**Randomized complete block (RCB) design**

We continue to compare more than two population means but now the samples are dependent in some manner.

Treatments are now applied to experimental units within a group or “block”. We saw this same idea before where samples were taken in a “paired” manner for two populations.

For example, “Which grocery store has lower prices on average?” was answered by sampling the SAME items at two different stores. This allowed for a more fair comparison because the same items were being compared. To put this in a more statistical wording, we are removing a source of variability in prices by sampling the same items rather than potentially different items.

Now, this pairing is extended to include more than two items which we now will call a “block”. The reason why this type of design is better than a CRD is it removes some of the variability that would be measured in SSE. We will have a nice demonstration of this later. Why is having a smaller SSE beneficial?

Look at F = MST/MSE

Examples:

1. Wheat varieties (location = block): This is a **modified** example from earlier in Section 6.4.

Which variety of hard red winter wheat (Newton, Roughrider, Tam 105, Colt, Arkan, Scout 66, and Kharkof) provides higher yield per acre on average for southwest Nebraska? Ten locations around southwest Nebraska have a plot of each wheat variety planted.

Location #1 may look like:



Location #7 may look like:



The data set could look like:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Wheat Variety** | | | | | | |
| Location | Newton | Roughrider | Tam 105 | Colt | Arkan | Scout 66 | Khakof |
| 1 | y11 | y12 | y13 | y14 | y15 | y16 | y17 |
| 2 | y21 | y22 | y23 | y24 | y25 | y26 | y27 |
| 3 | Y31 | y32 | y33 | y34 | y35 | y36 | y37 |
|  |  |  |  |  |  |  |  |
| 10 | y10,1 | y10, 2 | y10,3 | y10, 4 | y10, 5 | y10, 6 | y10, 7 |

where yij = observed yield per acre at location i for variety j.

Using our previous terminology:

* Response variable: Yield
* Factor: Wheat variety
* Factor levels: Newton, Roughrider, Tam 105, Colt, Arkan, Scout 66, and Kharkof
* Treatments: Newton, Roughrider, Tam 105, Colt, Arkan, Scout 66, and Kharkof
* Experimental unit: A plot of land at a location
* Block: Location

Notes:

* By using different locations in southwest Nebraska, inferences to the population of southwest Nebraska’s farms can be made.
* By recognizing that each location is a component of the variability in yield, a separate factor for location can be added to the ANOVA table. This factor will be referred to as a “block”.

1. Golf balls (golfballs.csv): Which type of golf ball travels the farthest when struck by a driver: Maxfli, Top-Flite, Titleist-B, or Titleist-P? **Distance C**

There are **many** factors that golfers use in deciding which is the “best” golf ball. One of these factors is which ball will travel the farthest.

For example, on Top-Flite’s website a few years ago (the web page no longer exists), I came across the following comparison:

|  |  |
| --- | --- |
|  | Length using driver |
| **Top-Flite** | Longest |
| **Titleist B** | -3 yards |
| **Titleist P** | -6 yards |
| **Maxfli** | -4 yards |

Unfortunately, Top-Flite did not provide the actual data to support their claim. Also, Top-Flite did not give any explanation about how they conducted their experiment that produced these results.

Despite not having their specific details, the web page inspired me to construct what would be an appropriate experimental design for this setting. Also, I simulated what a data set may potentially look like in this setting, which may lead to my conclusions differing from Top-Flite’s conclusions. The experimental design is as follows:

Twenty amateur golfers of varying abilities were selected as a random sample from the population of golfers (I am assuming that Top-Flite did not use robotic arms for their testing). Each golfer used a driver to strike each type of golf ball in a random order. The distance each ball traveled (carry distance, does not include roll) is recorded in yards. The experiment was conducted under ideal golfing conditions (72°F, no wind, 30% humidity, mostly sunny). The table below shows the random order each golfer hit a particular type of golf ball:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Golfer | **1** | Titleist-B | Maxfli | Titleist-P | Top-Flite |
| **2** | Titleist-P | Titleist-B | Top-Flite | Maxfli |
|  |  |  |  |  |
| **20** | Titleist-P | Top-Flite | Titleist-B | Maxfli |

The data “collected” is summarized in the table below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Golf Ball | | | |
|  |  | **Titleist-B** | **Maxfli** | **Titleist-P** | **Top-Flite** |
| **Golfer** | **1** | y11=190 | y12=211 | y13=203 | y14=205 |
| **2** | y21=161 | y22=170 | 166 | 170 |
| **3** | 250 | 262 | 232 | 255 |
|  |  |  |  |  |
| **20** | y20,1=167 | 186 | 172 | 182 |

Using our previous terminology:

* Response variable: Distance
* Factor: Golf ball type
* Factor levels: Titleist-B, Maxfli, Titleist-P, and Top-Flite
* Treatments: Titleist-B, Maxfli, Titleist-P, and Top-Flite
* Experimental unit: A golf ball
* Block: A golfer

Notes:

* By using different golfers, inferences to the population of amateur golfers can be made.
* By recognizing that each golfer is a component of the distance variability, a separate factor for golfer can be added to the ANOVA table. This factor will be referred to as a “block”.

Example: Golf balls (golfballs.R, golfballs.csv)

Read in the data:

> gb <- read.csv(file = "golfballs.csv")

> gb

Person TitleistB Maxfli TitleistP TopFlite

1 1 190 211 203 205

2 2 161 170 166 170

3 3 250 262 232 255

4 4 153 201 180 195

5 5 180 185 182 179

6 6 213 221 215 216

7 7 270 265 272 267

8 8 199 201 205 200

9 9 195 212 197 205

10 10 167 178 174 175

11 11 190 223 207 201

12 12 165 163 177 156

13 13 253 265 225 260

14 14 165 206 186 210

15 15 184 194 200 170

16 16 219 220 226 215

17 17 283 274 293 275

18 18 208 213 202 199

19 19 185 214 203 199

20 20 167 186 172 182

While the above data structure is convenient, it is not in the correct format for the aov()function Instead, we need the data in a format where columns represent the response, factor, and blocking variables:

> gb2 <- rbind(data.frame(Person = gb$Person, Type =

"TitleistB", Distance = gb$TitleistB),

data.frame(Person = gb$Person, Type = "Maxfli",

Distance = gb$Maxfli),

data.frame(Person = gb$Person, Type = "TitleistP",

Distance = gb$TitleistP),

data.frame(Person = gb$Person, Type = "TopFlite",

Distance = gb$TopFlite))

> head(gb2)

Person Type Distance

1 1 TitleistB 190

2 2 TitleistB 161

3 3 TitleistB 250

4 4 TitleistB 153

5 5 TitleistB 180

6 6 TitleistB 213

> tail(gb2)

Person Type Distance

75 15 TopFlite 170

76 16 TopFlite 215

77 17 TopFlite 275

78 18 TopFlite 199

79 19 TopFlite 199

80 20 TopFlite 182

There are other ways in R to “reshape” the data in this manner (see the reshape() and stack() functions).

Below is a plot of the data:

> boxplot(formula = Distance ~ Type, data = gb2, main =

"Box and dot plot", ylab = "Distance", xlab = "Golf

ball", pars = list(outpch=NA), col = NA)

> stripchart(x = gb2$Distance ~ gb2$Type, lwd = 2, col =

"red", method = "jitter", vertical = TRUE, pch = 1, add

= TRUE)



Does the plot suggest mean differences exist among the golf ball types?

Block was not represented in the above plot. Thus, this plot is kind of the equivalent to analyzing the data with a CRD. We can take into account the blocking information by subtracting off the mean block values for each response. Below is the code used to find “block adjusted” distances and a new version of the plot:

> distance.block.adj <- aggregate(x = Distance ~

Person, data = gb2, FUN = mean)

> distance.block.adj

Added after video recording: R has changed the syntax for aggregate(). In the video, I show formula = Distance ~ Person. Now, the proper syntax is x = Distance ~ Person. I made the correction here and in the program.

Person Distance

1 1 202.25

2 2 166.75

3 3 249.75

4 4 182.25

5 5 181.50

6 6 216.25

7 7 268.50

8 8 201.25

9 9 202.25

10 10 173.50

11 11 205.25

12 12 165.25

13 13 250.75

14 14 191.75

15 15 187.00

16 16 220.00

17 17 281.25

18 18 205.50

19 19 200.25

20 20 176.75

> gb3 <- merge(x = gb2, y = distance.block.adj, by.x =

"Person", by.y = "Person", suffixes =

c("",".person"))

> head(gb3)

Person Type Distance Distance.person

1 1 TitleistB 190 202.25

2 1 TitleistP 203 202.25

3 1 Maxfli 211 202.25

4 1 TopFlite 205 202.25

5 2 TitleistP 166 166.75

6 2 TopFlite 170 166.75

> tail(gb3)

Person Type Distance Distance.person

75 19 TitleistB 185 200.25

76 19 Maxfli 214 200.25

77 20 TitleistB 167 176.75

78 20 TitleistP 172 176.75

79 20 Maxfli 186 176.75

80 20 TopFlite 182 176.75

> gb3$Distance.adj <- gb3$Distance - gb3$Distance.person

> head(gb3)

Person Type Distance Distance.person Distance.adj

1 1 TitleistB 190 202.25 -12.25

2 1 TitleistP 203 202.25 0.75

3 1 Maxfli 211 202.25 8.75

4 1 TopFlite 205 202.25 2.75

5 2 TitleistP 166 166.75 -0.75

6 2 TopFlite 170 166.75 3.25

> boxplot(formula = Distance.adj ~ Type, data = gb3,

main = "Box and dot plot", ylab = "Distance

adjusted for person",xlab = "Golf ball", pars =

list(outpch=NA), col = NA)

> stripchart(x = gb3$Distance.adj ~ gb3$Type, lwd = 2,

col = "red", method = "jitter", vertical = TRUE,

pch = 1, add = TRUE)



Taking into account the person (block) dramatically changes how this plot looks. This plot shows the data with the block variation removed. Does the plot suggest mean differences exist among the golf ball types?

ANOVA Table:

Summarizes the necessary calculations:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source** | **SS** | **df** | **MS** | **F** |
| Treatments | SST | t – 1 | MST | MST/MSE |
| Blocks | SSB | b – 1 | MSB | MSB/MSE |
| Error | SSE | (t – 1)(b – 1) | MSE |  |
| Total | TSS | bt –1 |  |  |

where

* b = number of blocks
* t = number of treatments
* bt = total sample size
* SSB = sum of squares for blocks =  with  as the mean for block j
* MSB = mean square for blocks = SSB/(b – 1)
* SST = 
* MST = SST/(t – 1)
* SSE = 
* MSE = SSE/[(t – 1)(b – 1)]
* TSS = 

Notes:

* TSS = SSB + SST + SSE can be shown through using algebra.
* TSS is how we would usually calculate the numerator of the sample variance! Thus, we are just partitioning the total variability in the experiment. In other words, we are examining the different sources of variation in the experiment.
* Make sure to know the relationships between quantities in the ANOVA table!

When μ1 = = μt, one can show that the expected values of the random variable versions of MSE and MST are both .

When μ1 = = μt is NOT true, the expected value of the random variable version of MSE is still . However, the expected value of the random variable version of MST is now GREATER THAN . The amount that it is greater than grows as the differences among the means grows.

Thus, it is reasonable then to use Fobs = MST/MSE as a measure against μ1 = = μt. In fact, one can show that that the random variable version of MST/MSE has an F distribution if μ1 = = μt. Thus, we can make statements such as



where X has an F distribution with ν1 = t – 1 and ν2 = (t – 1)(b – 1) degrees of freedom. If we observe an Fobs value that is “large” relative to the F distribution, this leads us to believe that μ1 = = μt is NOT true. We then use  as the critical value for a hypothesis test of

Ho: μ1 = μ2 = … = μt (No difference in pop. means)

Ha: At least one pair of means are unequal

Test Statistic Method:

1. Ho: μ1 = μ2 = … = μt  
   Ha: At least one pair of means are unequal
2. Test statistic: Fobs = MST/MSE
3. Critical Value: 
4. If Fobs >  reject Ho; otherwise don’t reject Ho
5. Conclusion

Reject Ho: There is at least one pair of \_\_\_\_ means that are different.

Don’t Reject Ho: There is not sufficient evidence to conclude that any of the \_\_\_\_ means are different.

where \_\_\_\_ means to insert the factor under investigation

P-value Method:

1. Ho: μ1 = μ2 = … = μt  
   Ha: At least one pair of means are unequal
2. P-value is P(X > Fobs) where X is a random variable from a F distribution with ν1 = t – 1 and ν2 = (t – 1)(b – 1) degrees of freedom
3. State α
4. Reject Ho if p-value < α; otherwise don’t reject Ho
5. Conclusion

A similar type of hypothesis test could be performed with respect to the block means. This type of test is usually not of interest, so we will not discuss it here.

Example: Golf balls (golfballs.R, golfballs.csv)

ANOVA calculations:

> mod.fit <- aov(formula = Distance ~ Type +

factor(Person), data = gb2)

> summary(mod.fit)

Df Sum Sq Mean Sq F value Pr(>F)

Type 3 1791 596.9 6.3621 0.000853 \*\*\*

factor(Person) 19 82837 4359.8 46.4696 < 2.2e-16 \*\*\*

Residuals 57 5348 93.8

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Notice the use of factor() in the above code because person was coded numerically.

Added after video recording: If Type was coded numerically as well (e.g., 1, 2, 3, 4 for the 4 ball types), we would need to use factor(Type) in the formula argument.

Below is the hypothesis test using α = 0.05:

1. Ho: μTitleistB = μMaxfli = μTitleistP = μTopFlite

Ha: At least two golf ball type means are unequal

1. Test statistic: Fobs = 6.36
2. Critical Value:  = F0.05, 3, 57 = 2.77

> qf(p = 0.95, df1 = 3, df2 = 57)

[1] 2.766438

1. Because 6.36 > 2.77, reject Ho.
2. There is at least one pair of golf ball types that have differences in their mean distances.

Notice the p-value is P(X > 6.36) = 0.000853, where X has a F distribution with ν1 = 3 and ν2 = 57.

Consequences:

Some golf ball types are better (in terms of average distance) than others. A golfer may prefer one type to another. Of course, there are other factors a golfer may consider, such as: price, ball trajectory, etc…, in deciding what type of ball to purchase.

If a golfer wants to choose a ball solely on distance, which golf ball type should the golfer choose? We will answer this question shortly using multiple comparisons!

Notes:

* What if the data was analyzed as a CRD instead?

> mod.fit.CRD <- aov(formula = Distance ~ Type, data =

gb2)

> summary(mod.fit.CRD)

Df Sum Sq Mean Sq F value Pr(>F)

Type 3 1791 596.9 0.5144 0.6736

Residuals 76 88184 1160.3

The p-value for the hypothesis test of

Ho: μTitleistB = μMaxfli = μTitleistP = μTopFlite

Ha: At least two golf ball type means are unequal

is 0.6736. Thus, the variability contributed by each person was quite important due to the different answer we obtained from the RCB. Also, notice how these CRD results agree with what we saw in our first box and dot plot. We could go even further to see that

SSECRD = SSERCB + SSBRCB

88184 = 5348 + 82837

where the subscript denotes the particular design the sums of squares came from. Thus, much of the variability in SSECRD is due to the different people participating in the experiment. This helps to justify the need for blocks!

* What if we analyzed the data again as a CRD, but with the block adjusted version of the data?

> mod.fit.adjust<-aov(formula = Distance.adj ~ Type,

data = gb3)

> summary(mod.fit.adjust)

Df Sum Sq Mean Sq F value Pr(>F)

Type 3 1790.7 596.90 8.4828 6.247e-05 \*\*\*

Residuals 76 5347.8 70.37

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Notice the SSE, SST, and MST values are the same as they were before for the RCB! The MSE value is not quite correct because there are too many degrees of freedom for it.

ANOVA model

By using ANOVA methods with blocks, we are actually estimating a statistical linear model of the form:

Yij = μ + αi + βj + εij

for i = 1, …, t and j = 1, …, ni where

* μ is an overall mean
* αi = treatment effect
* βj = block effect
* εij = random error term that has a normal distribution with mean 0 and variance 

This model shows how we believe that the response Yij comes about. There is an overall mean for the treatments where each treatment and block is a little different, so we add effects corresponding to them. Because every response for a treatment and block combination will not be the same, we add an error term that accounts for perhaps other items that we are not taking into account.

Multiple comparisons

If the hypothesis test for equality of treatment means results in a “reject Ho”, there is at least one pair of means that are different. We are interested then in determining which means are different. The same types of multiple comparison procedures can be used here for the RCB, but now we need to account for the different design structure.

With respect to the CIs and hypothesis tests, the degrees of freedom used are (t – 1)(b – 1) for the corresponding distributional quantile.

Example: Golf balls (golfballs.R, golfballs.csv)

Added after video recording: If Type was coded numerically (e.g., 1, 2, 3, 4 for the 4 ball types), we would need to use "factor(Type)" in the which argument.

HSD using αE = 0.05:

> TukeyHSD(x = mod.fit, conf.level = 0.95, which = "Type")

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = Distance ~ Type + factor(Person), data = gb2)



$Type



diff lwr upr p adj

Maxfli-TitleistB 13.35 5.243792 21.4562076 0.0003159



TitleistP-TitleistB 6.00 -2.106208 14.1062076 0.2157180



TopFlite-TitleistB 6.85 -1.256208 14.9562076 0.1258010



TitleistP-Maxfli -7.35 -15.456208 0.7562076 0.0886358

TopFlite-Maxfli -6.50 -14.606208 1.6062076 0.1584679



TopFlite-TitleistP 0.85 -7.256208 8.9562076 0.9924559

> aggregate(formula = Distance.adj ~ Type, data = gb3, FUN

= mean)

Type Distance.adj

1 TitleistB -6.55

2 Maxfli 6.80

3 TitleistP -0.55

4 TopFlite 0.30

Notes:

* The which argument denotes the set of multiple comparisons to perform. Again, this needs to correspond to a factor specified in the aov() function.
* The aggregate() function is used here to help demonstrate what TukeyHSD() gives in the “diff” column of the output. Note that Maxfli – TitleistB = 6.80 – (-6.55) = 13.35.
* Plot summarizing the results:



Notice the labels are ordered by their adjusted mean values.

* Conclusions:
  + MaxFli and TitleistB are different with respect to their means. Because the confidence interval is positive for μMaxfli – μTitleistB, MaxFli’s mean distance is longer than TitleistB’s mean distance.
  + The low p-values for TopFlite vs. TilteistB and TitleistP vs. Maxfli suggest there exists marginal evidence for a difference between means. Some people may prefer to not mention this part.
* Compare these conclusions to what we saw in the box and dot plots for the block adjusted data.

Unfortunately, pairwise.t.test() can not properly account for a block in a RCB, so this function can not be used for LSD or Bonferroni multiple comparisons. Instead, we can use the agricolae package:

Added after video recording: If Type was coded numerically (e.g., 1, 2, 3, 4 for the 4 ball types), we would need to use "factor(Type)" in the trt argument.

> library(package = agricolae)

> # LSD

> save.LSD <- LSD.test(y = mod.fit, trt = "Type", alpha =

0.05, group = FALSE, p.adj = "none", main = "Golf ball

types")

> save.LSD

$statistics

MSerror Df Mean CV t.value LSD

93.82105 57 206.4 4.692891 2.002465 6.133596

$parameters

test p.ajusted name.t ntr alpha

Fisher-LSD none Type 4 0.05

$means

Distance std r LCL UCL Min Max

Maxfli 213.20 32.06342 20 208.8629 217.5371 163 274

TitleistB 199.85 37.80598 20 195.5129 204.1871 153 283

TitleistP 205.85 32.33749 20 201.5129 210.1871 166 293

TopFlite 206.70 33.73753 20 202.3629 211.0371 156 275

Q25 Q50 Q75

Maxfli 192.00 211.5 221.50

TitleistB 167.00 190.0 214.50

TitleistP 181.50 202.5 217.50

TopFlite 181.25 200.5 215.25

$comparison

difference pvalue signif. LCL

Maxfli - TitleistB 13.35 0.0001 \*\*\* 7.2164038

Maxfli - TitleistP 7.35 0.0197 \* 1.2164038

Maxfli - TopFlite 6.50 0.0382 \* 0.3664038

TitleistB - TitleistP -6.00 0.0550 . -12.1335962

TitleistB - TopFlite -6.85 0.0293 \* -12.9835962

TitleistP - TopFlite -0.85 0.7824 -6.9835962

UCL

Maxfli - TitleistB 19.4835962

Maxfli - TitleistP 13.4835962

Maxfli - TopFlite 12.6335962

TitleistB - TitleistP 0.1335962

TitleistB - TopFlite -0.7164038

TitleistP - TopFlite 5.2835962

$groups

NULL

attr(,"class")

[1] "group"

The LSD multiple comparison procedure leads to declaring the mean for Maxfli to be different from each of the other golf ball types using αI = 0.05. Also, TopFlite and TitleistB would be declared to have different means. The plot summarizing these results is:



> # Bonferroni

> save.Bon <- LSD.test(y = mod.fit, trt = "Type", alpha =

0.05, group = FALSE, p.adj = "bonferroni", main = "Golf

ball types")

Added after video recording: If Type was coded numerically (e.g., 1, 2, 3, 4 for the 4 ball types), we would need to use "factor(Type)" in the trt argument.

> save.Bon

$statistics

MSerror Df Mean CV t.value MSD

93.82105 57 206.4 4.692891 2.733463 8.372657

$parameters

test p.ajusted name.t ntr alpha

Fisher-LSD bonferroni Type 4 0.05

$means

Distance std r LCL UCL Min Max

Maxfli 213.20 32.06342 20 208.8629 217.5371 163 274

TitleistB 199.85 37.80598 20 195.5129 204.1871 153 283

TitleistP 205.85 32.33749 20 201.5129 210.1871 166 293

TopFlite 206.70 33.73753 20 202.3629 211.0371 156 275

Q25 Q50 Q75

Maxfli 192.00 211.5 221.50

TitleistB 167.00 190.0 214.50

TitleistP 181.50 202.5 217.50

TopFlite 181.25 200.5 215.25

$comparison

difference pvalue signif. LCL

Maxfli - TitleistB 13.35 0.0003 \*\*\* 4.977343

Maxfli - TitleistP 7.35 0.1182 -1.022657

Maxfli - TopFlite 6.50 0.2291 -1.872657

TitleistB - TitleistP -6.00 0.3302 -14.372657

TitleistB - TopFlite -6.85 0.1756 -15.222657

TitleistP - TopFlite -0.85 1.0000 -9.222657

UCL

Maxfli - TitleistB 21.722657

Maxfli - TitleistP 15.722657

Maxfli - TopFlite 14.872657

TitleistB - TitleistP 2.372657

TitleistB - TopFlite 1.522657

TitleistP - TopFlite 7.522657

$groups

NULL

attr(,"class")

[1] "group"

The Bonferroni multiple comparison procedure leads to declaring the same conclusions as HSD here using αE = 0.05.

The multcomp package also allows for multiple comparisons for an RCB. Please see the program for code showing how to use the package.

There are many, many other types of experimental designs.

* In the wheat varieties example, we may want to include additional factors like type of fertilizer or whether or not irrigation was used. This becomes a multi-factorial design.
* Latin square designs allow for the incorporation of two blocking mechanisms rather than just the one discussed in these notes.
* There can be more than one size of an experimental unit. A common type of design in this situation is called a “split-plot design”. For example, there may be physical constraints with respect to how different factors are applied to experimental units leading to different sized experimental units for the factors.
* More than one measurement could occur on each experimental unit. This leads to what is known as subsampling.

A course on experimental designs will discuss these examples and MORE!!!