Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on April 13, 2021

2021 Spring

Chapter 13. Randomized Block Designs

Plan

§ 13.1 Introduction

§ 13.2 The F Test for a Randomized Block Design

§ 13.A Appendix: Some Discussions and Extensions

Chapter 13. Randomized Block Designs

§ 13.1 Introduction

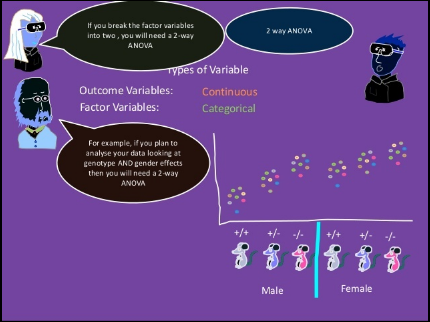
§ 13.2 The F Test for a Randomized Block Design

§ 13.A Appendix: Some Discussions and Extensions

Rationale:

Reducing variability by blocking[†]

[†] *Blocking* is the arranging of experimental units in groups (blocks) that are similar to one another.



If you break the factor variables into two , you will need a 2-way ANOVA

2 way ANOVA

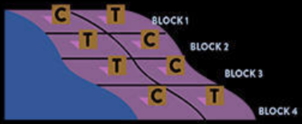
Types of Variable

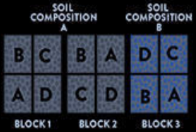
Outcome Variables: Continuous

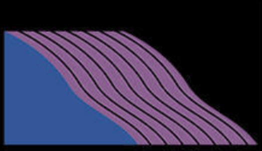
Factor Variables: Categorical

For example, if you plan to analyse your data looking at genotype AND gender effects then you will need a 2-way ANOVA

https://www.slideshare.net/KevinHamill2/
experimental-design-cartoon-part-5-sample-size







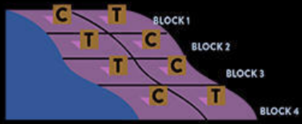
Goal Reducing variability caused by

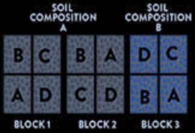
a *elevation*.
b *soil types*. } One-way ANOVA

v.s.

c *complete randomized design* Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design





Goal Reducing variability caused by

- a *elevation.*
- b *soil types.*

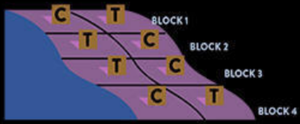
v.s.

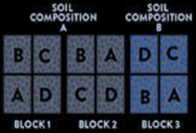
- c *complete randomized design*

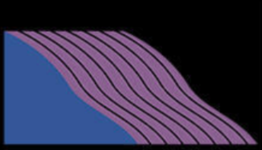
} One-way ANOVA

Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design







Goal Reducing variability caused by

- a *elevation*.
- b *soil types*.

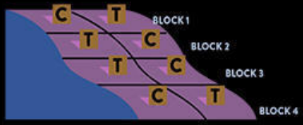
} One-way ANOVA

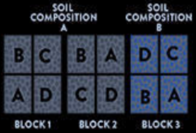
v.s.

- c *complete randomized design*

Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design





Goal Reducing variability caused by

- a *elevation*.
- b *soil types*.

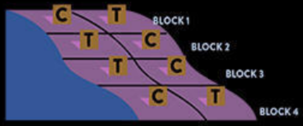
$\}$ One-way ANOVA

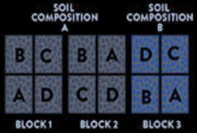
v.s.

- c *complete randomized design*

Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design





Goal Reducing variability caused by

- a *elevation*.
- b *soil types*.

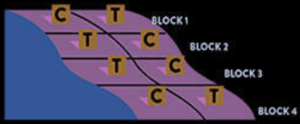
$\}$ One-way ANOVA

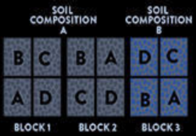
v.s.

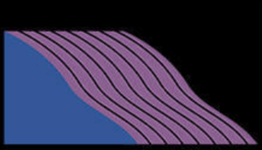
- c *complete randomized design*

Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design







Goal Reducing variability caused by

- a *elevation*.
- b *soil types*.

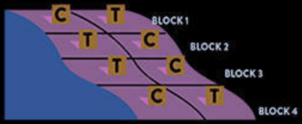
$\}$ One-way ANOVA

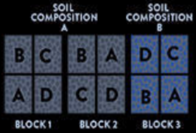
v.s.

- c *complete randomized design*

Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design





Goal Reducing variability caused by

- a *elevation*.
- b *soil types*.

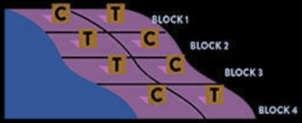
} One-way ANOVA

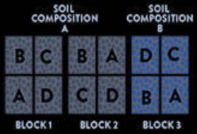
v.s.

- c *complete randomized design*

Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design





Goal Reducing variability caused by

- a *elevation*.
- b *soil types*.

} One-way ANOVA

v.s.

- c *complete randomized design*

Two-way ANOVA

https://www.sare.org/Learning-Center/Bulletins/How-to-Conduct-Research-on-Your-Farm-or-Ranch/Text-Version/Basics-of-Experimental-Design

Plan

§ 13.1 Introduction

§ 13.2 The F Test for a Randomized Block Design

§ 13.A Appendix: Some Discussions and Extensions

Chapter 13. Randomized Block Designs

§ 13.1 Introduction

§ 13.2 The F Test for a Randomized Block Design

§ 13.A Appendix: Some Discussions and Extensions

Setup Y_{ij} indep. $\sim N(\mu_j + \beta_i, \sigma^2)$, i.e., $Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$, ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$

Setup Y_{ij} indep. $\sim N(\mu_j + \beta_i, \sigma^2)$, i.e., $Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$, ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$

Setup Y_{ij} indep. $\sim N(\mu_j + \beta_i, \sigma^2)$, i.e., $Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$, ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$

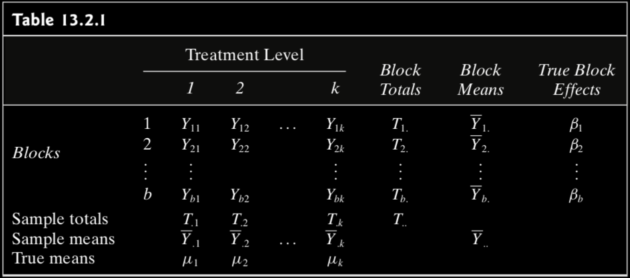


Table 13.2.1

		Treatment Level				*Block Totals*	*Block Means*	*True Block Effects*
		1	*2*		*k*			
Blocks	1	Y_{11}	Y_{12}	$\dots$	Y_{1k}	$T_{1.}$	$\overline{Y}_{1.}$	β_1
	2	Y_{21}	Y_{22}		Y_{2k}	$T_{2.}$	$\overline{Y}_{2.}$	β_2
	$\vdots$	$\vdots$			$\vdots$	$\vdots$	$\vdots$	$\vdots$
	b	Y_{b1}	Y_{b2}		Y_{bk}	$T_{b.}$	$\overline{Y}_{b.}$	β_b
Sample totals		$T_{.1}$	$T_{.2}$		$T_{.k}$	$T_{..}$		
Sample means		$\overline{Y}_{.1}$	$\overline{Y}_{.2}$	$\dots$	$\overline{Y}_{.k}$		$\overline{Y}_{..}$	
True means		μ_1	μ_2		μ_k			

Recall For one-way ANOVA,

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{.j}\right) + \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{.j}\right)^2 + \text{zero cross term} + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{.j}\right)^2 + b\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSTR
\end{aligned}$$

Recall For one-way ANOVA,

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{.j}\right) + \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{.j}\right)^2 + \text{zero cross term} + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{.j}\right)^2 + b\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSTR
\end{aligned}$$

Recall For one-way ANOVA,

$$Y_{ij} = \mu_j + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{.j}\right) + \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{.j}\right)^2 + \text{zero cross term} + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{.j}\right)^2 + b\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSTR
\end{aligned}$$

$$
\begin{array}{ccccc}
SSTOT & = & SSE & + & SSTR \\
 & \Downarrow & & & \\
\dfrac{SSTOT}{\sigma^2} & = & \dfrac{SSE}{\sigma^2} & + & \dfrac{SSTR}{\sigma^2} \\
\wr & & \wr & & \wr \\
\chi^2(bk-1) & & \chi^2(bk-k) & \perp & \chi^2(k-1) \\
\text{Under } H_0 & & \checkmark & & \text{Under } H_0
\end{array}
$$

$$H_0 : \mu_1 = \cdots = \mu_k$$

Symmetry

If

$$Y_{ij} = \beta_i + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{i.}\right) + \left(\overline{Y}_{i.} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + \text{zero cross term} + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + k\sum_{i=1}^{b}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSB
\end{aligned}$$

Symmetry If

$$Y_{ij} = \beta_i + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{i.}\right) + \left(\overline{Y}_{i.} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + \text{zero cross term} + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + k\sum_{i=1}^{b}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSB
\end{aligned}$$

Symmetry If

$$Y_{ij} = \beta_i + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{i.}\right) + \left(\overline{Y}_{i.} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + \text{zero cross term} + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.}\right)^2 + k\sum_{i=1}^{b}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSB
\end{aligned}$$

$$
\begin{array}{ccccc}
SSTOT & = & SSE & + & SSB \\
 & \Downarrow & & & \\
\dfrac{SSTOT}{\sigma^2} & = & \dfrac{SSE}{\sigma^2} & + & \dfrac{SSB}{\sigma^2} \\
\wr & & \wr & & \wr \\
\chi^2(bk-1) & & \chi^2(bk-b) & \perp & \chi^2(b-1) \\
\text{Under } \widetilde{H}_0 & & \checkmark & & \text{Under } \widetilde{H}_0
\end{array}
$$

$$\widetilde{H}_0 : \beta_1 = \cdots = \beta_b$$

Similarly

If

$$Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right) + \left(\overline{Y}_{i.} - \overline{Y}_{..}\right) + \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right)^2 + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&\quad + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 + \text{zero cross terms} \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right)^2 + k\sum_{i=1}^{b}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 + b\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSB + SSTR
\end{aligned}$$

Similarly

If

$$Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right) + \left(\overline{Y}_{i.} - \overline{Y}_{..}\right) + \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right)^2 + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&\quad + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 + \text{zero cross terms} \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right)^2 + k\sum_{i=1}^{b}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 + b\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSB + SSTR
\end{aligned}$$

Similarly If

$$Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$$

$$\Downarrow$$

$$\begin{aligned}
SSTOT &= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{..}\right)^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left[\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right) + \left(\overline{Y}_{i.} - \overline{Y}_{..}\right) + \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)\right]^2 \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right)^2 + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\
&\quad + \sum_{i=1}^{b}\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 + \text{zero cross terms} \\
&= \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}\right)^2 + k\sum_{i=1}^{b}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 + b\sum_{j=1}^{k}\left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 \\
&= SSE + SSB + SSTR
\end{aligned}$$

$$\begin{array}{ccccccc}
SSTOT & = & SSE & + & SSB & + & SSTR \\
 & \Downarrow & & & & & \\
\frac{SSTOT}{\sigma^2} & = & \frac{SSE}{\sigma^2} & + & \frac{SSB}{\sigma^2} & + & \frac{SSTR}{\sigma^2} \\
\wr & & \wr & & \wr & & \wr \\
\chi^2(bk-1) & & \chi^2((k-1)(b-1)) & \perp & \chi^2(b-1) & \perp & \chi^2(k-1) \\
\text{Under } H_0 \text{ or } \widetilde{H}_0 & & \checkmark & & \text{under } \widetilde{H}_0 & & \text{under } H_0
\end{array}$$

$$\widetilde{H}_0 : \beta_1 = \cdots = \beta_b \qquad \text{and} \qquad H_0 : \mu_1 \cdots = \mu_k$$

$$H_0 : \mu_1 \cdots = \mu_k$$

$$\Downarrow$$

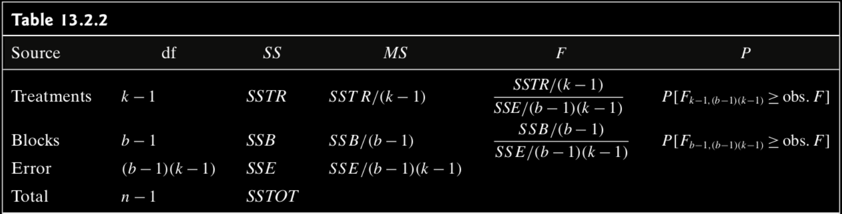


Table 13.2.2

Source	df	SS	MS	F	P
Treatments	$k-1$	$SSTR$	$SSTR/(k-1)$	$\dfrac{SSTR/(k-1)}{SSE/(b-1)(k-1)}$	$P[F_{k-1,(b-1)(k-1)} \geq \text{obs. } F]$
Blocks	$b-1$	SSB	$SSB/(b-1)$	$\dfrac{SSB/(b-1)}{SSE/(b-1)(k-1)}$	$P[F_{b-1,(b-1)(k-1)} \geq \text{obs. } F]$
Error	$(b-1)(k-1)$	SSE	$SSE/(b-1)(k-1)$		
Total	$n-1$	$SSTOT$			

$$\Uparrow$$

$$\widetilde{H}_0 : \beta_1 = \cdots = \beta_b$$

Computing formulas

$$C = \frac{T_{\cdot\cdot}^2}{bk}$$

$$SSTR = b\sum_{j=1}^{k}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)^2 = b\sum_{j=1}^{k}\overline{Y}_{\cdot j}^2 - bk\overline{Y}_{\cdot\cdot}^2 = \frac{1}{b}\sum_{j=1}^{k}\overline{T}_{\cdot j}^2 - C.$$

$$SSB = k\sum_{i=1}^{b}\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2 = k\sum_{i=1}^{b}\overline{Y}_{i\cdot}^2 - bk\overline{Y}_{\cdot\cdot}^2 = \frac{1}{k}\sum_{j=1}^{k}\overline{T}_{i\cdot}^2 - C.$$

$$SSTOT = \sum_{i=1}^{b}\sum_{j=1}^{k}\left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{b}\sum_{j=1}^{k}Y_{ij}^2 - bk\overline{Y}_{\cdot\cdot}^2 = \sum_{i=1}^{b}\sum_{j=1}^{k}Y_{ij}^2 - C.$$

$$SSE = SSTOT - SSTR - SSB$$

E.g. Two methods to test wines: whether these two procedures produce the same measurements?

	DRS-FTIR	Standard
White wine 1	112.9	115.1
White wine 2	123.1	125.6
Red wine 1	135.2	132.4
Red wine 2	140.2	143.7

Test at $\alpha = 0.05$

$$H_0 : \mu_{DRS} = \mu_{STD} \quad \text{v.s.} \quad H_1 : \mu_{DRS} \neq \mu_{STD}$$

and

$$\widetilde{H}_0 : \mu_{W1} = \mu_{W2} = \mu_{R1} = \mu_{R2} \quad \text{v.s.} \quad \widetilde{H}_1 : \text{not equal}$$

E.g. Two methods to test wines: whether these two procedures produce the same measurements?

	DRS-FTIR	Standard
White wine 1	112.9	115.1
White wine 2	123.1	125.6
Red wine 1	135.2	132.4
Red wine 2	140.2	143.7

Test at $\alpha = 0.05$

$$H_0 : \mu_{DRS} = \mu_{STD} \quad \text{v.s.} \quad H_1 : \mu_{DRS} \neq \mu_{STD}$$

and

$$\widetilde{H}_0 : \mu_{W1} = \mu_{W2} = \mu_{R1} = \mu_{R2} \quad \text{v.s.} \quad \widetilde{H}_1 : \text{not equal}$$

```
> # Case Study 13.2.1
> # install.packages(”ggpubr”)
> DIRS <- c(112.9, 123.1, 135.2, 140.2)
> STD <- c(115.1, 125.6, 132.4, 143.7)
> Wines <- c(”W1”, ”W2”, ”R1”, ”R2”)
> # Create a data frame
> my_data <- data.frame(
+  method = rep(c(”DIRS”, ”STD”),
     each =4),
+  types = c(Wines,Wines),
+  concentration = c(DIRS, STD)
+ )
> # Show data
> print(my_data)
  method types concentration
1   DIRS   W1         112.9
2   DIRS   W2         123.1
3   DIRS   R1         135.2
4   DIRS   R2         140.2
5    STD   W1         115.1
6    STD   W2         125.6
7    STD   R1         132.4
8    STD   R2         143.7
```

```
> # Compute t-test with equal variances
> res <- t.test(concentration ~ method,
+               data = my_data,
+               var.equal = TRUE)
> res

  Two Sample t-test

data:  concentration by method
t = -0.15721, df = 6, p-value = 0.8802
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.362 19.662
sample estimates:
mean in group DIRS mean in group STD
            127.85             129.20
```

```
> # Compute t-test with unequal variances
> res <- t.test(concentration ~ method,
+               data = my_data,
+               var.equal = FALSE)
> res

  Welch Two Sample t-test

data:  concentration by method
t = -0.15721, df = 5.9968, p-value = 0.8802
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.3647 19.6647
sample estimates:
mean in group DIRS mean in group STD
            127.85             129.20
```

```
> # The following one-way ANOVA is equivalent
> # to the two-sample t test
> library(car)
> model3 = lm(concentration ~ method,
+             data=my_data)
> Anova(model3)
Anova Table (Type II tests)

Response: concentration
          Sum Sq Df F value Pr(>F)
method      3.64  1  0.0247 0.8802
Residuals 884.87  6
```

1. Classical method
2. Welch approximation
3. one-way ANOVA

⇓

The same answer
(p-value)

Concl. Fail to reject H_0

```
> # Compute t-test with equal variances
> res <- t.test(concentration ~ method,
+               data = my_data,
+               var.equal = TRUE)
> res

  Two Sample t-test

data:  concentration by method
t = -0.15721, df = 6, p-value = 0.8802
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.362 19.662
sample estimates:
mean in group DIRS mean in group STD
            127.85            129.20
```

```
> # Compute t-test with unequal variances
> res <- t.test(concentration ~ method,
+               data = my_data,
+               var.equal = FALSE)
> res

  Welch Two Sample t-test

data:  concentration by method
t = -0.15721, df = 5.9968, p-value = 0.8802
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.3647 19.6647
sample estimates:
mean in group DIRS mean in group STD
            127.85            129.20
```

```
> # The following one-way ANOVA is equivalent
> # to the two-sample t test
> library(car)
> model3 = lm(concentration ~ method,
+             data=my_data)
> Anova(model3)
Anova Table (Type II tests)

Response: concentration
          Sum Sq Df F value Pr(>F)
method      3.64  1  0.0247 0.8802
Residuals 884.87  6
```

1. Classical method
2. Welch approximation
3. one-way ANOVA

⇓

The same answer (p-value)

Concl. Fail to reject H_0

```
> # Compute t−test with equal
      variances
> res <− t.test(concentration ~ method,
+               data = my_data,
+               var.equal = TRUE)
> res

  Two Sample t−test

data:  concentration by method
t = −0.15721, df = 6, p−value = 0.8802
alternative hypothesis: true difference in
      means is not equal to 0
95 percent confidence interval:
 −22.362 19.662
sample estimates:
mean in group DIRS mean in group STD
          127.85            129.20
```

```
> # Compute t−test with unequal
      variances
> res <− t.test(concentration ~ method,
+               data = my_data,
+               var.equal = FALSE)
> res

  Welch Two Sample t−test

data:  concentration by method
t = −0.15721, df = 5.9968, p−value =
      0.8802
alternative hypothesis: true difference in
      means is not equal to 0
95 percent confidence interval:
 −22.3647 19.6647
sample estimates:
mean in group DIRS mean in group STD
          127.85            129.20
```

```
> # The following one−way ANOVA is
      equivalent
> # to the two−sample t test
> library(car)
> model3 = lm(concentration ~ method,
+             data=my_data)
> Anova(model3)
Anova Table (Type II tests)

Response: concentration
          Sum Sq Df F value Pr(>F)
method      3.64  1  0.0247 0.8802
Residuals 884.87 6
```

1. Classical method
2. Welch approximation
3. one-way ANOVA

$\Downarrow$

The same answer
(p-value)

Concl. Fail to reject H_0

```
> # Compute t-test with equal variances
> res <- t.test(concentration ~ method,
+               data = my_data,
+               var.equal = TRUE)
> res

  Two Sample t-test

data:  concentration by method
t = -0.15721, df = 6, p-value = 0.8802
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.362 19.662
sample estimates:
mean in group DIRS mean in group STD
            127.85            129.20
```

```
> # Compute t-test with unequal variances
> res <- t.test(concentration ~ method,
+               data = my_data,
+               var.equal = FALSE)
> res

  Welch Two Sample t-test

data:  concentration by method
t = -0.15721, df = 5.9968, p-value = 0.8802
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.3647 19.6647
sample estimates:
mean in group DIRS mean in group STD
            127.85            129.20
```

```
> # The following one-way ANOVA is equivalent
> # to the two-sample t test
> library(car)
> model3 = lm(concentration ~ method,
+             data=my_data)
> Anova(model3)
Anova Table (Type II tests)

Response: concentration
          Sum Sq Df F value Pr(>F)
method      3.64  1  0.0247 0.8802
Residuals 884.87  6
```

1. Classical method
2. Welch approximation
3. one-way ANOVA

⇓

The same answer
(p-value)

Concl. Fail to reject H_0

```
> # Now let's carry out two−way ANOVA
> library(car)
> model = lm(concentration ~ method + types,
+            data=my_data)
> Anova(model)
Anova Table (Type II tests)

Response: concentration
          Sum Sq Df F value Pr(>F)
method      3.65  1  0.9154 0.409258
types     872.92  3 73.0787 0.002652 **
Residuals  11.94  3
```

```
> # Now let's try one−way ANOVA
> model2 = lm(concentration ~ types,
+             data=my_data)
> Anova(model2)
Anova Table (Type II tests)

Response: concentration
          Sum Sq Df F value Pr(>F)
types     872.92  3  74.657 0.0005739 ***
Residuals  15.59  4
−−−
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Fail to reject H_0
2. Reject $\widetilde{H}_0$

```
> # Now let's carry out two−way ANOVA
> library(car)
> model = lm(concentration ~ method + types,
+              data=my_data)
> Anova(model)
Anova Table (Type II tests)

Response: concentration
           Sum Sq Df F value Pr(>F)
method      3.65  1  0.9154 0.409258
types     872.92  3 73.0787 0.002652 **
Residuals  11.94  3
```

```
> # Now let's try one−way ANOVA
> model2 = lm(concentration ~ types,
+              data=my_data)
> Anova(model2)
Anova Table (Type II tests)

Response: concentration
           Sum Sq Df F value    Pr(>F)
types      872.92  3  74.657 0.0005739 ***
Residuals   15.59  4
−−−
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Fail to reject H_0
2. Reject $\widetilde{H}_0$

E.g. 2 `https://rcompanion.org/rcompanion/d_08.html`

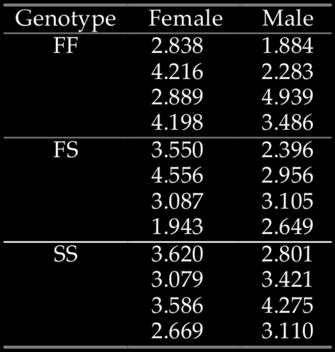
Test at $\alpha = 0.05$

$$H_0 : \mu_F = \mu_M \quad v.s. \quad H_1 : \mu_F \neq \mu_F$$

and

$$\widetilde{H}_0 : \mu_{FF} = \mu_S = \mu_{SS} \quad v.s. \quad \widetilde{H}_1 : \text{not all equal}$$

E.g. 2 `https://rcompanion.org/rcompanion/d_08.html`



Genotype	Female	Male
FF	2.838	1.884
	4.216	2.283
	2.889	4.939
	4.198	3.486
FS	3.550	2.396
	4.556	2.956
	3.087	3.105
	1.943	2.649
SS	3.620	2.801
	3.079	3.421
	3.586	4.275
	2.669	3.110

Test at $\alpha = 0.05$

$$H_0 : \mu_F = \mu_M \quad v.s. \quad H_1 : \mu_F \neq \mu_F$$

and

$$\widetilde{H}_0 : \mu_{FF} = \mu_S = \mu_{SS} \quad v.s. \quad \widetilde{H}_1 : \text{not all equal}$$

```
> Data
   id    Sex Genotype Activity
1   1   male       ff    1.884
2   2   male       ff    2.283
3   3   male       fs    2.396
4   4 female       ff    2.838
5   5   male       fs    2.956
6   6 female       ff    4.216
7   7 female       ss    3.620
8   8 female       ff    2.889
9   9 female       fs    3.550
10 10   male       fs    3.105
11 11 female       fs    4.556
12 12 female       fs    3.087
13 13   male       ff    4.939
14 14   male       ff    3.486
15 15 female       ss    3.079
16 16   male       fs    2.649
```

```
17 17 female       fs    1.943
18 19 female       ff    4.198
19 20 female       ff    2.473
20 22 female       ff    2.033
21 24 female       fs    2.200
22 25 female       fs    2.157
23 26   male       ss    2.801
24 28   male       ss    3.421
25 29 female       ff    1.811
26 30 female       fs    4.281
27 32 female       fs    4.772
28 34 female       ss    3.586
29 36 female       ff    3.944
30 38 female       ss    2.669
31 39 female       ss    3.050
32 41   male       ss    4.275
33 43 female       ss    2.963
34 46 female       ss    3.236
35 48 female       ss    3.673
36 49   male       ss    3.110
```

```
> # Two-way ANOVA
> model = lm(Activity ~ Sex + Genotype,
+            data=Data)
> Anova(model, type="II")
Anova Table (Type II tests)

Response: Activity
           Sum Sq Df F value Pr(>F)
Sex        0.0681  1  0.0888 0.7676
Genotype   0.2772  2  0.1808 0.8354
Residuals 24.5285 32
> # One-way ANOVA
> model_Sex = lm(Activity ~ Sex,
+                data=Data)
> Anova(model_Sex, type="II")
Anova Table (Type II tests)

Response: Activity
           Sum Sq Df F value Pr(>F)
Sex        0.0681  1  0.0933 0.7619
Residuals 24.8057 34
> # One-way ANOVA
> model_Genotype = lm(Activity ~ Genotype,
+                data=Data)
> Anova(model_Genotype, type="II")
Anova Table (Type II tests)

Response: Activity
           Sum Sq Df F value Pr(>F)
Genotype   0.2772  2   0.186 0.8312
Residuals 24.5965 33
```

Tuckey’s pairwise comparison

Replace $Q_{\alpha,k,b(k-k)}$ by $Q_{\alpha,k,(b-1)(k-1)}$

```
> # Tukey’s pairwise comparison (One-
      way)
> model1 = aov(Activity ~ Genotype,
+                      data=Data)
> TukeyHSD(model1, ”Genotype”,
      ordered = TRUE)
  Tukey multiple comparisons of means
    95% family-wise confidence level
    factor levels have been ordered

Fit: aov(formula = Activity ~ Genotype,
       data = Data)

$Genotype
              diff        lwr       upr      p
                adj
fs-ff 0.05483333 -0.8100204 0.919687
      0.9867505
ss-ff 0.20741667 -0.6574370 1.072270
      0.8272105
ss-fs 0.15258333 -0.7122704 1.017437
      0.9021607
```

```
> # Tukey’s pairwise comparison (Two
      -way)
> model2 = aov(Activity ~ Sex +
      Genotype,
+             data=Data)
> TukeyHSD(model2, ”Genotype”,
      ordered = TRUE)
  Tukey multiple comparisons of means
    95% family-wise confidence level
    factor levels have been ordered

Fit: aov(formula = Activity ~ Sex +
      Genotype, data = Data)

$Genotype
              diff        lwr        upr     p
                adj
fs-ff 0.05483333 -0.8234920 0.9331586
      0.987114
ss-ff 0.20741667 -0.6709086 1.0857420
      0.831554
ss-fs 0.15258333 -0.7257420 1.0309086
      0.904729
```

Remark By two-way ANOVA, or through blocking one factor, we obtain

1. larger p-values:

   more conservative to reject H_0.

2. wider C.I.'s:

   more conservative on our estimates.

Remark By two-way ANOVA, or through blocking one factor, we obtain

1. larger p-values:
   more conservative to reject H_0.

2. wider C.I.'s:
   more conservative on our estimates.

Remark By two-way ANOVA, or through blocking one factor, we obtain

1. larger p-values:
   more conservative to reject H_0.
2. wider C.I.'s:
   more conservative on our estimates.

Remark By two-way ANOVA, or through blocking one factor, we obtain

1. larger p-values:
   more conservative to reject H_0.

2. wider C.I.'s:
   more conservative on our estimates.

Remark By two-way ANOVA, or through blocking one factor, we obtain

1. larger p-values:
   more conservative to reject H_0.
2. wider C.I.'s:
   more conservative on our estimates.

Plan

§ 13.1 Introduction

§ 13.2 The F Test for a Randomized Block Design

§ 13.A Appendix: Some Discussions and Extensions

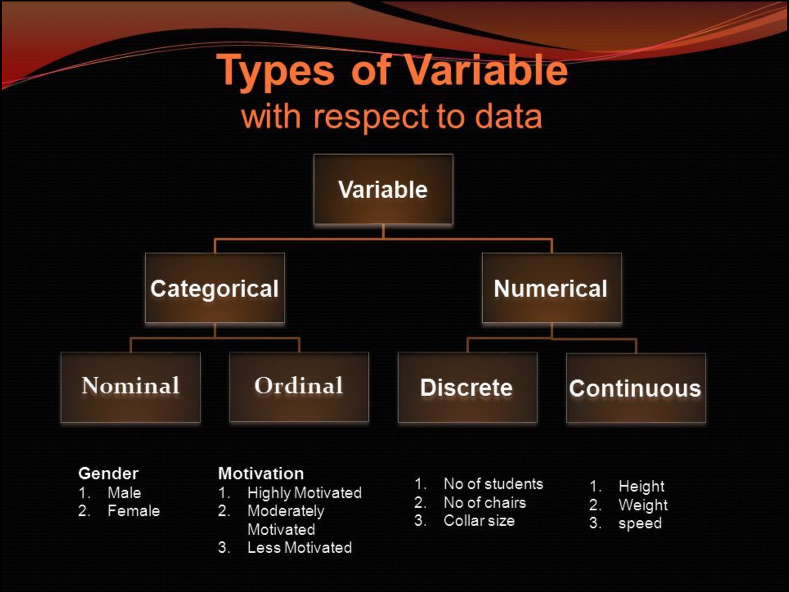
Chapter 13. Randomized Block Designs

§ 13.1 Introduction

§ 13.2 The F Test for a Randomized Block Design

§ 13.A Appendix: Some Discussions and Extensions

§ 13.A Appendix: Some Discussions and Extensions



Types of Variable

with respect to data

Variable

- Categorical
  - Nominal
  - Ordinal
- Numerical
  - Discrete
  - Continuous

Gender
1. Male
2. Female

Motivation
1. Highly Motivated
2. Moderately Motivated
3. Less Motivated

1. No of students
2. No of chairs
3. Collar size

1. Height
2. Weight
3. speed

	Numerical values	Categorical values
Sample 1	$y_{11} \cdots y_{1m}$	$c_{11} \cdots c_{1p}$
Sample 2	$y_{21} \cdots y_{2m}$	$c_{21} \cdots c_{2p}$
$\vdots$	$\vdots \quad \vdots$	$\vdots \quad \vdots$
Sample n	$y_{n1} \cdots y_{nm}$	$c_{n1} \cdots c_{np}$

n samples

m

p

		Dependent variable	
		Continuous	Categorical
Independent variable	Continuous	Linear Regression	Logistic regression
	Categorical	t-test/ANOVA	Chi-square test

Indep. v.s. Dependent

1. Categorical v.s. Continuous

1.1 $p = 1, m = 1,$	One-way ANOVA
1.2 $p = 2, m = 1,$	Two-way ANOVA
1.3 $p \geq 3, m = 1,$	p-way ANOVA
1.4 $p = 1, m \geq 2,$	One-way MANOVA[a]
1.5 $p = 2, m \geq 2,$	Two-way MANOVA
1.6 $p \geq 3, m \geq 2,$	p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

2.1 $m_{ind} = 1, m_{dep} = 1,$	Simple linear regression
2.2 $m_{ind} \geq 2$	Multiple linear regression
2.3 $m_{dep} \geq 2$	Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous
    1.1 $p = 1, m = 1,$ One-way ANOVA
    1.2 $p = 2, m = 1,$ Two-way ANOVA
    1.3 $p \geq 3, m = 1,$ p-way ANOVA

    1.4 $p = 1, m \geq 2,$ One-way MANOVA[a]
    1.5 $p = 2, m \geq 2,$ Two-way MANOVA
    1.6 $p \geq 3, m \geq 2,$ p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous
    2.1 $m_{ind} = 1, m_{dep} = 1,$ Simple linear regression
    2.2 $m_{ind} \geq 2$ Multiple linear regression
    2.3 $m_{dep} \geq 2$ Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous
   - 1.1 $p = 1$, $m = 1$, One-way ANOVA
   - 1.2 $p = 2$, $m = 1$, Two-way ANOVA
   - 1.3 $p \geq 3$, $m = 1$, p-way ANOVA

   - 1.4 $p = 1$, $m \geq 2$, One-way MANOVA[a]
   - 1.5 $p = 2$, $m \geq 2$, Two-way MANOVA
   - 1.6 $p \geq 3$, $m \geq 2$, p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous
   - 2.1 $m_{ind} = 1$, $m_{dep} = 1$, Simple linear regression
   - 2.2 $m_{ind} \geq 2$ Multiple linear regression
   - 2.3 $m_{dep} \geq 2$ Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous

1.1 $p = 1$, $m = 1$,	One-way ANOVA
1.2 $p = 2$, $m = 1$,	Two-way ANOVA
1.3 $p \geq 3$, $m = 1$,	p-way ANOVA
1.4 $p = 1$, $m \geq 2$,	One-way MANOVA[a]
1.5 $p = 2$, $m \geq 2$,	Two-way MANOVA
1.6 $p \geq 3$, $m \geq 2$,	p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

2.1 $m_{ind} = 1$, $m_{dep} = 1$,	Simple linear regression
2.2 $m_{ind} \geq 2$	Multiple linear regression
2.3 $m_{dep} \geq 2$	Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous
   - 1.1 $p = 1$, $m = 1$, One-way ANOVA
   - 1.2 $p = 2$, $m = 1$, Two-way ANOVA
   - 1.3 $p \geq 3$, $m = 1$, p-way ANOVA

   - 1.4 $p = 1$, $m \geq 2$, One-way MANOVA[a]
   - 1.5 $p = 2$, $m \geq 2$, Two-way MANOVA
   - 1.6 $p \geq 3$, $m \geq 2$, p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous
   - 2.1 $m_{ind} = 1$, $m_{dep} = 1$, Simple linear regression
   - 2.2 $m_{ind} \geq 2$ Multiple linear regression
   - 2.3 $m_{dep} \geq 2$ Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous

1.1 $p = 1,\ m = 1,$	One-way ANOVA
1.2 $p = 2,\ m = 1,$	Two-way ANOVA
1.3 $p \geq 3,\ m = 1,$	p-way ANOVA
1.4 $p = 1,\ m \geq 2,$	One-way MANOVA[a]
1.5 $p = 2,\ m \geq 2,$	Two-way MANOVA
1.6 $p \geq 3,\ m \geq 2,$	p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

2.1 $m_{ind} = 1,\ m_{dep} = 1,$	Simple linear regression
2.2 $m_{ind} \geq 2$	Multiple linear regression
2.3 $m_{dep} \geq 2$	Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous

1.1	$p = 1,\ m = 1,$	One-way ANOVA
1.2	$p = 2,\ m = 1,$	Two-way ANOVA
1.3	$p \geq 3,\ m = 1,$	p-way ANOVA
1.4	$p = 1,\ m \geq 2,$	One-way MANOVA[a]
1.5	$p = 2,\ m \geq 2,$	Two-way MANOVA
1.6	$p \geq 3,\ m \geq 2,$	p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

2.1	$m_{ind} = 1,\ m_{dep} = 1,$	Simple linear regression
2.2	$m_{ind} \geq 2$	Multiple linear regression
2.3	$m_{dep} \geq 2$	Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous
    - 1.1 $p = 1,\ m = 1,$ One-way ANOVA
    - 1.2 $p = 2,\ m = 1,$ Two-way ANOVA
    - 1.3 $p \geq 3,\ m = 1,$ p-way ANOVA

    - 1.4 $p = 1,\ m \geq 2,$ One-way MANOVA[a]
    - 1.5 $p = 2,\ m \geq 2,$ Two-way MANOVA
    - 1.6 $p \geq 3,\ m \geq 2,$ p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous
    - 2.1 $m_{ind} = 1,\ m_{dep} = 1,$ Simple linear regression
    - 2.2 $m_{ind} \geq 2$ Multiple linear regression
    - 2.3 $m_{dep} \geq 2$ Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous
   - 1.1 $p = 1,\ m = 1,$ One-way ANOVA
   - 1.2 $p = 2,\ m = 1,$ Two-way ANOVA
   - 1.3 $p \geq 3,\ m = 1,$ p-way ANOVA

   - 1.4 $p = 1,\ m \geq 2,$ One-way MANOVA[a]
   - 1.5 $p = 2,\ m \geq 2,$ Two-way MANOVA
   - 1.6 $p \geq 3,\ m \geq 2,$ p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous
   - 2.1 $m_{ind} = 1,\ m_{dep} = 1.$ Simple linear regression
   - 2.2 $m_{ind} \geq 2$ Multiple linear regression
   - 2.3 $m_{dep} \geq 2$ Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous

1.1 $p = 1$, $m = 1$,	One-way ANOVA
1.2 $p = 2$, $m = 1$,	Two-way ANOVA
1.3 $p \geq 3$, $m = 1$,	p-way ANOVA
1.4 $p = 1$, $m \geq 2$,	One-way MANOVA[a]
1.5 $p = 2$, $m \geq 2$,	Two-way MANOVA
1.6 $p \geq 3$, $m \geq 2$,	p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

2.1 $m_{ind} = 1$, $m_{dep} = 1$,	Simple linear regression
2.2 $m_{ind} \geq 2$	Multiple linear regression
2.3 $m_{dep} \geq 2$	Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous

   1.1 $p = 1$, $m = 1$, One-way ANOVA

   1.2 $p = 2$, $m = 1$, Two-way ANOVA

   1.3 $p \geq 3$, $m = 1$, p-way ANOVA

   1.4 $p = 1$, $m \geq 2$, One-way MANOVA[a]

   1.5 $p = 2$, $m \geq 2$, Two-way MANOVA

   1.6 $p \geq 3$, $m \geq 2$, p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

   2.1 $m_{ind} = 1$, $m_{dep} = 1$, Simple linear regression

   2.2 $m_{ind} \geq 2$ Multiple linear regression

   2.3 $m_{dep} \geq 2$ Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous

1.1 $p = 1,\ m = 1,$	One-way ANOVA
1.2 $p = 2,\ m = 1,$	Two-way ANOVA
1.3 $p \geq 3,\ m = 1,$	p-way ANOVA
1.4 $p = 1,\ m \geq 2,$	One-way MANOVA[a]
1.5 $p = 2,\ m \geq 2,$	Two-way MANOVA
1.6 $p \geq 3,\ m \geq 2,$	p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

2.1 $m_{ind} = 1,\ m_{dep} = 1,$	Simple linear regression
2.2 $m_{ind} \geq 2$	Multiple linear regression
2.3 $m_{dep} \geq 2$	Multivariate linear regression

Indep. v.s. Dependent

1. Categorical v.s. Continuous
    - 1.1 $p = 1$, $m = 1$, One-way ANOVA
    - 1.2 $p = 2$, $m = 1$, Two-way ANOVA
    - 1.3 $p \geq 3$, $m = 1$, p-way ANOVA

    - 1.4 $p = 1$, $m \geq 2$, One-way MANOVA[a]
    - 1.5 $p = 2$, $m \geq 2$, Two-way MANOVA
    - 1.6 $p \geq 3$, $m \geq 2$, p-way ANOVA

2. Continuous v.s. Continuous
    - 2.1 $m_{ind} = 1$, $m_{dep} = 1$, Simple linear regression
    - 2.2 $m_{ind} \geq 2$ Multiple linear regression
    - 2.3 $m_{dep} \geq 2$ Multivariate linear regression

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

Indep. v.s. Dependent

1. Categorical v.s. Continuous

1.1	$p = 1$, $m = 1$,	One-way ANOVA
1.2	$p = 2$, $m = 1$,	Two-way ANOVA
1.3	$p \geq 3$, $m = 1$,	p-way ANOVA
1.4	$p = 1$, $m \geq 2$,	One-way MANOVA[a]
1.5	$p = 2$, $m \geq 2$,	Two-way MANOVA
1.6	$p \geq 3$, $m \geq 2$,	p-way ANOVA

---

[a] MANOVA refers to the multivariate analysis of variance
ANOVA refers to the univariate analysis of variance.

2. Continuous v.s. Continuous

2.1	$m_{ind} = 1$, $m_{dep} = 1$,	Simple linear regression
2.2	$m_{ind} \geq 2$	Multiple linear regression
2.3	$m_{dep} \geq 2$	Multivariate linear regression

E.g. One example for MANOVA[1].

---

[1] `http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance`

E.g. One example for MANOVA[1].

The Iris Dataset

Collected by Ronald Fisher in 1936

---

[1] `http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance`

Iris Versicolor **Iris Setosa** **Iris Virginica**

```
> library(datasets)
> data(iris)
> summary(iris)
  Sepal.Length  Sepal.Width   Petal.Length  Petal.Width
       Species
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa
       :50
 1st Qu.:5.100  1st Qu.:2.800  1st Qu.:1.600  1st Qu.:0.300  versicolor
       :50
 Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica
       :50
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
> my_data <- iris
> my_data
    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1            5.1         3.5          1.4         0.2  setosa
2            4.9         3.0          1.4         0.2  setosa
3            4.7         3.2          1.3         0.2  setosa
4            4.6         3.1          1.5         0.2  setosa
5            5.0         3.6          1.4         0.2  setosa
6            5.4         3.9          1.7         0.4  setosa
7            4.6         3.4          1.4         0.3  setosa
8            5.0         3.4          1.5         0.2  setosa
9            4.4         2.9          1.4         0.2  setosa
10           4.9         3.1          1.5         0.1  setosa
```

```
> # Compute MAOVA test now
> res.man <- manova(cbind(Sepal.Length, Petal.Length) ~ Species, data =
      iris)
> summary(res.man)
           Df Pillai approx F num Df den Df Pr(>F)
Species     2 0.9885   71.829      4    294 < 2.2e-16 ***
Residuals 147
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # Look to see which differ
> summary.aov(res.man)
 Response Sepal.Length :
             Df Sum Sq Mean Sq F value Pr(>F)
Species       2 63.212 31.606 119.26 < 2.2e-16 ***
Residuals   147 38.956  0.265
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response Petal.Length :
             Df Sum Sq Mean Sq F value Pr(>F)
Species       2 437.10 218.551 1180.2 < 2.2e-16 ***
Residuals   147  27.22   0.185
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1:w
```

Concl.: Two variables are highly significantly different among species.