

Randomization Tests: Two-Factorial & Blocking Designs

In this section we will consider randomization tests for two-factor completely randomized designs (CRDs) and randomized complete block designs (RCBDs). Recall that in doing these tests we need to consider how the experimental treatments were initially randomized to the experimental units (i.e. follow the design) and what is the hypothesis we wish to test. For an introduction on the subject, besides Peter Hoff's notes, see *Statistical Inference* by P. Garthwaite, I. Jolliffe, and B. Jones. For a more detailed reference see *Randomization Tests* by Eugene S. Edgington.

- Two-Factor CRD:

Consider the following data from Montgomery — “An experiment is conducted to study the influence of operating temperature and three types of face-plate glass in the light output of an oscilloscope tube.” The experimental units were randomly assigned to *Glass Type* and *Temperature*, so we have a two-factor CRD design.

Table 1: Two-Factor Layout

Glass Type	Temperature		
	100	125	150
1	580, 568, 570	1090, 1087, 1085	1392, 1380, 1386
2	550, 530, 579	1070, 1035, 1000	1328, 1312, 1299
3	546, 575, 599	1045, 1053, 1066	867, 904, 899

```
> rm(list = ls())
>
> y <- c(580, 1090, 1392,
+       568, 1087, 1380,
+       570, 1085, 1386,
+       550, 1070, 1328,
+       530, 1035, 1312,
+       579, 1000, 1299,
+       546, 1045, 867,
+       575, 1053, 904,
+       599, 1066, 889)
>
> glass <- factor(rep( c("GT1", "GT2", "GT3"), each=9))
> temp <- factor(rep(c("100", "125", "150"), 9))
```

As was mentioned in the beginning of Chapter 4, we can treat this as **one-factor** CRD with $3 \times 3 = 9$ levels. In doing so, we can test for difference between any of the 9 levels. So we might have the following hypotheses:

H_0 : There is **no difference** between the different levels of the treatment.

H_1 : There is **a difference** between the levels treatment.

Here we could use an F-statistic as our test statistic and since the design was a CRD we can randomly permute the data **with no restriction** through out the two-factor layout in Table 1. We can randomly sample a permutation, calculate our test statistic, and compare to the observed test statistic.

```
> ##### One-factor randomization test with 9 levels
> glass.temp <- paste(glass, temp, sep=".")
> glass.temp[1:6]
[1] "GT1.100" "GT1.125" "GT1.150" "GT1.100" "GT1.125" "GT1.150"
>
> g.obs <- anova(lm(y ~ as.factor(glass.temp)))[1,4]
> g.obs
[1] 824.7704
> ## test
> g.sim <- real()
> for(nsim in 1:5000){
+ x.sim <- sample(glass.temp)
+ g.sim[nsim] <- anova(lm(y ~ as.factor(x.sim)))[1,4]
+ }
>
> p.value <- mean(g.sim>=g.obs)
> p.value
[1] 0
> max(g.sim)
[1] 8.069625
```

Let's examine the ANOVA table and compare the result with the p-value from the population test:

```
> anova(lm(y ~ as.factor(glass.temp)))
Analysis of Variance Table

Response: y
              Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(glass.temp)  8 2411751  301469  824.77 < 2.2e-16
```

– **Now** let's consider a different hypothesis test:

H_0 : There is **no difference** between the different levels of the **glass type**.

H_1 : There is **a difference** between the levels **glass type**.

So we wish to examine the main effects due to *glass type*. Since we are only interested in this effect, we should **only permute within each column** of Table 1. This holds fixed the effects due to *temperature*. Note that even though the design is CRD, the hypothesis we wish to test restricts the permutation. Let's conduct the test:

```
> ## Two-factor test for glass type.
> D <- data.frame(y, glass, temp)
> g.obs <- anova(lm(y ~ as.factor(glass) + as.factor(temp), data=D))[1,4]
> g.obs
[1] 5.585111
>
```

```

> # selct x
> x <- D[,2:3]
> x.sim <- x
>
> ## test
> g.sim <- real()
> for(nsim in 1:5000){
+
+ x.sim[x.sim$temp==100,]$glass <- sample(x[x$temp==100,]$glass)
+ x.sim[x.sim$temp==125,]$glass <- sample(x[x$temp==125,]$glass)
+ x.sim[x.sim$temp==150,]$glass <- sample(x[x$temp==150,]$glass)
+
+
+ g.sim[nsim] <- anova(lm(y ~ as.factor(x.sim$glass) + as.factor(x.sim$temp)))[1,4]
+ }
>
> p.value <- mean(g.sim>g.obs)
> p.value
[1] 0.0112
> max(g.sim)
[1] 7.863377

```

– Now let's consider the test for *temperature*:

H_0 : There is **no difference** between the different levels of **temperature**.

H_1 : There is **a difference** between the levels of **temperature**.

Since we are only interested in the effect due to *temperature*, we should **only permute within each row** of Table 1. This holds fixed the effects due to *glass type*. Again, even though the design is CRD, the hypothesis we wish to test restricts the permutation. Let's conduct the test:

```

> ## Two-factor test for temperature.
> g.obs <- anova(lm(y ~ as.factor(glass) + as.factor(temp), data=D))[2,4]
> g.obs
[1] 72.94317
>
> # selct x
> x <- D[,2:3]
> x.sim <- x
>
> ## test
> g.sim <- real()
> for(nsim in 1:5000){
+
+ x.sim[x.sim$glass=="GT1",]$temp <- sample(x[x$glass=="GT1",]$temp)
+ x.sim[x.sim$glass=="GT2",]$temp <- sample(x[x$glass=="GT2",]$temp)
+ x.sim[x.sim$glass=="GT3",]$temp <- sample(x[x$glass=="GT3",]$temp)
+
+
+ g.sim[nsim] <- anova(lm(y ~ as.factor(x.sim$glass) + as.factor(x.sim$temp)))[2,4]
+ }

```

```

>
>
> p.value <- mean(g.sim>=g.obs)
> p.value
[1] 0
> max(g.sim)
[1] 18.66026

```

Now let's compare these p-values to those obtained from the population test:

```

> anova(lm(y ~ as.factor(glass) + as.factor(temp), data=D))
Analysis of Variance Table

```

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(glass)	2	150865	75432	5.5851	0.01092
as.factor(temp)	2	1970335	985167	72.9432	1.957e-10
Residuals	22	297131	13506		

- Note that some authors have suggested using unrestricted permutation for testing main effects, but “[t]his has disadvantages for testing whether one factor affected the responses, since the influence of other factors may bias results.” (*Statistical Inference* by P. Garthwaite, I. Jolliffe, and B. Jones).
- So far we have considered randomization tests on the main effects using the F-statistic as our randomization test statistic. From the model:

$$y_{i,j,k} = \mu + glass_i + temp_j + (glass : temp)_{i,j} + \epsilon_{i,j,k},$$

We have:

$$\begin{aligned}
 * F_{glass} &= \frac{\sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 (\bar{y}_{i..} - \bar{y}_{...})^2 / (3-1)}{MSE}, \\
 * F_{temp} &= \frac{\sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 (\bar{y}_{.j.} - \bar{y}_{...})^2 / (3-1)}{MSE}.
 \end{aligned}$$

But can we use a randomization test to test the interaction term $(glass : temp)_{i,j}$? The answer is not directly, if we want to hold constant the two main effects. Approaches have been constructed based on fitting a model and permuting the residuals, but we will not discuss that here (P. Garthwaite, I. Jolliffe, and B. Jones).

- * But notice something, when we consider the design as a one-factor with 9 levels, we can do randomization tests of interest, using unrestricted randomization, based on those 9 cells! For example, we can look for difference among the means of levels for one factor (F_1) across the levels of another factor (F_2) — i.e. interactions! But, we are not separating out the additive main effects from the non-additive part. So some test like:

$$g(\mathbf{Y}) = (\bar{y}_{11.} - \bar{y}_{21.}) - (\bar{y}_{12.} - \bar{y}_{22.}).$$

- **Now let's consider a CRBD.** Recall what the point of blocking is:

- There is factor F_2 which has a large source of variation when considering the relationship between an outcome (y) and a factor F_1 , but we are not really interested in F_2 . It is a **nuisance factor**.
- We can't control F_2 but we can break it up into roughly homogenous blocks and account for it in our analysis.

- Within each block, if we randomly assign **all** the levels of F_1 , then the blocking is said to be **complete**.
- Can we have replication with a complete randomized block design? Of course we can!

Let's consider the **nitrogen timing experiment** outlined on page 127 in Chapter 4:

- Within each row (block) all 6 timings were randomized — thus the design is complete.
- To test for an effect of the timing via a randomization test, we can only permute the timings within each since we need to permute based on the design!

Let's conduct the test:

```
> #####
> ## nitrogen permutation test
> D <- read.table("http://faculty.unlv.edu/westveld/Teaching/Sta713/Data/nitrogen.txt")
>
> ## randomization test for timing treatment while controlling for the rows (i.e. the blocks)
> g.obs <- anova(lm(D$y ~ as.factor(D$trt) + as.factor(D$rw)))[1,4]
> g.obs
[1] 5.591686
>
> # select x
> x <- D[,2:3]
> x.sim <- x
>
> ## test
> g.sim <- real()
> for(nsim in 1:5000){
+
+ x.sim[x.sim$rw==1,]$trt <- sample(x[x$rw==1,]$trt)
+ x.sim[x.sim$rw==2,]$trt <- sample(x[x$rw==2,]$trt)
+ x.sim[x.sim$rw==3,]$trt <- sample(x[x$rw==3,]$trt)
+ x.sim[x.sim$rw==4,]$trt <- sample(x[x$rw==4,]$trt)
+
+ g.sim[nsim] <- anova(lm(D$y ~ as.factor(x.sim$trt) + as.factor(x.sim$rw)))[1,4]
+ }
>
> p.value <- mean(g.sim>=g.obs)
> p.value
[1] 0.0018
> max(g.sim)
[1] 7.322397
```

And the population test comparison:

```
> anova(lm(D$y ~ as.factor(D$trt) + as.factor(D$rw)))
Analysis of Variance Table

Response: D$y
          Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(D$trt)  5 201.316   40.263   5.5917 0.004191
as.factor(D$rw)   3  197.004   65.668   9.1198 0.001116
Residuals       15  108.008    7.201
```