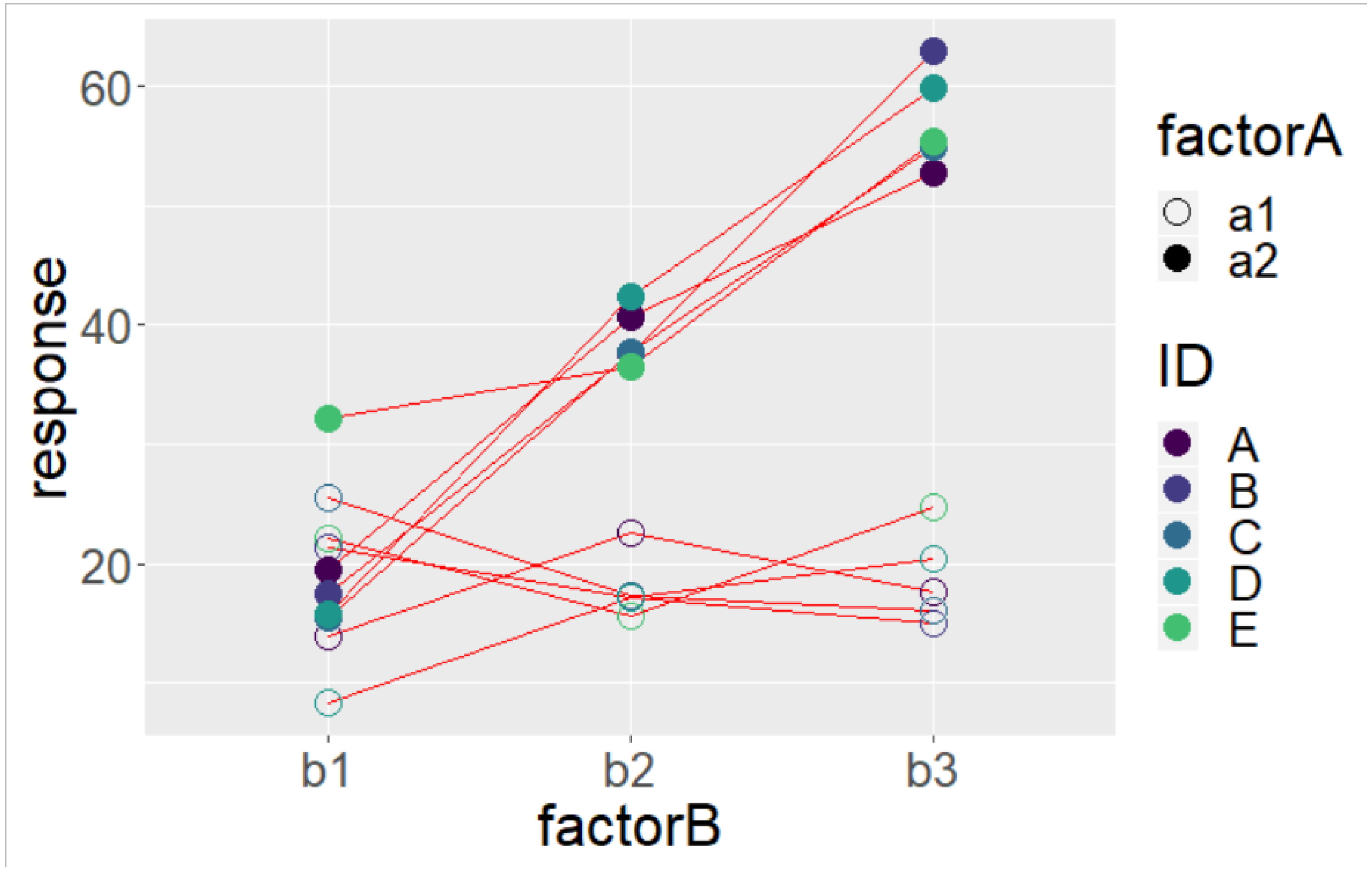
```
out$ANOVA
Effect            DFn DFd SSn  SSd F     p         p<.05 ges
1 factorA           1 24 3123 566 132.3 2.988e-11 *     0.846
2 factorB           2 24 1770 566  37.5 4.117e-08 *     0.757
3 factorA:factorB 2 24 1673 566  35.4 6.871e-08 *     0.747
```

Output Variables by ezANOVA():



DFn	Degrees of Freedom in the numerator (a.k.a. DFeffect).
DFd	Degrees of Freedom in the denominator (a.k.a. DFerror).
SSn	Sum of Squares in the numerator (a.k.a. SSeffect).
SSd	Sum of Squares in the denominator (a.k.a. SSerror).
F	F-value.
p	p-value (probability of the data given the null hypothesis).
p<.05	Highlights p-values less than the traditional alpha level of .05.
ges	Generalized Eta-Squared measure of effect size (see in references below: Bakeman, 2005).

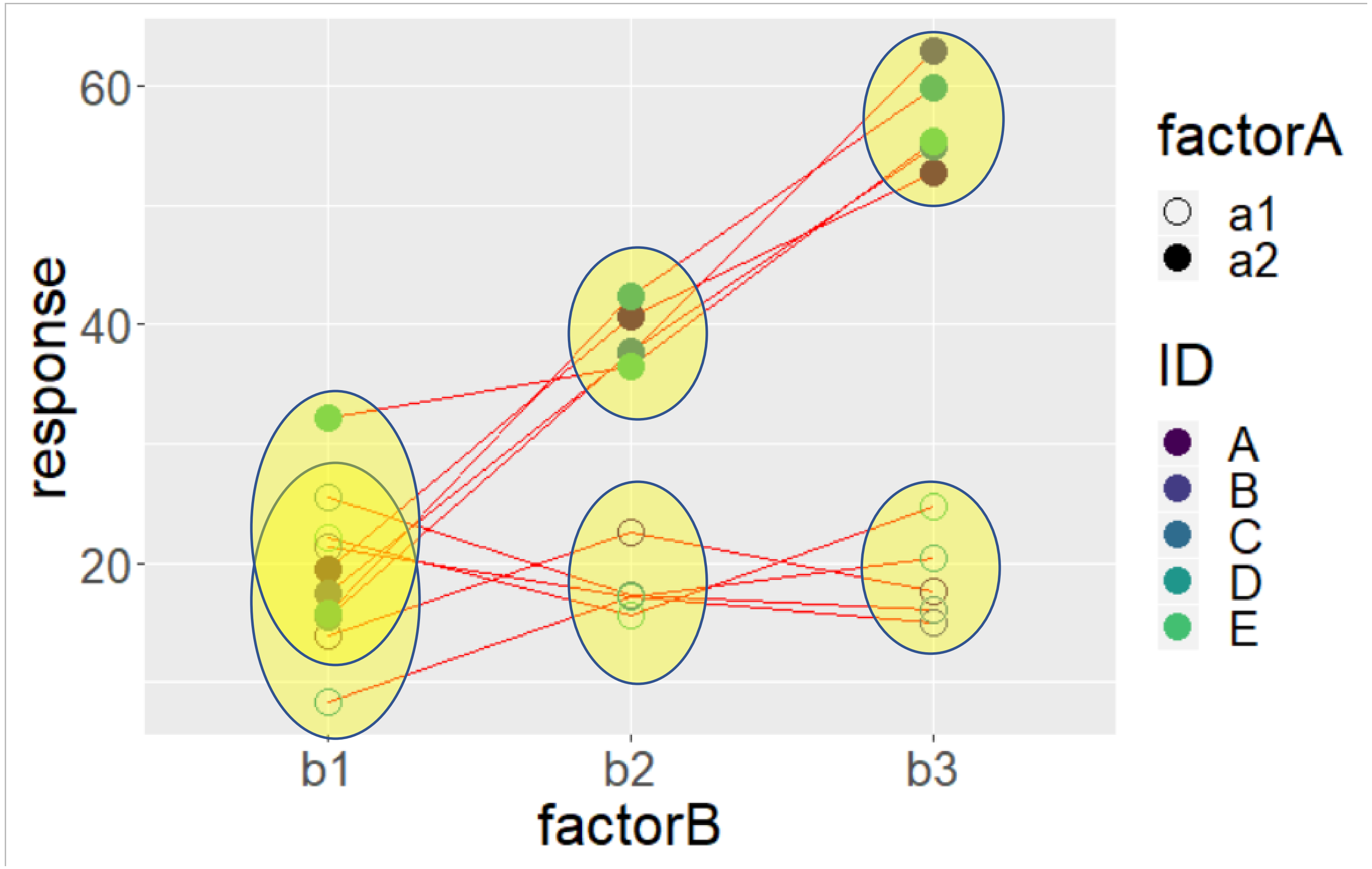
Two-way ANOVA: Related Measures on factor A and factor B



N=5

Measurements are related across groups.

Group means are not independent! We can’t unpair them!!

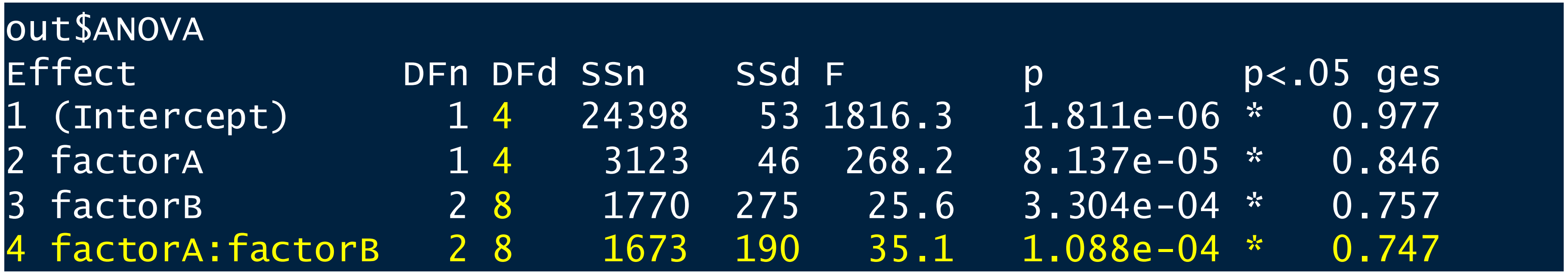


Every replicate (color) is independent

Two-way ANOVA : Repeated Measures on factorA and factorB

```
out <- ezANOVA(
  …
within = c(factorA,
           factorB),
…
)
```

within: “levels of this factor vary within samples”

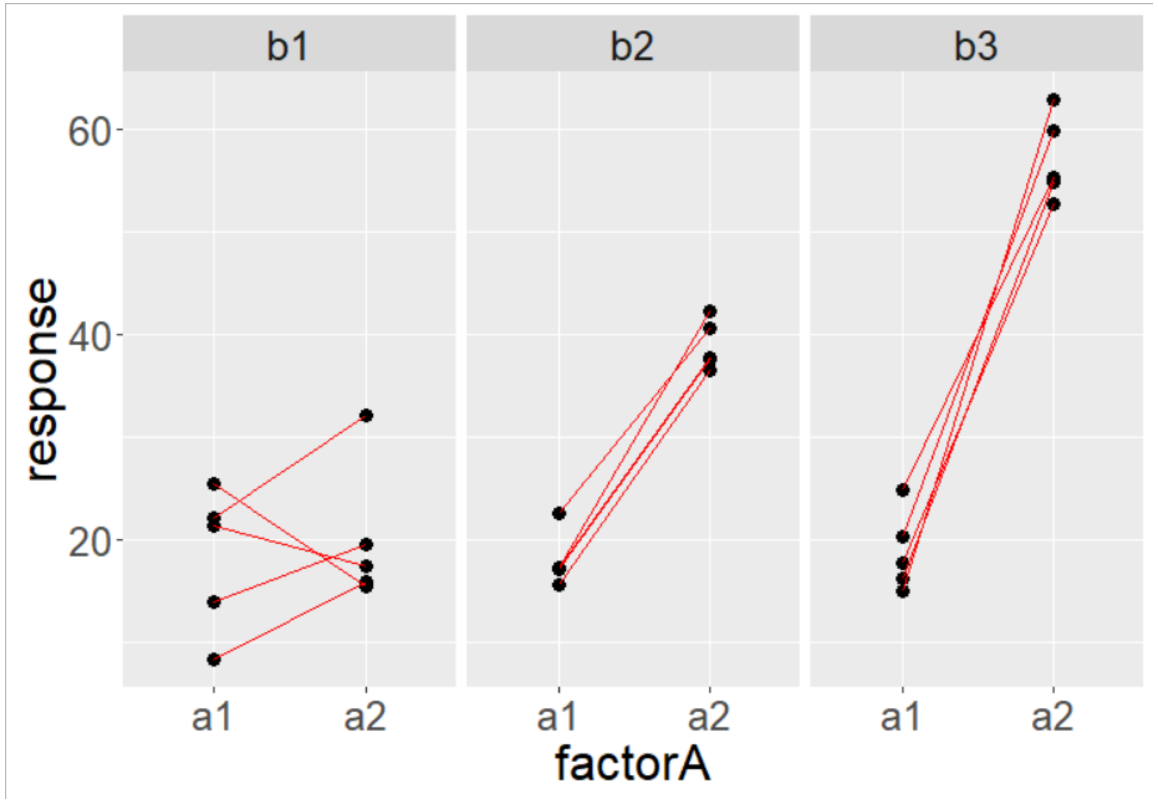


```
out$ANOVA
Effect             DFn DFd SSn    SSd F        p          p<.05 ges
1 (Intercept)        1 4   24398   53 1816.3   1.811e-06 *    0.977
2 factorA            1 4    3123   46  268.2   8.137e-05 *    0.846
3 factorB            2 8    1770  275   25.6   3.304e-04 *    0.757
4 factorA:factorB    2 8    1673  190   35.1   1.088e-04 *    0.747
```

Two-way ANOVA mixed

Related Measures on factor A

Completely Randomized on factor B



N=15

A groups are related measures

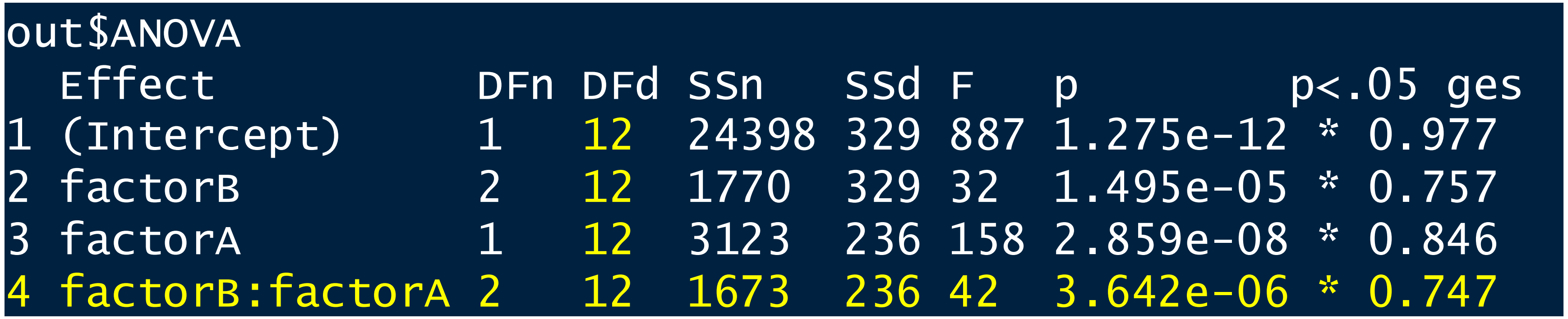
B groups are independent

Two-way ANOVA: mixed

Related Measures on factor A
Completely Randomized on factor B

```
out <- ezANOVA(
  …
 between= factorB,
 within = factorA
…
)
```

between: “levels of this factor vary between samples”
within: “levels of this factor vary within samples”

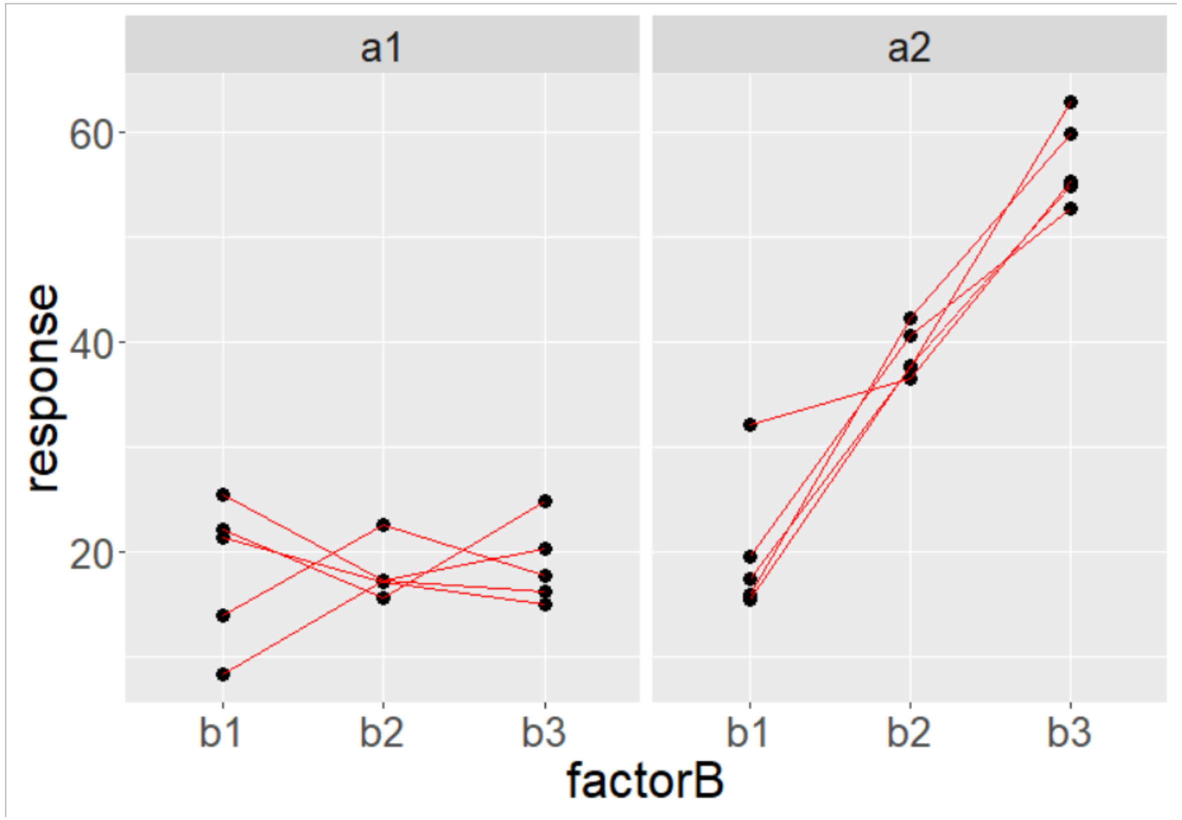


```
out$ANOVA
  Effect          DFn DFd SSn   SSd F   p         p<.05 ges
1 (Intercept)     1   12  24398 329 887 1.275e-12 * 0.977
2 factorB         2   12  1770  329 32  1.495e-05 * 0.757
3 factorA         1   12  3123  236 158 2.859e-08 * 0.846
4 factorB:factorA 2   12  1673  236 42  3.642e-06 * 0.747
```

Two-way ANOVA mixed

Completely Randomized on factor A

Related Measures on factor B



N=10

A groups are independent

B groups are related measures

Two-way ANOVA mixed

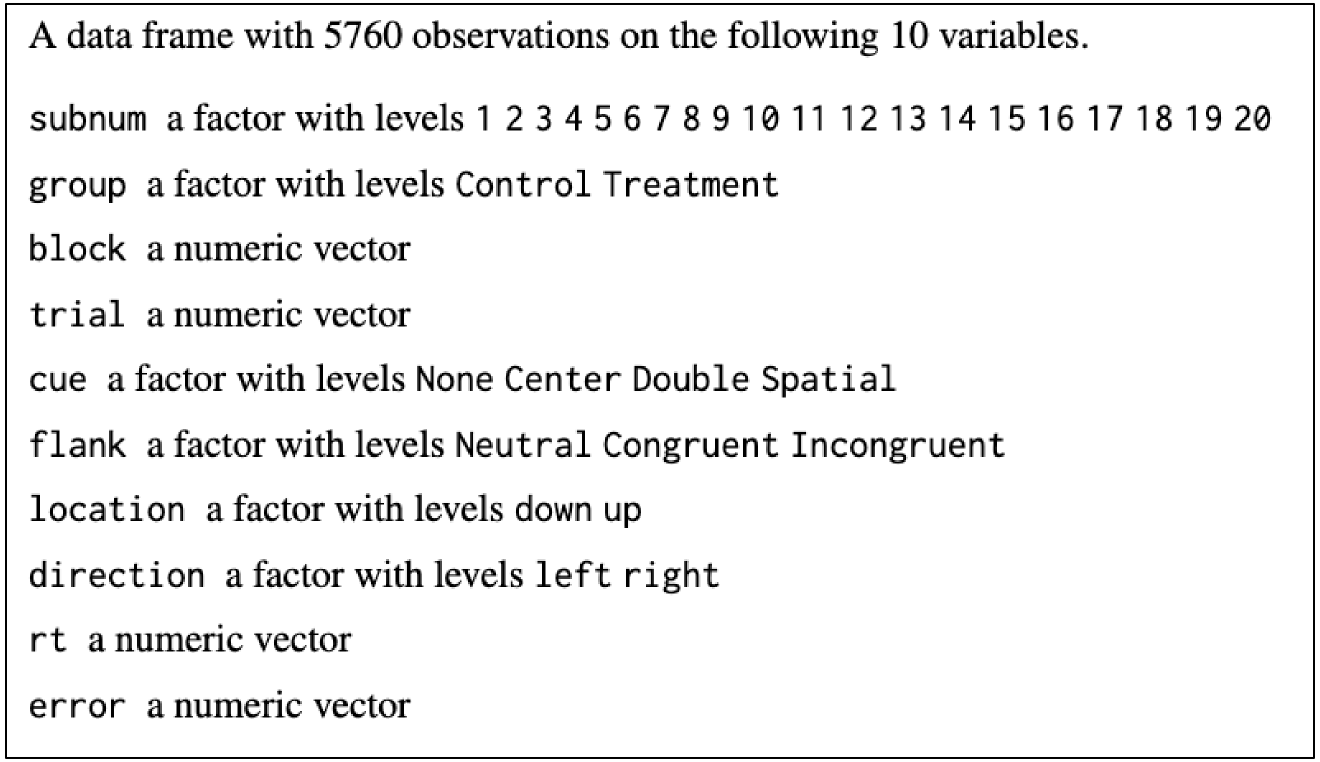
Completely Randomized on factor A
Related Measures on factor B

```
out <- ezANOVA(
  …
 between= factorA,
 within = factorB,
…
)
```

between: “levels of this factor vary between samples”
within: “levels of this factor vary within samples”

```
out$ANOVA
Effect          DFn DFd SSn   SSd F    p         p<.05 ges
1 (Intercept)     1 8   24398 100 1945 7.700e-11 * 0.977
2 factorA         1 8    3123 100  249 2.595e-07 * 0.846
3 factorB         2 16   1770 466   30 3.564e-06 * 0.757
4 factorA:factorB 2 16   1673 466   28 5.092e-06 * 0.747
```

Simulated Example data from Attention Network Test (ANT)



A data frame with 5760 observations on the following 10 variables.

`subnum` a factor with levels `1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20`

`group` a factor with levels `Control Treatment`

`block` a numeric vector

`trial` a numeric vector

`cue` a factor with levels `None Center Double Spatial`

`flank` a factor with levels `Neutral Congruent Incongruent`

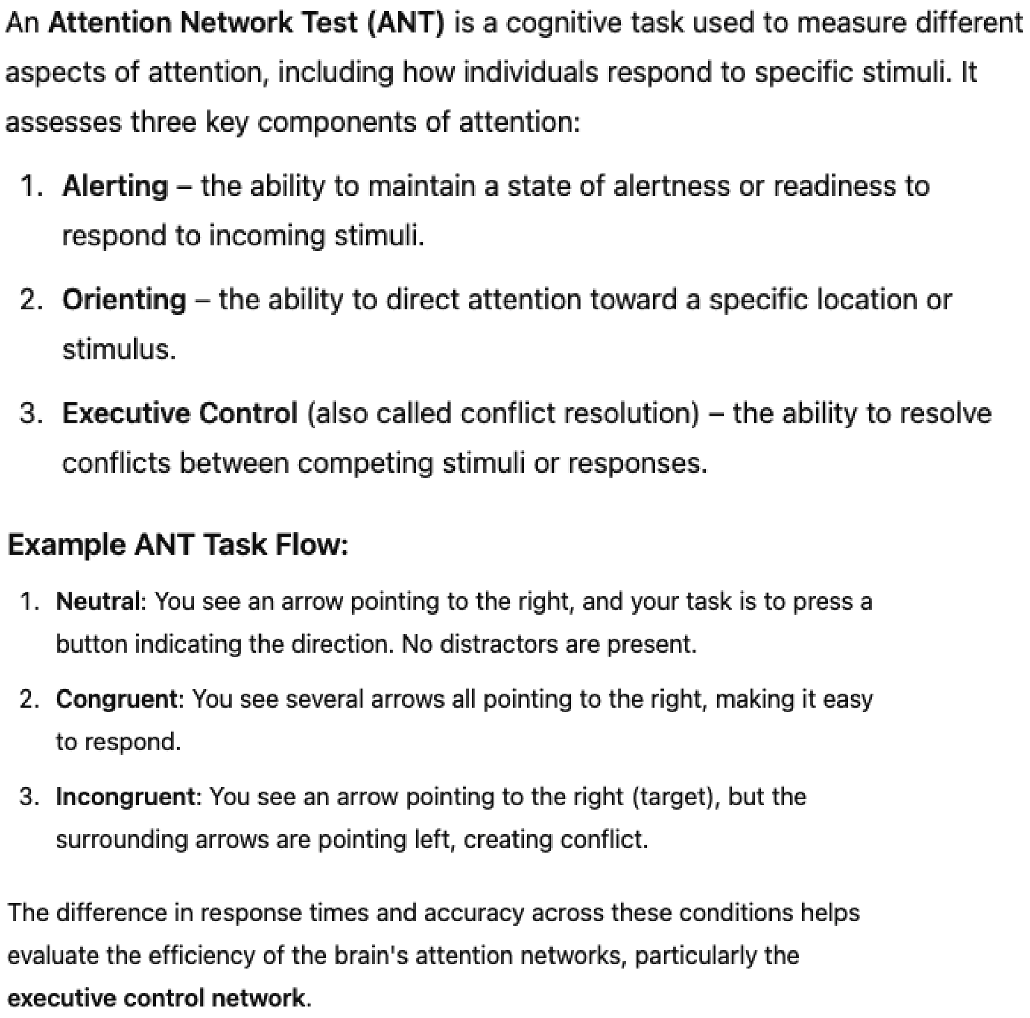
`location` a factor with levels `down up`

`direction` a factor with levels `left right`

`rt` a numeric vector

`error` a numeric vector

- Within-Ss variables ("cue" and "flank");
- Between-Ss variable ("group");
- Dependent variables (response time, "rt", and whether an error was made, "error")



An **Attention Network Test (ANT)** is a cognitive task used to measure different aspects of attention, including how individuals respond to specific stimuli. It assesses three key components of attention:

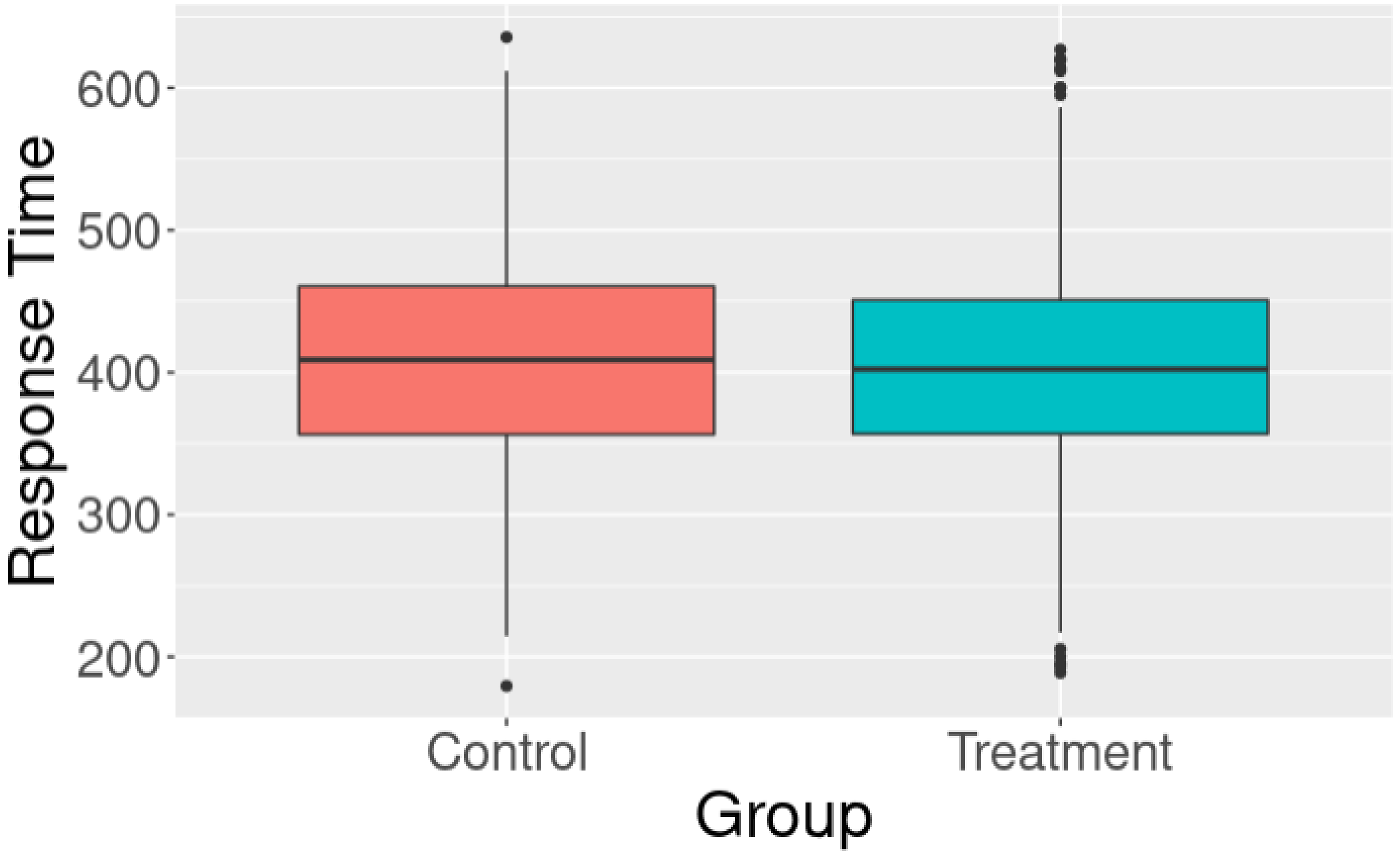
1. **Alerting** – the ability to maintain a state of alertness or readiness to respond to incoming stimuli.
2. **Orienting** – the ability to direct attention toward a specific location or stimulus.
3. **Executive Control** (also called conflict resolution) – the ability to resolve conflicts between competing stimuli or responses.

Example ANT Task Flow:

1. **Neutral**: You see an arrow pointing to the right, and your task is to press a button indicating the direction. No distractors are present.
2. **Congruent**: You see several arrows all pointing to the right, making it easy to respond.
3. **Incongruent**: You see an arrow pointing to the right (target), but the surrounding arrows are pointing left, creating conflict.

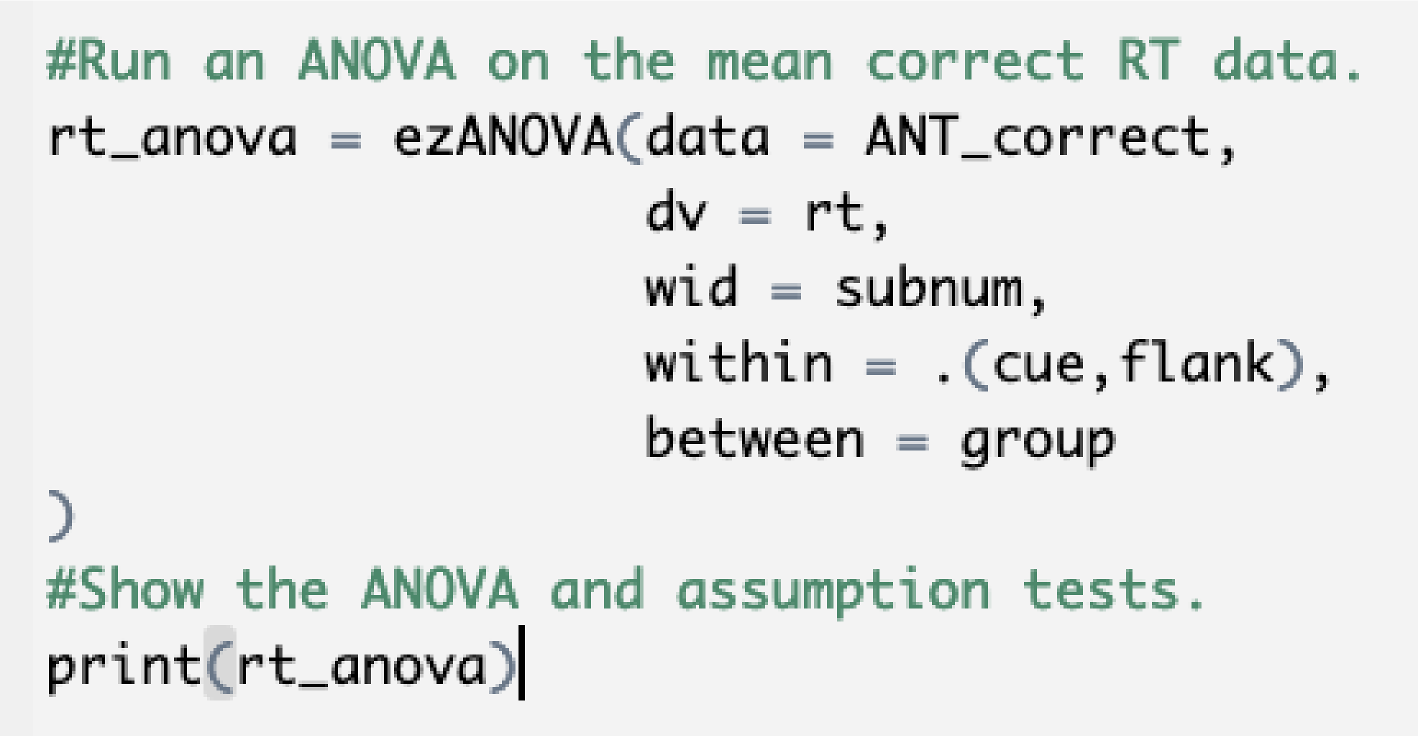
The difference in response times and accuracy across these conditions helps evaluate the efficiency of the brain's attention networks, particularly the **executive control network**.

Response Time vs. Group



Mixed two-way ANOVA

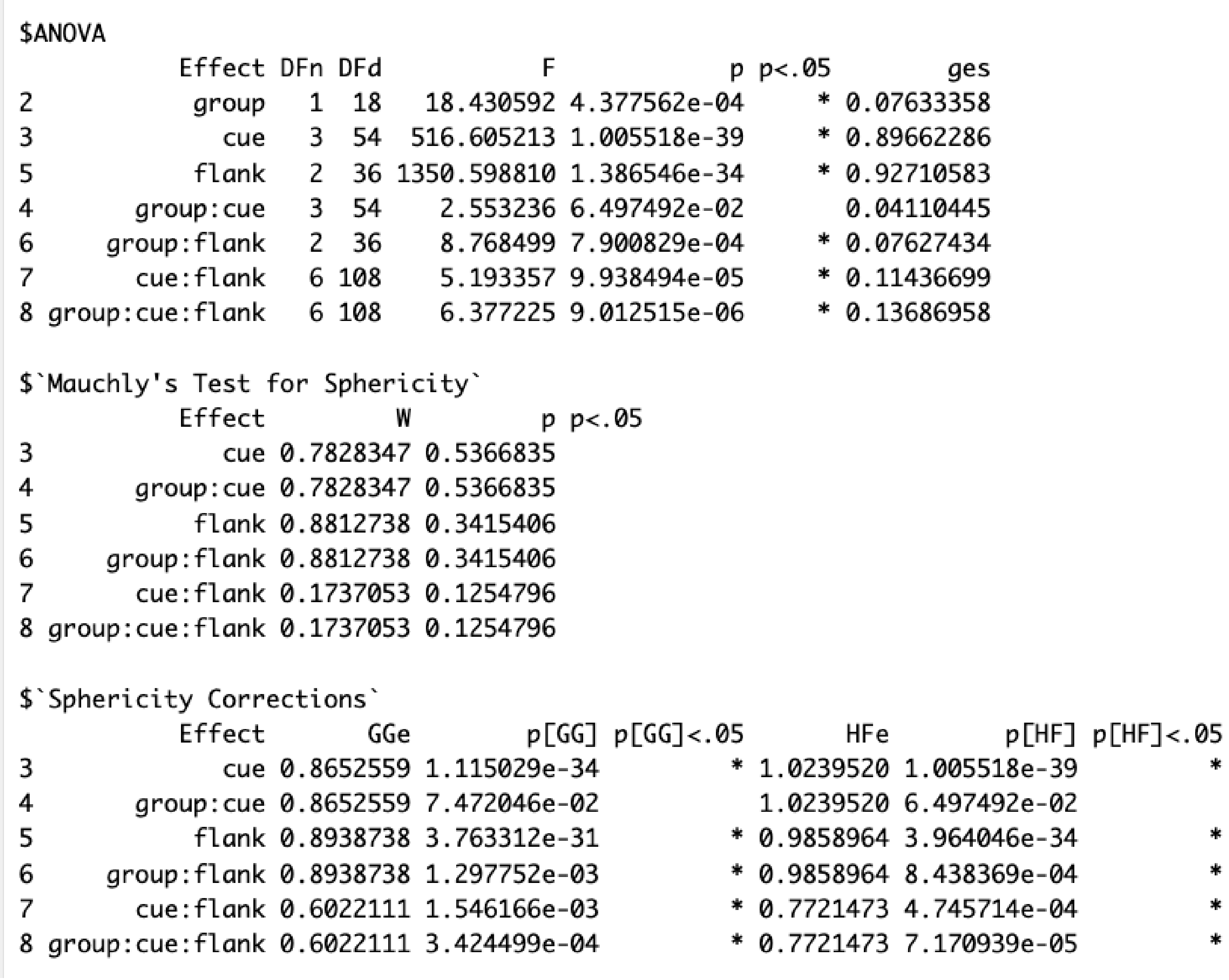
- Response: response time (rt)
- Completely Randomized on group
- Related Measures on cue and flank



```
#Run an ANOVA on the mean correct RT data.
rt_anova = ezANOVA(data = ANT_correct,
                   dv = rt,
                   wid = subnum,
                   within = .(cue,flank),
                   between = group
)
#Show the ANOVA and assumption tests.
print(rt_anova)
```

between: “levels of this factor vary between samples”
within: “levels of this factor vary within samples”

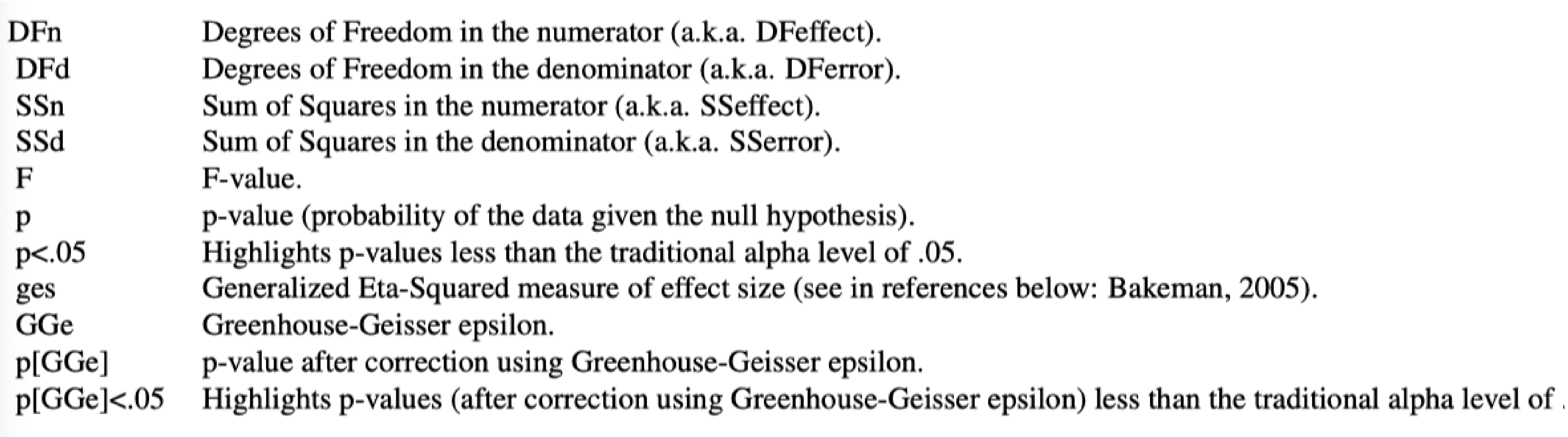
Mixed two-way ANOVA results by ezANOVA()



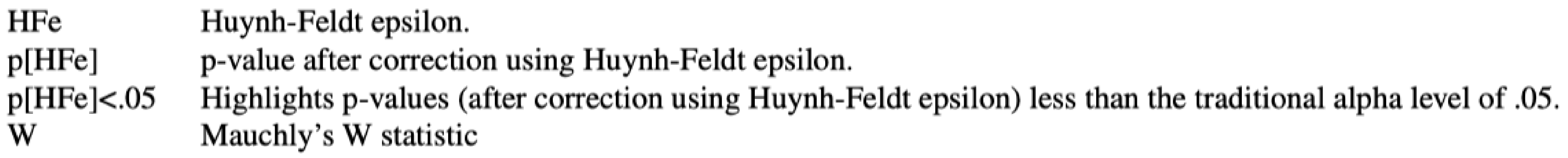
```
$ANOVA
           Effect DFn DFd           F            p p<.05        ges
2           group   1  18   18.430592 4.377562e-04     * 0.07633358
3             cue   3  54  516.605213 1.005518e-39     * 0.89662286
5           flank   2  36 1350.598810 1.386546e-34     * 0.92710583
4       group:cue   3  54    2.553236 6.497492e-02       0.04110445
6     group:flank   2  36    8.768499 7.900829e-04     * 0.07627434
7       cue:flank   6 108    5.193357 9.938494e-05     * 0.11436699
8 group:cue:flank   6 108    6.377225 9.012515e-06     * 0.13686958

$`Mauchly's Test for Sphericity`
           Effect         W         p p<.05
3             cue 0.7828347 0.5366835
4       group:cue 0.7828347 0.5366835
5           flank 0.8812738 0.3415406
6     group:flank 0.8812738 0.3415406
7       cue:flank 0.1737053 0.1254796
8 group:cue:flank 0.1737053 0.1254796

$`Sphericity Corrections`
           Effect       GGe        p[GG] p[GG]<.05       HFe        p[HF] p[HF]<.05
3             cue 0.8652559 1.115029e-34         * 1.0239520 1.005518e-39         *
4       group:cue 0.8652559 7.472046e-02           1.0239520 6.497492e-02
5           flank 0.8938738 3.763312e-31         * 0.9858964 3.964046e-34         *
6     group:flank 0.8938738 1.297752e-03         * 0.9858964 8.438369e-04         *
7       cue:flank 0.6022111 1.546166e-03         * 0.7721473 4.745714e-04         *
8 group:cue:flank 0.6022111 3.424499e-04         * 0.7721473 7.170939e-05         *
```

Sphericity refers to the condition where the variances of the differences between all combinations of related (paired) groups are equal.

Output variables by ezANOVA():

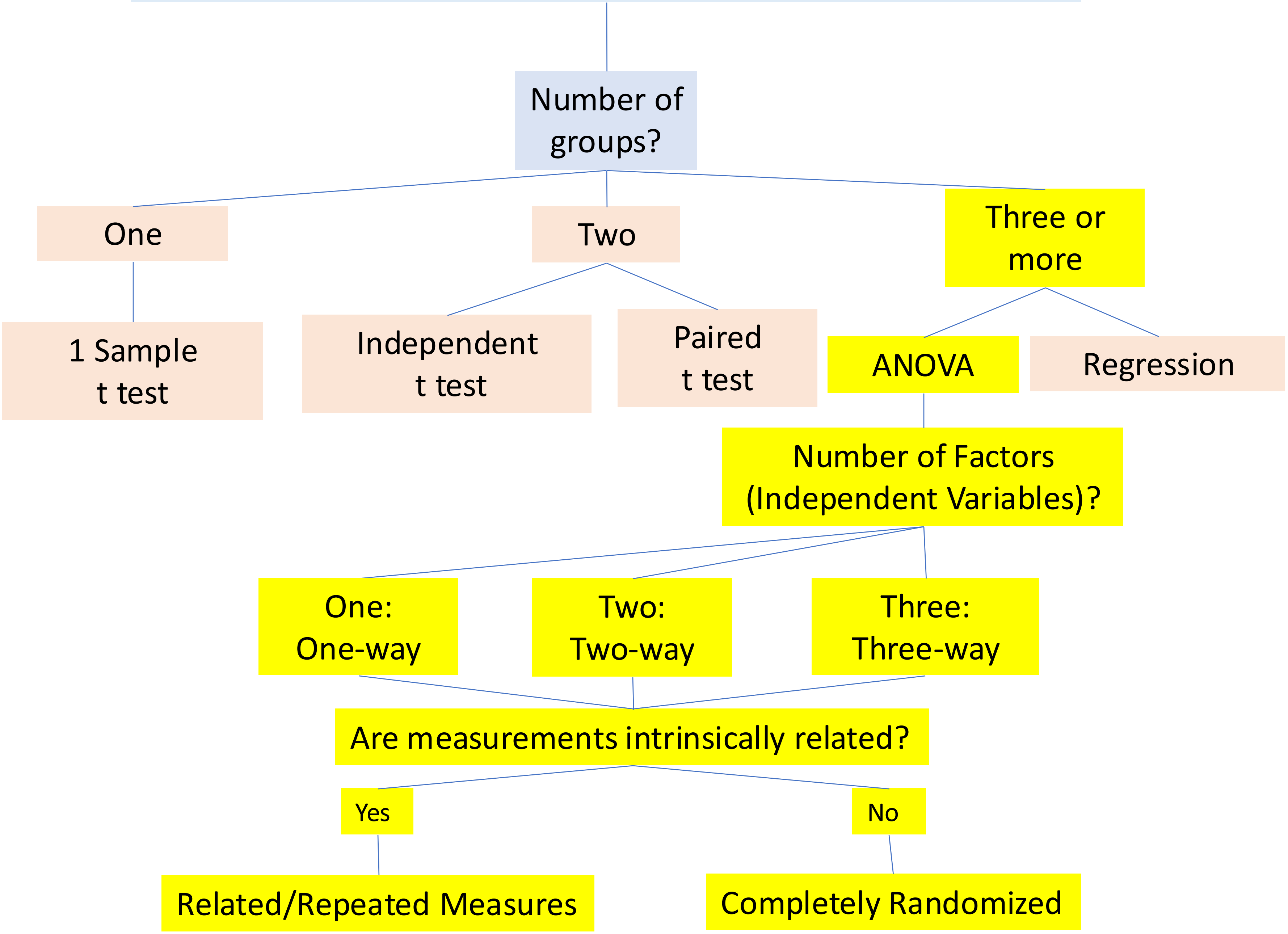


DFn	Degrees of Freedom in the numerator (a.k.a. DFeffect).
DFd	Degrees of Freedom in the denominator (a.k.a. DFerror).
SSn	Sum of Squares in the numerator (a.k.a. SSeffect).
SSd	Sum of Squares in the denominator (a.k.a. SSerror).
F	F-value.
p	p-value (probability of the data given the null hypothesis).
p<.05	Highlights p-values less than the traditional alpha level of .05.
ges	Generalized Eta-Squared measure of effect size (see in references below: Bakeman, 2005).
GGe	Greenhouse-Geisser epsilon.
p[GGe]	p-value after correction using Greenhouse-Geisser epsilon.
p[GGe]<.05	Highlights p-values (after correction using Greenhouse-Geisser epsilon) less than the traditional alpha level of



HFe	Huynh-Feldt epsilon.
p[HFe]	p-value after correction using Huynh-Feldt epsilon.
p[HFe]<.05	Highlights p-values (after correction using Huynh-Feldt epsilon) less than the traditional alpha level of .05.
W	Mauchly’s W statistic

Outcome/Response/Dependent Variable is Continuous



Number of groups?

One

1 Sample t test

Two

Independent t test

Paired t test

Three or more

ANOVA

Regression

Number of Factors (Independent Variables)?

One: One-way

Two: Two-way

Three: Three-way

Are measurements intrinsically related?

Yes

Related/Repeated Measures

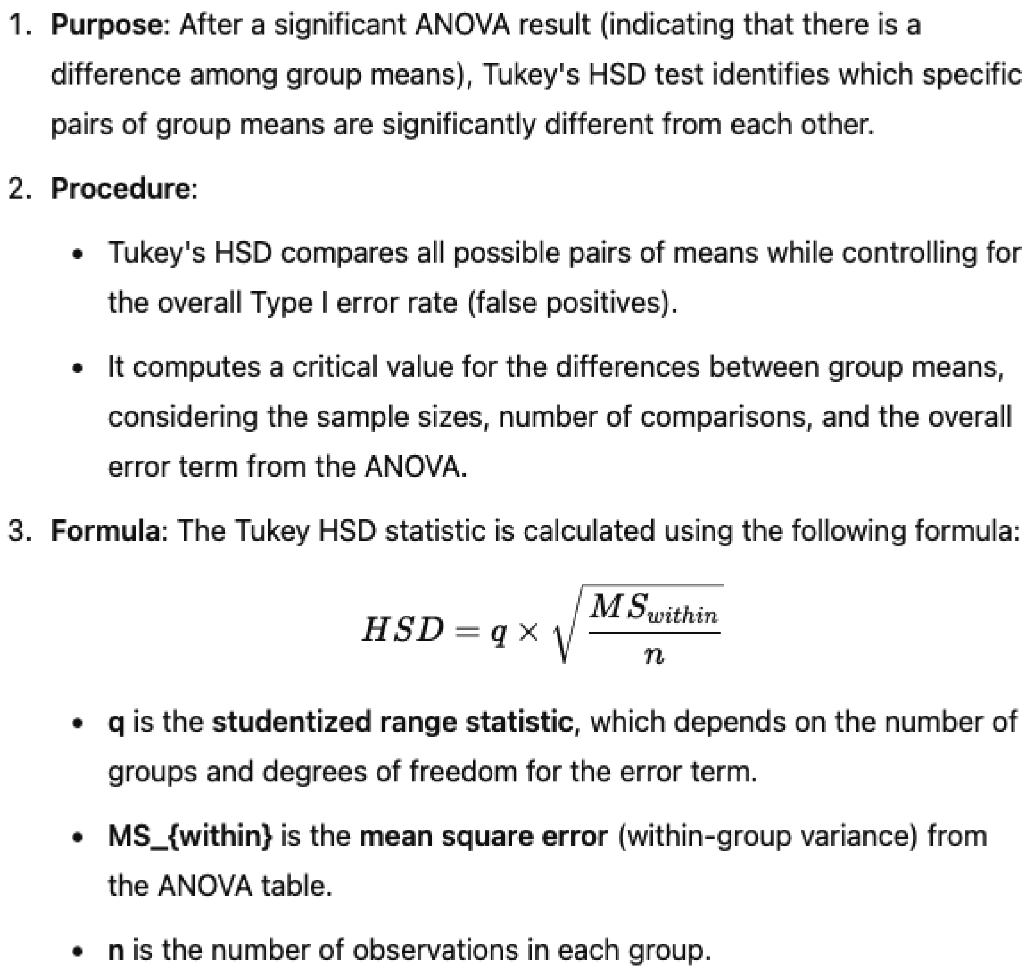
No

Completely Randomized

Steps to Conduct ANOVA

1. **State Hypotheses**:
    - Null Hypothesis (H0): There is no effect of the factors on the dependent variable.
    - Alternative Hypothesis (H1): There is an effect of at least one factor.
2. **Collect Data**: Ensure you have a balanced design if possible (equal sample sizes in each group).
3. **Analyze Data**: Use statistical software or manual calculations to perform the ANOVA. The output will include F-values and p-values for the main effects and interactions.
4. **Interpret Results**:
    - If the p-value is less than the significance level (commonly 0.05), you reject the null hypothesis.
    - Assess interaction effects to see if they significantly influence the dependent variable.
5. **Post-Hoc Tests**: If significant differences are found, post-hoc tests (like Tukey’s HSD) can help identify which groups differ.

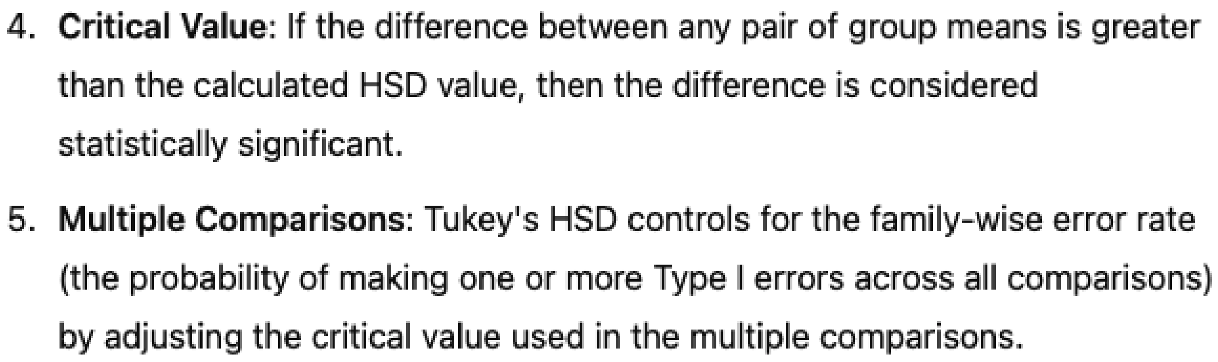
Tukey’s Honestly Significant Difference Test



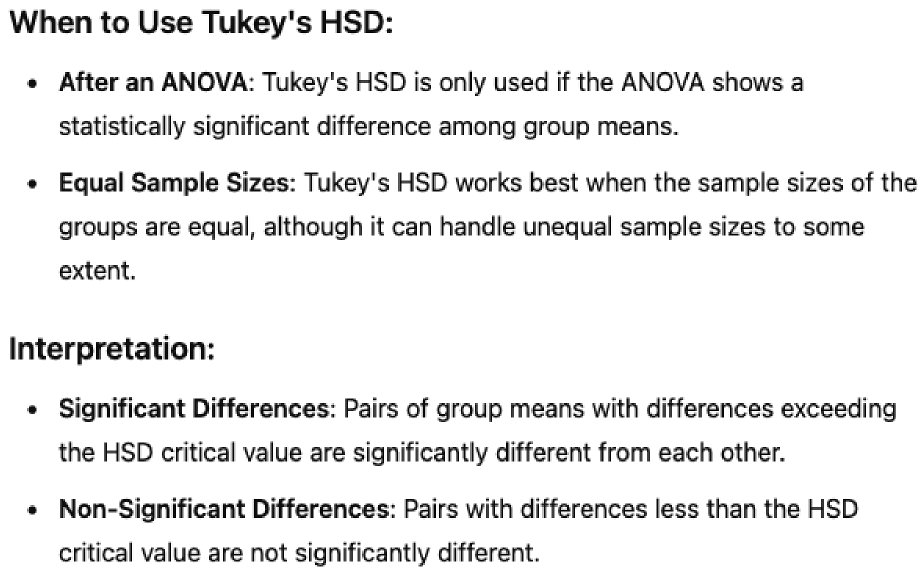
1. **Purpose**: After a significant ANOVA result (indicating that there is a difference among group means), Tukey's HSD test identifies which specific pairs of group means are significantly different from each other.
2. **Procedure**:
    - Tukey's HSD compares all possible pairs of means while controlling for the overall Type I error rate (false positives).
    - It computes a critical value for the differences between group means, considering the sample sizes, number of comparisons, and the overall error term from the ANOVA.
3. **Formula**: The Tukey HSD statistic is calculated using the following formula:

$$HSD = q \times \sqrt{\frac{MS_{within}}{n}}$$

    - **q** is the **studentized range statistic**, which depends on the number of groups and degrees of freedom for the error term.
    - **MS_{within}** is the **mean square error** (within-group variance) from the ANOVA table.
    - **n** is the number of observations in each group.



4. **Critical Value**: If the difference between any pair of group means is greater than the calculated HSD value, then the difference is considered statistically significant.
5. **Multiple Comparisons**: Tukey's HSD controls for the family-wise error rate (the probability of making one or more Type I errors across all comparisons) by adjusting the critical value used in the multiple comparisons.



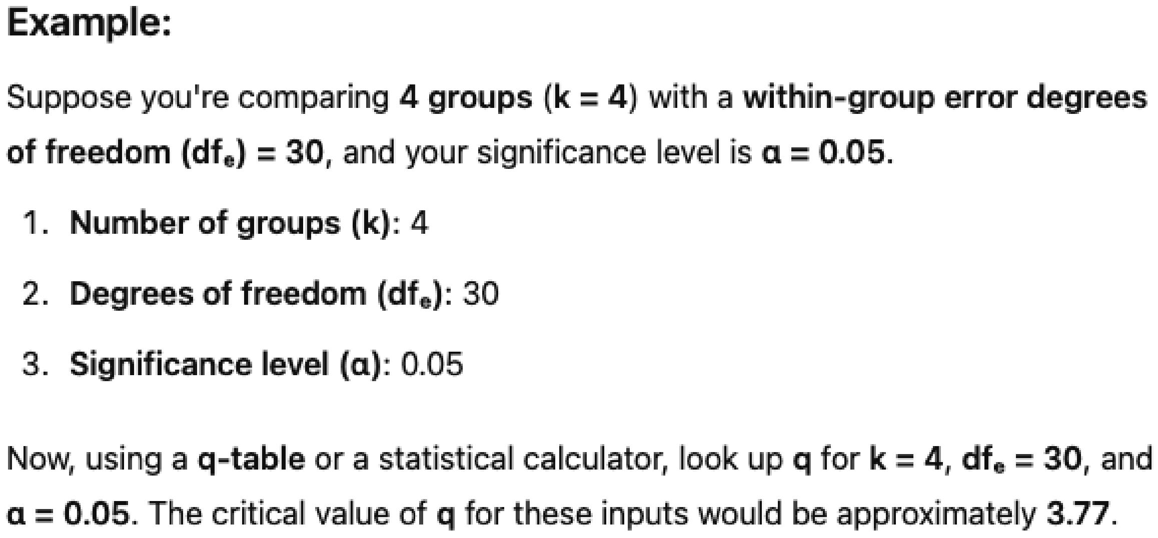
When to Use Tukey's HSD:

- **After an ANOVA**: Tukey's HSD is only used if the ANOVA shows a statistically significant difference among group means.
- **Equal Sample Sizes**: Tukey's HSD works best when the sample sizes of the groups are equal, although it can handle unequal sample sizes to some extent.

Interpretation:

- **Significant Differences**: Pairs of group means with differences exceeding the HSD critical value are significantly different from each other.
- **Non-Significant Differences**: Pairs with differences less than the HSD critical value are not significantly different.

Tukey’s Honestly Significant Difference Test

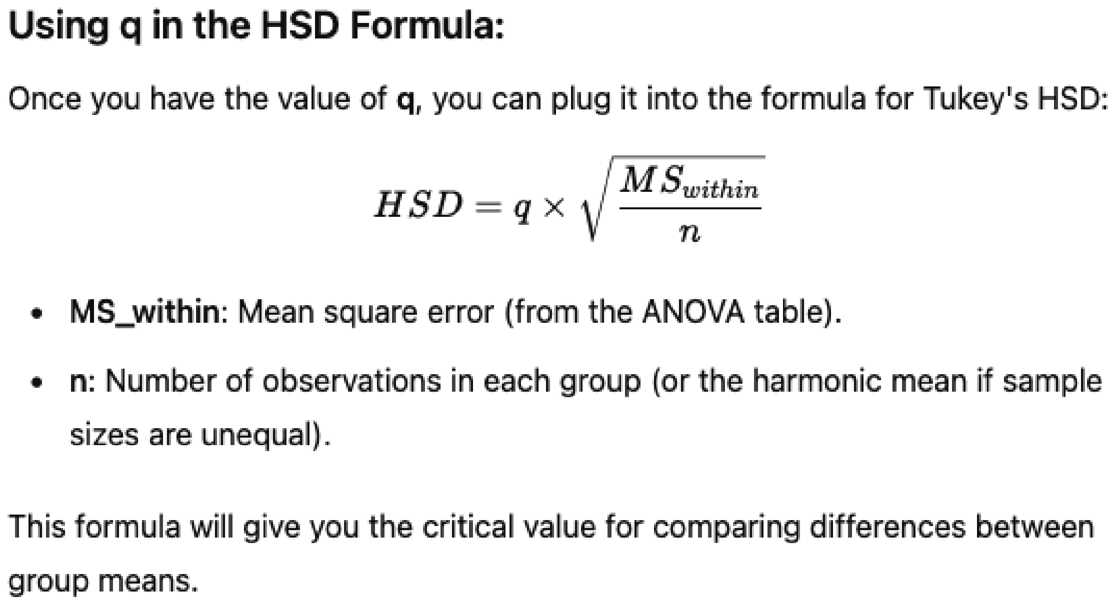


Example:

Suppose you're comparing **4 groups** (**k = 4**) with a **within-group error degrees of freedom (df_e) = 30**, and your significance level is **α = 0.05**.

1. **Number of groups (k)**: 4
2. **Degrees of freedom (df_e)**: 30
3. **Significance level (α)**: 0.05

Now, using a **q-table** or a statistical calculator, look up **q** for **k = 4**, **df_e = 30**, and **α = 0.05**. The critical value of **q** for these inputs would be approximately **3.77**.



Using q in the HSD Formula:

Once you have the value of **q**, you can plug it into the formula for Tukey's HSD:

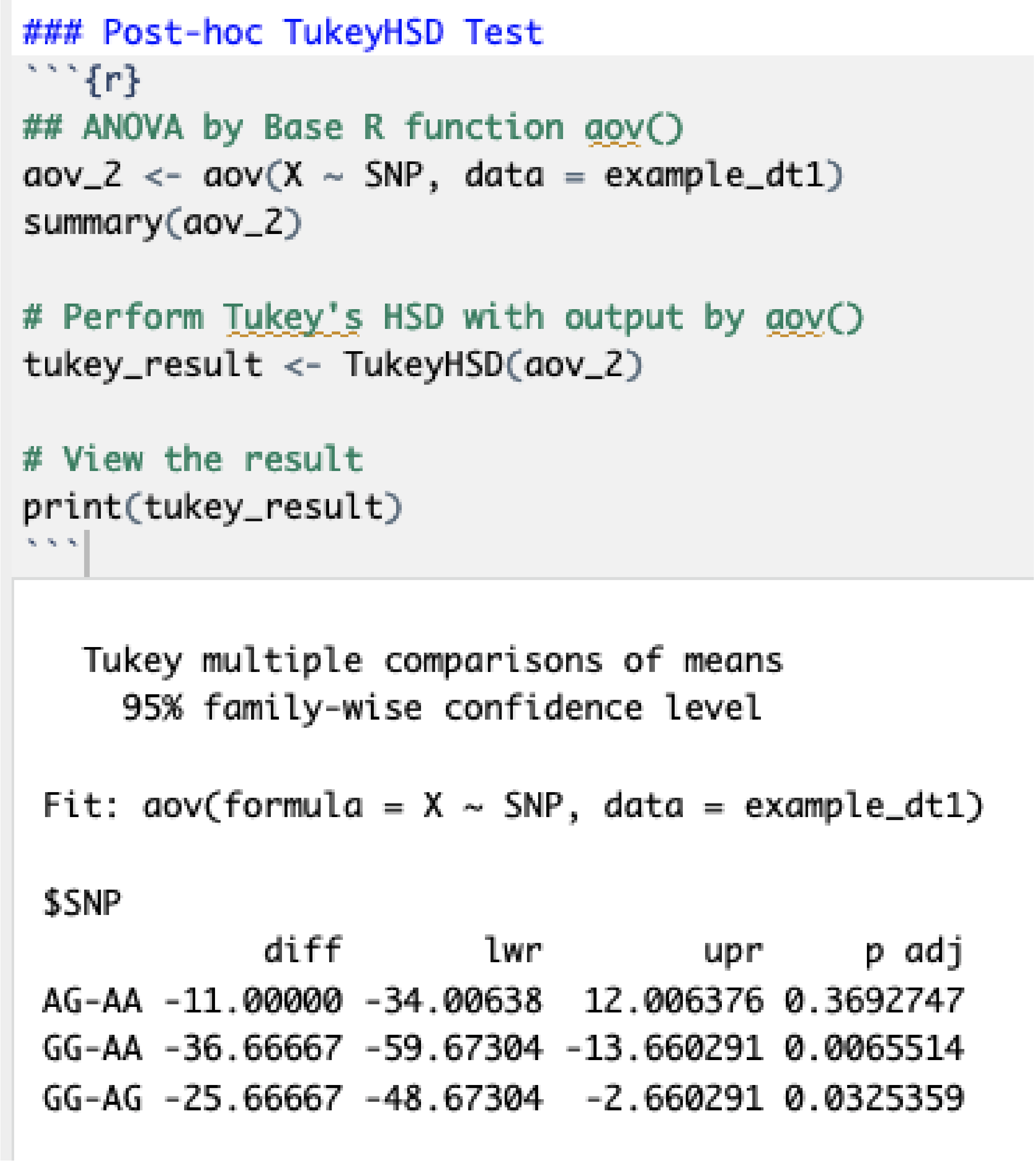
$$HSD = q \times \sqrt{\frac{MS_{within}}{n}}$$

- **MS_within**: Mean square error (from the ANOVA table).
- **n:** Number of observations in each group (or the harmonic mean if sample sizes are unequal).

This formula will give you the critical value for comparing differences between group means.

The critical value of **q** is typically obtained from statistical tables of the **studentized range distribution** (also called **q-distribution**).

Example Tukey’s HSD Test



~~~
### Post-hoc TukeyHSD Test
```{r}
## ANOVA by Base R function aov()
aov_2 <- aov(X ~ SNP, data = example_dt1)
summary(aov_2)

# Perform Tukey's HSD with output by aov()
tukey_result <- TukeyHSD(aov_2)

# View the result
print(tukey_result)
~~~
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = X ~ SNP, data = example_dt1)

$SNP
           diff       lwr        upr     p adj
AG-AA -11.00000 -34.00638  12.006376 0.3692747
GG-AA -36.66667 -59.67304 -13.660291 0.0065514
GG-AG -25.66667 -48.67304  -2.660291 0.0325359
```
```

# Common Mistakes with ANOVA

- Overdesigned. Testing too many factors & levels simultaneously.
- Treating technical replicates as independent.
- Not controlling Family-wise Error Rate (FWER) in post-hoc tests.
- Running Completely Randomized analysis on Related Measures designs
  - Running post-hoc randomized tests with related measures
- Never doing *a priori* power / sample size analysis (Week 11 Lecture)

# Best Practices

1

Make a tidy data frame

- One variable per column. One column for the unique subject ID
- Missing data? Exclude or impute.

2

Visualize the data

- Response vs. Factors

3

Use ez::ezANOVA() or aov()

- Completely Randomized (between) or Related Measures (within)?

4

Post-hoc test as needed by TukeyHSD()

# When data assumptions such as normal distributions cannot be satisfied?

Permutation test (Week 12 Lecture)

# In-Class Exercise 2 :

## Two-way ANOVA