Randomized Block Design

Dr Austin R Brown

Kennesaw State University

Introduction

▶ In the last section, we were introduced to the completely randomized design.

▶ Recall, one of the key characteristics of a CRD is that the treatments are assigned to the experimental units at random.

▶ Additionally, in CRD, we are assuming that we don't have any extraneous sources of variation that could affect the response variable.

  ▶ In other words, we are assuming that the experimental units are homogeneous and that there is no other source of variation (read, independent/explantory/lurking variable) that could affect the response variable.

▶ In practice, this is rarely the case.

Example: Comparing Exercise Interventions in Adults with Type 2 Diabetes

▶ Let's consider an example. Suppose researchers at KSU are interested in understanding how different types of exercise impact blood sugar control (as measured by HbA1c percentage) in adults with Type 2 diabetes.

▶ The three different exercise interventions the researchers would like to compare are:
1. Aerobic Exercise (e.g., treadmill walking/jogging)
2. Resistance Training (e.g., weight lifting)
3. Combined Aerobic and Resistance Training

▶ Let's go through the process of designing an experiment to compare these three exercise interventions.

Designing the Experiment

▶ The first step is to **Define the Objective**. I like to more generally think of this as specifying the research question.
  ▶ This sets up the roadmap for how the experiment will be conducted and ultimately what information we are trying to glean from the experiment's results.

▶ Here, we want to know if there is a difference in HbA1c percentage among the three exercise interventions in the population of adults with Type 2 diabetes.

Designing the Experiment

▶ Next, we want to identify the outcome. In this case, the outcome is the HbA1c percentage and is measured as a continuous variable.

▶ For context, HbA1c percentage is a measure of the average blood sugar levels over the past 2-3 months. It is a key measure in the management of diabetes.

▶ Normal values for HbA1c are typically less than 5.7%. Values between 5.7% and 6.4% are considered prediabetic, and values greater than 6.5% are considered diabetic.

Designing the Experiment

▶ Next, we want to identify the independent variables and possible lurking variables.

▶ Our primary independent variable is the type of exercise intervention. This is a categorical variable with three levels: Aerobic Exercise, Resistance Training, and Combined Aerobic and Resistance Training.

▶ When we think about lurking variables, there are a few that potentially come to mind. One of which is baseline body composition, as typically measured by BMI category.

   ▶ E.g., Underweight, Normal Weight, Overweight, Obese

Designing the Experiment

▶ A person's body composition can impact how they respond to exercise interventions. For example, individuals with higher body fat percentages may respond differently to exercise interventions than individuals with lower body fat percentages.

▶ Since our population of interest is adults with Type 2 diabetes in general, we know that there is likely to be a wide range of body compositions in our sample.

   ▶ In other words, our experimental units are not homogeneous with respect to body composition.

▶ This is where the randomized block design comes into play.

Randomized Block Design

▶ The randomized block design is a type of experimental design that is used when the experimental units are not homogeneous with respect to a particular variable.

   ▶ BMI category in this case

▶ So what we do in this case is partition our sample into homogeneous subgroups called **blocks**. A blocking variable isn't necessarily of interest to us in our experiment, but it is a known, measurable, and likely source of variation that would be beneficial for us to control.

▶ Then, we randomly assign levels of the treatment variable, exercise intervention in this case, to experimental units within each block.

   ▶ This is a similar process to CRD randomization, but undertaken on each block independently.

Randomized Block Design

▶ We also need to be clear on the **Data Collection Process and Procedures**.

▶ In this case, we would want to know the participants' BMI category at the onset of the study and for them to be part of the exercise intervention long enough to see measurable results, if appropriate.

▶ For us, we will have them participate for 12 weeks. We will assess their BMI category at the onset of the study and then at the end of the 12 weeks, we will measure each participant's HbA1c percentage.

▶ We will also recruit 30 participants from each of the four BMI categories (the blocks). Then within each block, we will randomly assign them to one of the three exercise interventions for a total sample size of 120.

Randomized Block Design: Descriptive Analysis

▶ How will using RBD change the descriptive and inferential methods we use to answer our research question?
  ▶ Only a little bit! The general processes are the same!

▶ Let's start with our descriptive analyses. As before, we will perform some quantitative summaries (e.g., calculate sample means and sample standard deviations) and also generate some descriptive visualizations.

▶ Our data is contained in the `Exercise and Diabetes.xlsx` file in D2L.

Randomized Block Design: Descriptive Analysis

```r
library(tidyverse)
library(readxl)
library(rstatix)
## Read in File ##
exercise <- read_excel("Exercise and Diabetes.xlsx")
## Descriptive Analysis - Means and Standard Deviations ##
exercise |>
  group_by(`BMI Category`,`Exercise Treatment`) |>
  get_summary_stats(HbA1c, type = "mean_sd") |>
  arrange(`Exercise Treatment`)
```

Randomized Block Design: Descriptive Analysis

```
# A tibble: 12 x 6
   `BMI Category` `Exercise Treatment` variable     n  mean    sd
   <chr>          <chr>                <fct>    <dbl> <dbl> <dbl>
 1 Normal Weight  Aerobic              HbA1c       10  7.28 0.214
 2 Obese          Aerobic              HbA1c       10  7.80 0.207
 3 Overweight     Aerobic              HbA1c       10  7.54 0.228
 4 Underweight    Aerobic              HbA1c       10  7.30 0.305
 5 Normal Weight  Combined             HbA1c       10  7.11 0.219
 6 Obese          Combined             HbA1c       10  7.43 0.4
 7 Overweight     Combined             HbA1c       10  7.20 0.168
 8 Underweight    Combined             HbA1c       10  7.15 0.18
 9 Normal Weight  Resistance           HbA1c       10  6.94 0.174
10 Obese          Resistance           HbA1c       10  7.39 0.231
11 Overweight     Resistance           HbA1c       10  7.02 0.261
12 Underweight    Resistance           HbA1c       10  6.68 0.197
```
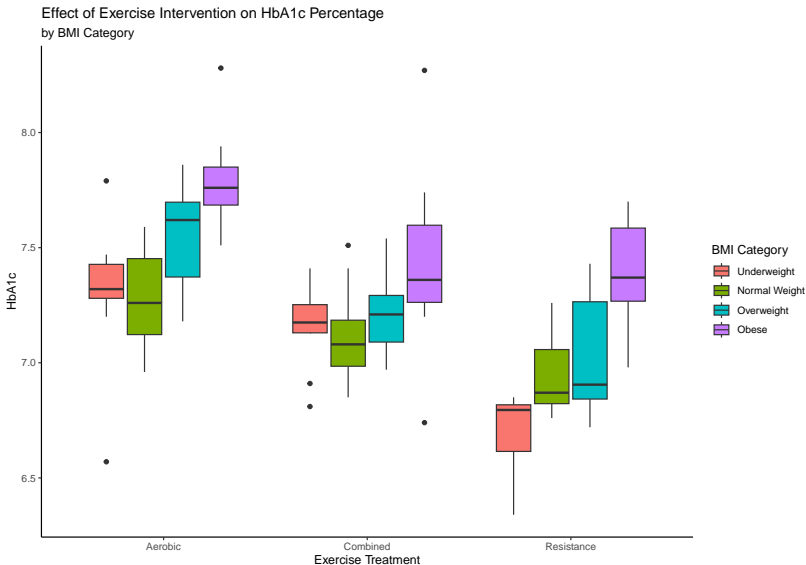
Randomized Block Design: Descriptive Analysis

▶ As we can see here, the mean HbA1c tends to increase within each BMI category.

▶ Additionally, within the exercise treatments, it looks like HbA1c tends to decrease as we move from aerobic to resistance to combined. This may potentially indicate that the combined treatment is the most effective at reducing HbA1c levels, controlling for BMI category.

▶ Let's visually assess potential differences in HbA1c using a boxplot with `ggplot2`!

Randomized Block Design: Descriptive Analysis

```r
exercise |>
  mutate(`BMI Category` = factor(`BMI Category`,levels=c("Underweight",
                                                          "Normal Weight",
                                                          "Overweight",
                                                          "Obese"))) |>
  ggplot(aes(x=`Exercise Treatment`,y=HbA1c,fill=`BMI Category`)) +
  geom_boxplot() +
  labs(title = "Effect of Exercise Intervention on HbA1c Percentage",
       subtitle = "by BMI Category") +
  theme_classic()
```

Randomized Block Design: Descriptive Analysis

Effect of Exercise Intervention on HbA1c Percentage
by BMI Category

Randomized Block Design: Descriptive Analysis

▶ As we can see, the boxplot shows that the combined and resistance treatments tend to produce lower HbA1c levels than the aerobic treatment.

▶ Additionally, it appears that the underweight and normal weight categories had lower HBA1c levels comparing the combined and resistance groups.

▶ So in general, it feels like the answer to our question is that the combined and resistance exercise interventions may be effective compared to the aerobic group.
  ▶ Now let's see if we have statistical evidence to further confirm our suspicion.

Randomized Block Design: Inferential Analysis

▶ The model for the analysis of an RCB design is:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

▶ where β_j represents the block effects and τ_i represents the treatment effects. In general, we have $i = 1, 2, \dots t$ levels of the treatment variable and $j = 1, 2, \dots b$ levels of the blocking factor. This implies that we have $t \times b$ combinations of treatment and blocking effects.

▶ Supposing we have an equal number of r replicates per each combination of treatment and blocking effect, we have a total of $t \times b \times r$ observations in our sample.

▶ As before, we assume $\varepsilon_{ij} \sim N(0, \sigma^2)$ which we can assess in a similar manner as CRD.

Randomized Block Design: Inferential Analysis

▶ For our omnibus test of this model, our main test of concern is the test for the treatments effect. The null and alternative hypothesis for this effect is essentially the same as the CRD model:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$$

H_1 : At least one pair of group means are not equal

▶ The question is though: how is the test constructed? Here's where there are some noteworthy differences.

Randomized Block Design: Inferential Analysis

▶ The way the total amount of variance is estimated in a sample is through a measure called **Sums of Squares Total (SST)**.

▶ As before, SST is just the sum of the squared differences between each observation and the grand mean. In the case of RBD:

$$SST = \sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{..})^2$$

Randomized Block Design: Inferential Analysis

▶ The logic behind all supervised learning methods (i.e., those in which we have an outcome and predictors/independent variables/explanatory variables) is that the total amount of variation in the outcome can be partitioned into two categories:

  ▶ Explained variation (that is, variation in the response accounted for by the predictors)
  ▶ Unexplained variation (that is, variation in the response not accounted for by the predictors)

▶ With regards to SST, this means that we can break its value down into those two sources:

$$SST = \text{Explained Variation} + \text{Unexplained Variation}$$

Randomized Block Design: Inferential Analysis

▶ This concept and equation are referred to as the **_Fundamental ANOVA Identity_**.

▶ In CRD, Explained Variation was SSTreat and Unexplained Variation was SSE.

▶ For us here in RBD, we need to also account for the blocking effect in the computation of Explained Variation.

Randomized Block Design: Inferential Analysis

▶ So:

$$\text{Explained Variation} = SSTreat + SSBlocks$$

$$\text{Explained Variation} = b\sum_{i=1}^{t}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + t\sum_{j=1}^{b}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$$

▶ Here $\bar{y}_{i\cdot}$ represents the sample mean of the ith treatment, $\bar{y}_{\cdot j}$ represents the sample mean of the jth block, and $\bar{y}_{\cdot\cdot}$ represents the overall sample mean.

Randomized Block Design: Inferential Analysis

▶ Complibling all of this into a standard ANOVA table:

Table 1: ANOVA Table for Randomized Block Design with One Blocking Factor

Source	SS	df	Mean Square	F-Ratio
Treatments	SSTreat	$t-1$	$\frac{SSTreat}{t-1}$	$\frac{MSTreat}{MSE}$
Blocks	SSBlocks	$b-1$	$\frac{SSBlock}{b-1}$	
Error	SSE	$(t-1)(b-1)$	$\frac{SSE}{(t-1)(b-1)}$	
Total	SST	$N-1$		

Randomized Block Design: Inferential Analysis

▶ Therefore:

$$F_{Stat} = \frac{MSTreat}{MSE} \sim F(t-1, (t-1)(b-1))$$

▶ So as before, if our $F_{Stat} > F_{cv}$ for some given α level, we would say our data more strongly support the alternative hypothesis.

▶ Let's see how we obtain this information using R:

Randomized Block Design: Inferential Analysis

```r
## Fit Blocked ANOVA Model ##
block_mod <- aov(HbA1c ~ `Exercise Treatment` + `BMI Category`,
                 data=exercise)
## Extract ANOVA Table ##
library(broom)
block_mod |>
  tidy()
```

```
# A tibble: 3 x 6
  term                  df sumsq meansq statistic  p.value
  <chr>              <dbl> <dbl>  <dbl>     <dbl>    <dbl>
1 `Exercise Treatment`   2  4.40   2.20      36.5 5.56e-13
2 `BMI Category`         3  4.34   1.45      24.0 4.02e-12
3 Residuals            114  6.87 0.0603        NA       NA
```

Randomized Block Design: Inferential Analysis

▶ Note, remember that before we can feel confident in interpreting these results, we must fist perform all relevant checks of the assumptions.

  ▶ I do this in the R script, not here for the sake of breveity. Please run the code in the associated R script.

▶ You can see all of the techniques we use are exactly the same except one. Instead of using Levene's Test, I'm using the Breush-Pagan Test to assess the constant variance assumption.

▶ The null and alternative hypothesis for both tests are the same (e.g., H_0: Variance is constant, H_1: Variance is not constant). They just work slightly differently with the B-P test being more broadly applicable to all types of linear models, not just ANOVA methods.

Randomized Block Design: Inferential Analysis

▶ Four notes: First, in this output, the Residuals row is the same as the Error row in table 1.

▶ Second, the Totals row is omitted.

▶ Third, we can obviously perform a test of significance for our blocking effect as well. We don't typically do this in RBD because aren't necessarily concerned about its significance; we merely wanted to control for its effect.

▶ Fourth and finally, we assume that the block and the treatment are *independent* of each other. This may or may not be a reasonable assumption.

Randomized Block Design: Inferential Analysis

▶ Contextually we can see that since the p-value for the Exercise Treatment is less than our typically used threshold of 0.05, this would indicate that the data more strongly support H_1, which is in alignment with what we saw in the descriptive analysis.

▶ This would indicate that at least two of the exercise treatment groups differ in terms of mean HbA1c. But which pair(s)?

▶ To answer this question, we can use Tukey's HSD just like before:

Randomized Block Design: Inferential Analysis

```
## Tukey HSD using tukey_hsd function from rstatix ##
exercise |>
  rename(Exercise_Treatment = `Exercise Treatment`,
         BMI_Category = `BMI Category`) |>
  tukey_hsd(HbA1c ~ Exercise_Treatment+BMI_Category) |>
  filter(term == "Exercise_Treatment") |>
  select(group1,group2,estimate,p.adj)
```

Randomized Block Design: Inferential Analysis

```
# A tibble: 3 x 4
  group1   group2     estimate   p.adj
  <chr>    <chr>         <dbl>    <dbl>
1 Aerobic  Combined     -0.254 2.97e- 5
2 Aerobic  Resistance   -0.468 2.6 e-13
3 Combined Resistance   -0.215 4.55e- 4
```

Randomized Block Design: Inferential Analysis

▶ Note, in the prior code, I had to use the rename function since the column names contained spaces (the tukey_hsd function doesn't work with columns that contain spaces in their names).

▶ Then, I used the tukey_hsd function which is part of the rstatix package to avoid having to recreate the aov object with the switched column names.

   ▶ It does the same thing as TukeyHSD but just has nicer output by default.

▶ **Contextually**, we can see that all of the p-values are less than 0.05, which would indicate that all the HbA1c group means may differ substantially from each other in this population, controlling for BMI category.

   ▶ As we saw, people in the Aerobic treatment group had the greatest group mean, followed by Resistance, and then Combined.

Randomized Block Design: Limitation

▶ In this study, what are some limitations?

▶ For starters, we didn't record the participant's initial HbA1c level, so it isn't clear if the level we observed after 12 weeks was meaningfully different from where they began.

▶ More generally, we also assume in RBD that we don't have an interaction between the blocks and the treatments, which may not be a reasonable assumption.

▶ Finally, if we misspecify a blocking factor which doesn't substantially reduce SSE, then we are inadvertly decreasing statistical power for our treatment test via the reduction of degrees of freedom for SSE.