STAT 3119

Week 10: 10/31/2019 @GWU

Outline

- Two-factor ANCOVA (Ch 22.4)
- ANCOVA for Randomized complete block designs
- Additional Consideration for ANCOVA (Ch 22.5)
- Recitation: discussion on last HW and mid-term (everyone did great!)

Single-factor ANCOVA : one factor (r levels)

The standard **ANCOVA model for a single-factor study** with r fixed levels (22.3) is

$$Y_{ij} = \mu_{.} + \tau_i + \gamma(X_{ij} - \overline{X}_{..}) + \epsilon_{ij} \qquad (22.3)$$

where

- $\mu_{.}$ = an overall mean
- τ_i = the fixed treatment effects subject to the restriction $\sum \tau_i = 0$.
- γ = the regression coefficient for the relation between Y and X.
- Error term ϵ_{ij} = independent $N(0, \sigma^2)$, $i = 1, ..., r; j = 1, ..., n_i$.

We estimate the model parameters based on it equivalent **Regression** model: We can express the ANCOVA model (22.3) as follows

$$Y_{ij} = \mu_{.} + \tau_1 I_{ij1} + ... + \tau_{r-1} I_{ij,r-1} + \gamma x_{ij} + \epsilon_{ij}$$

where

- $x_{ij} = X_{ij} - \overline{X}_{..}$, centered observations
- we use $r-1$ indicator functions to represent the r treatment levels, $I_{ij,m} = 1$ if observation is from level i, -1 if observation is from level r and 0 otherwise, $m = 1, ..., r-1$.

Two-factor ANCOVA (Ch 22.4) : two factors+ a continuous ovariate

From Chapter 19, we study the 2-factor balanced study by the following ANOVA model:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where

- $\mu_{..}$ overall mean,

- α_i and β_j are the main effect of factor A and B; $(\alpha\beta)_{ij}$ interaction effects.
- Zero sum constraints: $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, and

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0\,, \quad i = 1..., a\,; \, j = 1, ..., b$$

- ϵ_{ijk} are iid $N(0, \sigma^2)$
- For a balanced study, $k = 1, ..., n$; For unbalanced design, $k = 1, ..., n_{ij}$ (To be discussed in Chapter 23).

The ANCOVA model for a two-factor study is as follows:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma(X_{ijk} - \overline{X}_{...}) + \epsilon_{ijk}$$

Similarly, we will use regression approach to get the model parameter estimates, which does not require the study to be 'balanced'.

Estimation with regression approach

We need to set up the correct form for the regression model, so the regression coefficients will correspond to the same parameter as in the ANCOVA model.

For simplicity, the textbook only illustrates a balanced two-factor study with one concomitant variable when both factors A and B are at two levels, i.e., when $a = b = 2$. Then we only need one indicator function for each factor. The regression model counterpart to covariance model then is:

$$Y_{ijk} = \mu_{..} + \alpha_1 I_{ijk1} + \beta_1 I_{ijk2} + (\alpha\beta)_{11} I_{ijk1} I_{ijk2} + \gamma x_{ijk} + \varepsilon_{ijk} \qquad (22.27)$$

where:

$$I_1 = \begin{cases} 1 & \text{if case from level 1 for factor } A \\ -1 & \text{if case from level 2 for factor } A \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if case from level 1 for factor } B \\ -1 & \text{if case from level 2 for factor } B \end{cases}$$

$$x_{ijk} = X_{ijk} - \overline{X}_{...}$$

Note that the regression coefficients in (22.27) are the analysis of variance factor effects α_1, β_1, and $(\alpha\beta)_{11}$ and the concomitant variable coefficient γ.

From the regression model, we will get the $\hat{\mu}_{..}$, $\hat{\alpha}_1$, $\hat{\beta}_1$ and $\hat{\alpha}_{11}$. Then we can obtain from the treatment effects from the zero-sum constraints, e.g. $\hat{\alpha}_2 = -\hat{\alpha}_1$; $\hat{\beta}_2 = -\hat{\beta}_1$; then interaction effects $\widehat{\alpha\beta}_{12} = \widehat{\alpha\beta}_{21} = -\widehat{\alpha\beta}_{11}$ and $\widehat{\alpha\beta}_{22} = \widehat{\alpha\beta}_{11}$.

Hypothesis Testing: We compare the above full regression model with a reduced model,

- Testing for factor A effects requires that $\alpha_1 = 0$ in the reduced model.
- Testing for factor B effects requires that $\beta_1 = 0$ in the reduced model.
- When testing for AB interactions, we set $\alpha\beta_{11} = 0$ in the reduced model.

Estimation:

- Estimation of factor A and factor B effects and the adjusted treatment means can also be done from the regression.
- We can use similar multiple comparison procedures as in the one-factor ANCOVA by replacing $n_T = nab$ and $r = ab$.

In general, the regression approach can be applied to factor A with $a(a \geq 2)$ levels and factor B with $b(b \geq 2)$ levels. We just need to use more indicator functions: i.e., $(a-1)$ indicator functions for factor A and $(b-1)$ indicator functions for factor B; then also add their interaction terms in the regression model. And for a given factor, each indicator function will have 3 possible values $(1, -1, \text{or } 0)$, $I_{ij,m} = 1$ if it is from level i, $I_{ij,m} = -1$ if the observation is from the last level and 0 otherwise.

Data Example (page 934)

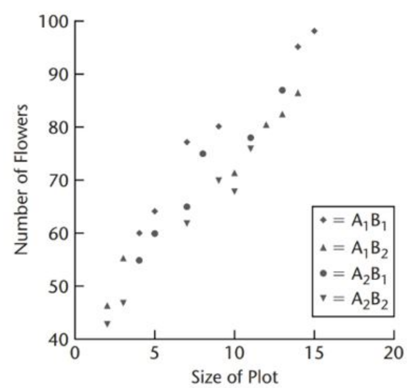
A gardener conducted an experiment to study the effects of flower variety (factor A: varieties LP, WB) and moisture level (factor B: low, high) on yield of salable flowers (Y).

- Because the plots were not of the same size, the gardener wished to use plot size (X) as the covariate.
- Six replications were made for each treatment. A portion of the data are presented in Table 22.6. Figure 22.7 contains a symbolic scatter plot of the data.
- The model assumptions of linear relations between Y and the concomitant variable X, as well as of parallel slopes for the four treatments, appear to be reasonable here.

TABLE 22.6 Data—Salable Flowers Example.

Factor A (flower variety) i	Factor B (moisture level) j			
	B_1 (low)		B_2 (high)	
	Y_{i1k}	X_{i1k}	Y_{i2k}	X_{i2k}
A_1 (variety LP)	98	15	71	10
	60	4	80	12
	...	...	...	...
	64	5	55	3
A_2 (variety WB)	55	4	76	11
	60	5	68	10
	...	...	...	...
	78	11	70	9

FIGURE 22.7 Symbolic Scatter Plot of Salable Flowers and Size of Plot—Salable Flowers Example.



Analysis using R

1. Read the data

```r
# read data from week5 folder online
Ex22B  =read.table(
        url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week-10/CH22TA06.txt"))

#Ex22B  =read.table("CH22TA06.txt") # if data in local computer

head(Ex22B, 7)
```

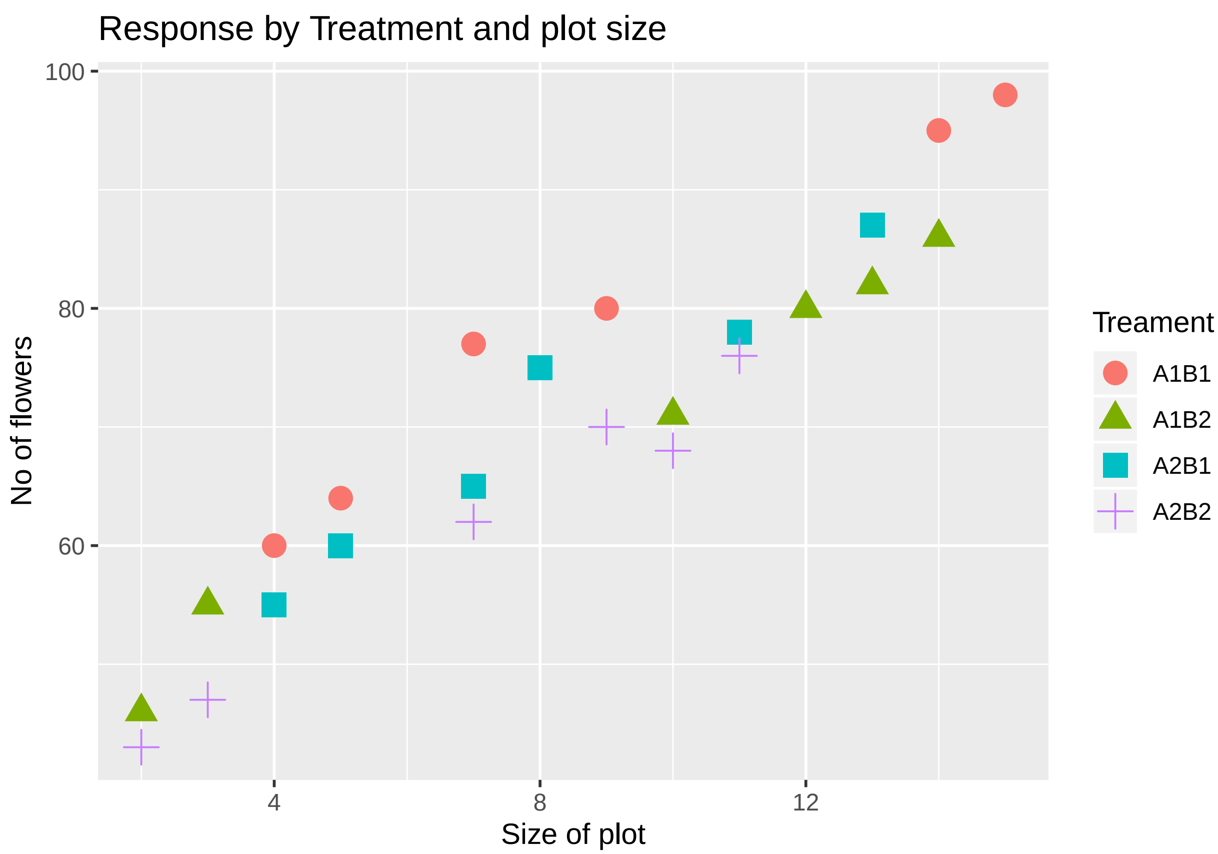
```
##   V1 V2 V3 V4 V5
## 1 98 15  1  1  1
## 2 60  4  1  1  2
## 3 77  7  1  1  3
## 4 80  9  1  1  4
## 5 95 14  1  1  5
## 6 64  5  1  1  6
## 7 55  4  2  1  1
```

```
names(Ex22B) =  c("Num_flowers", "Plot_size", "FactorA","FactorB", "Units")

# Add a treatment variable (1,2)*(2,3), get 2,3,4,6 for 4 combinations
Ex22B$Treament = Ex22B$FactorA*(Ex22B$FactorB+1)

# make the factor as a factor variable
Ex22B$FactorA =  as.factor(Ex22B$FactorA)
Ex22B$FactorB =  as.factor(Ex22B$FactorB)
Ex22B$Treament =  as.factor(Ex22B$Treament)
levels( Ex22B$Treament) <- c("A1B1","A1B2","A2B1", "A2B2")

str(Ex22B )
```

```
## 'data.frame':    24 obs. of  6 variables:
##  $ Num_flowers: int  98 60 77 80 95 64 55 60 75 65 ...
##  $ Plot_size  : int  15 4 7 9 14 5 4 5 8 7 ...
##  $ FactorA    : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 2 2 2 2 ...
##  $ FactorB    : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Units      : int  1 2 3 4 5 6 1 2 3 4 ...
##  $ Treament   : Factor w/ 4 levels "A1B1","A1B2",..: 1 1 1 1 1 1 3 3 3 3 ...
```

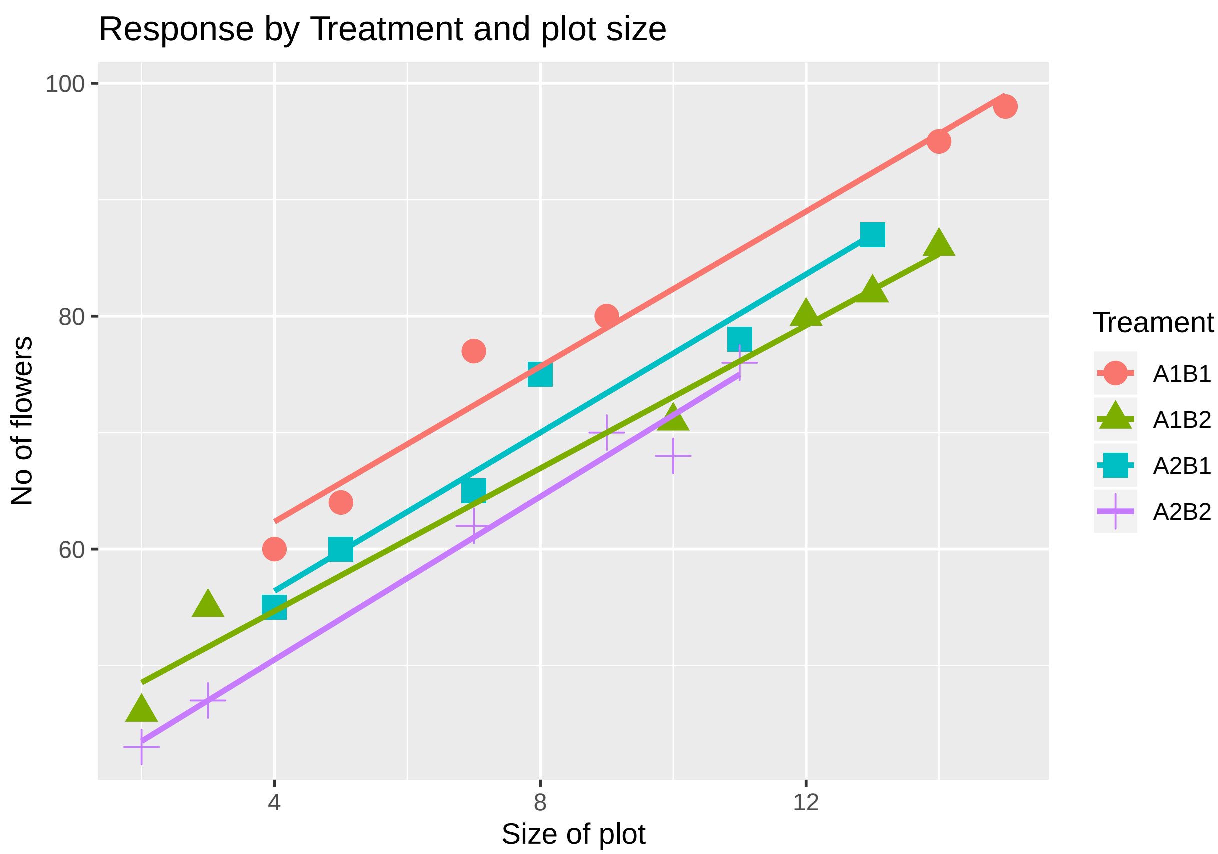
2A. Plot the response by covariate , group by factor A and B

```
library(ggplot2)
P= ggplot(Ex22B, aes(x= Plot_size, y= Num_flowers, color= Treament, shape= Treament ))+
  geom_point(size=4) +
  labs(title="Response by Treatment and plot size",
       x="Size of plot", y="No of flowers")
P
```



Note: We can see the four groups much clearly compared the gray-scale plot. We can also add the within group LS lines to visually check the ANCOVA assumptions of 'linearity' and 'constant slope' by treatment group.

```
# Add the smoother layer: LS fit
P+  geom_smooth(method = lm, se= F)
```



Note: The assumptions seem reasonable so we can proceed with ANCOVA analysis.

Example: ANCOVA by regression

1. Run the correct regression

```
IndicatorA = (Ex22B$FactorA=="1")*1 + (Ex22B$FactorA=="2")*(-1)
IndicatorB = (Ex22B$FactorB=="1")*1 + (Ex22B$FactorB=="2")*(-1)

# center the observation
(meanX= mean( Ex22B$Plot_size))
```

```
## [1] 8.25
```

```
X.centered = Ex22B$Plot_size - meanX

LM.full  =  lm( Num_flowers~ IndicatorA*IndicatorB + X.centered, data=Ex22B )
summary(LM.full)
```

```
##
## Call:
## lm(formula = Num_flowers ~ IndicatorA * IndicatorB + X.centered,
##     data = Ex22B)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8306 -1.7500 -0.5228  1.0000  5.0000
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            70.0000     0.5119 136.751  < 2e-16 ***
## IndicatorA              2.0423     0.5211   3.919 0.000921 ***
## IndicatorB              3.6808     0.5129   7.176 8.09e-07 ***
## X.centered              3.2769     0.1300  25.203 4.59e-16 ***
## IndicatorA:IndicatorB   0.8192     0.5129   1.597 0.126719
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.508 on 19 degrees of freedom
## Multiple R-squared:  0.9765, Adjusted R-squared:  0.9716
## F-statistic: 197.4 on 4 and 19 DF,  p-value: 3.433e-15
```

Note: We can then get the Fitted Regression Function and model parameter estimates (and their SE),

$$\hat{Y} = 70.0 + 2.0423 I_A + 3.6808 I_B + .8192 I_A I_B + 3.2769 x$$

The regression model shows that both factor A and factor B had significant effects. But their interaction was not significant. Also the covariate X is significantly related to the response Y.

2. Testing for each factor (each coefficient = 0)

1. Approach 1: Set a given coefficient zero to get a reduced model, then use anova() to compare two models

```r
# testing interaction, remove interaction term
LM.Nointeaction  =  lm( Num_flowers~ IndicatorA+IndicatorB + X.centered, data=Ex22B )
anova(LM.Nointeaction, LM.full)
```

```
## Analysis of Variance Table
##
## Model 1: Num_flowers ~ IndicatorA + IndicatorB + X.centered
## Model 2: Num_flowers ~ IndicatorA * IndicatorB + X.centered
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     20 135.52
## 2     19 119.48  1    16.042 2.5511 0.1267
```

Results: we can find the F-test statistic= 2.55, following $F(1, 19)$ distribution, with p-value=0.13.

2. Approach 2: We get the "Type III" sum of squares directly: Type III SS are "partial" SS, unlike what we have learned for standard ANOVA.

In ANOVA or general linear models, there are 3 types of SS.

- Type I, also called 'sequential' sum of squares: In essence the factors are tested in the order they are listed in the model, e.g: A+ B+ A:B . For balanced two factor studies, this is what we used in ANOVA table. The effects of A & B are orthogonal and it does not matter which factor is considered first. For unbalanced multi-factor study data, that the order of terms in the model matters. Type I SS for all the effects add up to the total SS.

- **Type III SS** are "partial" SS, every term in the model is tested in light of all other terms in the model. SS gives the sum of squares that would be obtained for each variable if it were entered last into the model, i.e., the "extra" SS in our textbook. Type III SS do not depend on the order of the factors in the model and the SS for all the effects do not add up to the total SS.

- Type II SS are similar to Type III, except that they preserve the principle of marginality. This means that main factors are tested in light of one another, but not in light of the interaction term that involves them.

We can use the **car** package in R and use **Anova()** functions to get the partial SS (type III). Since the factor effects components are not orthogonal here and we don't have the usual ANOVA SS partition.

```
options(contrasts = c("contr.sum", "contr.poly"))
library(car)
```

```
## Loading required package: carData
```

```
Anova(LM.full, type="III")
```

```
## Anova Table (Type III tests)
##
## Response: Num_flowers
##                      Sum Sq Df    F value    Pr(>F)
## (Intercept)          117600  1 18700.8527 < 2.2e-16 ***
## IndicatorA               97  1    15.3617 0.0009211 ***
## IndicatorB              324  1    51.4988 8.093e-07 ***
## X.centered             3995  1   635.2118 4.586e-16 ***
## IndicatorA:IndicatorB    16  1     2.5511 0.1267191
## Residuals               119 19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: Using **Anova()**, we get all the extra SS and their associated F-tests at once. These output are the same as Table 22.7 part (b).

Make inference on the factor effects

Following the example (page 937), since there was no significant interaction, it is valid to estimate the effect of factor A and B separately. In this example, we are interested to estimate the simultanous CI with 95% family confidence coefficients, $L_1 = \alpha_1 - \alpha_2 = \alpha_1 - (-\alpha_1) = 2\alpha_1$, and $L_2 = \beta_1 - \beta_2 = 2\beta$.

- From regression out, we have point estimate and SE, then

$$\hat{L}_1 = 2\hat{\alpha}_1 = 2(2.04234) = 4.08$$

$$\hat{L}_2 = 2\hat{\beta}_1 = 2(3.68078) = 7.36$$

$$s\{\hat{L}_1\} = 2s\{\hat{\alpha}_1\} = 2(.52108) = 1.042$$

$$s\{\hat{L}_2\} = 2s\{\hat{\beta}_1\} = 2(.51291) = 1.026$$

- We utilize the Bonferroni simultaneous estimation procedure for g = 2 comparisons, to get the two Cis:

R:

```
qt( 1-0.05/(2*2), df= 19)
```

```
## [1] 2.43344
```

$$1.5 = 4.08 - 2.433(1.042) \leq \alpha_1 - \alpha_2 \leq 4.08 + 2.433(1.042) = 6.6$$

$$4.9 = 7.36 - 2.433(1.026) \leq \beta_1 - \beta_2 \leq 7.36 + 2.433(1.026) = 9.9$$

Note: If other adjusted means or linear combinatons of model parameter estimates are used, we can always plug in the estimates from regression output, and derive and calculate its variance estimate directly using the variance-covariance matrix from the regression by **vcov(model.fit)**.

ANCOVA analysis for Randomized Complete Block Designs

This is also straightforward extension from the ANOVA model for RCBD design previously discussed in Chapter 21:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \epsilon_{ij}$$

- $\mu_{..}$ is the overall mean
- (similar to α_i as before) ρ_i are the constant for the block effects with $\sum_i \rho_i = 0$.
- (similar to β_j as before) τ_j is the main treatment effect with $\sum_j \tau_j = 0$.
- ϵ_{ij} independent $N(0, \sigma^2)$
- $i = 1, ..., n_b$ (no. of blocks), $j = 1, ..., r$ (treatment levels).

After adding the linear relationship with covariate, we have the ANCOVA model (22.29) for RCBD:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \gamma(X_{ij} - \overline{X}_{..}) + \epsilon_{ij}$$

Both the **estimation** and **additional inference** are based on regression approach (setting $n_b - 1$ indicator functions for block effects and $r - 1$ indicator functions for treatment effects), and we can compare the full and reduced regression models to get the hypothesis testing results , following the example in the textbook on page 938.

Additional considerations (1): ANCOVA vs. Block Design

Sometimes, a choice exists between: (1) a completely randomized design, with covariance analysis used to reduce the experimental errors and (2) a randomized block design, with the blocks formed by means of the covariate.

Generally, Block design is preferred because

1. If the regression between the response variable and the concomitant (blocking) variable is linear, a randomized block design and covariance analysis are about equally efficient. If the regression is not linear, the ANCOVA analysis with linear assumptions may not as effective as a randomized block design.
2. Randomized block designs are essentially free of assumptions about the nature of the relationship between the blocking variable and the response variable. We randomize all the treatment within the blocks to remove the confounding effects, while ANCOVA analysis assumes a definite form of relationship.
3. However, randomized block designs have somewhat fewer degrees of freedom available for experimental error than with ANCOVA analysis for a completely randomized design. RCBD used $(n_b - 1)$ df and ANCOVA use 1 df for linear relationship. This is OK if the sample size of the whole study is large. For small studies, we can't use too many block levels, whick makes the analysis less efficient.

Additional considerations (2): Use of Differences or change from baseline

In many studies, a prestudy observation X and a poststudy observation Y on the same variable are available for each unit. e.g.,

- X may be the score for a subject's attitude toward a company prior to reading its annual report, and Y may be the score after reading the report.
- X may be the lab values before or after a medical treatment

In the one-factor ANCOVA model without centering the covariate:

$$Y_{ij} = \mu_. + \tau_i + \gamma X_{ij} + \epsilon_{ij}$$

If $\gamma = 1$, then we can move X_{ij}, e.g. baseline or pre-treatment level to the left side, then define a new response variable as the difference $d_{ij} = Y_{ij} - X_{ij}$, then we can use a regular ANOVA model on d_ij,

$$Y_{ij} - X_{ij} \sim d_{ij} = \mu_. + \tau_i + \epsilon_{ij}$$

Example:

In the cracker promotion example, Y = current sales, X= previous sales,

```r
# read data from week5 folder online
Ex22  =read.table(
        url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week8/CH22TA01.txt"))

names(Ex22) =  c("New_Sales", "Old_Sales", "Treatment","Units")

# make the treatment as a factor variable
Ex22$Treatment =  as.factor(Ex22$Treatment)
```

```r
# get the residual MSE
Indicator1 = (Ex22$Treatment=="1")*1 + (Ex22$Treatment=="3")*(-1)
Indicator2 = (Ex22$Treatment=="2")*1 + (Ex22$Treatment=="3")*(-1)

# center the observation
X.centered = Ex22$Old_Sales - mean( Ex22$Old_Sales)

# Model 1: ANCOVA on Y
ANCOVA.fit<-lm( New_Sales~ Indicator1 + Indicator2 + X.centered, data=Ex22 )
summary(ANCOVA.fit)
```

```
##
## Call:
## lm(formula = New_Sales ~ Indicator1 + Indicator2 + X.centered,
##     data = Ex22)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4348 -1.2739 -0.3362  1.6710  2.4869
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.8000     0.4835  69.908 6.37e-16 ***
## Indicator1    6.0174     0.7083   8.496 3.67e-06 ***
## Indicator2    0.9420     0.6987   1.348    0.205
## X.centered    0.8986     0.1026   8.759 2.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.873 on 11 degrees of freedom
## Multiple R-squared:  0.9403, Adjusted R-squared:  0.9241
## F-statistic: 57.78 on 3 and 11 DF,  p-value: 5.082e-07
```

```r
# Model 2: ANOVA on difference
Ex22$Diff <- Ex22$New_Sales - Ex22$Old_Sales
ANOVA.Diff.fit<-lm( Diff~ Indicator1 + Indicator2 , data=Ex22 )
summary(ANOVA.Diff.fit)
```

```
##
## Call:
## lm(formula = Diff ~ Indicator1 + Indicator2, data = Ex22)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -2.8   -0.9   -0.6    1.6    3.2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.8000     0.4830  18.218 4.14e-10 ***
## Indicator1    6.2000     0.6831   9.076 1.01e-06 ***
## Indicator2    0.8000     0.6831   1.171    0.264
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.871 on 12 degrees of freedom
## Multiple R-squared:  0.9129, Adjusted R-squared:  0.8984
## F-statistic: 62.91 on 2 and 12 DF,  p-value: 4.356e-07
```

```
# we can also use aov()
summary(aov( Diff~ Treatment, data=Ex22))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment    2  440.4   220.2   62.91 4.36e-07 ***
## Residuals   12   42.0     3.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Then we can see

1) For ANCOVA model, the residual error MSE= $(1.873)^2 = 3.508$. The coefficient $\gamma = 0.899$ is quite close to 1.
2) For ANOVA model on the difference, the residual error MSE= $(1.871)^2 = 3.501$.

Thus, the two procedures are almost the same, in terms of estimation efficiency on the residual error variance.

However,

1) This only applies when the regression coefficient γ is close to 1. Otherwise, the ANCOVA on Y_{ij} is more efficient to estimate the treatment difference, compared to ANOVA on d_{ij}.
2) This only applies when X and Y are the same variable , thus meaningful to take their difference. e.g. it does not apply to the last example, when Y is the numbers of flowers and X is the size of plot.

Additional considerations (3)

- With observational studies, due to confounding, ANCOVA analysis may be used to adjust for confounding effects and reduce bias of the results. However, there may be lots of unknown or unmeasured confounding variables. Therefore, even we adjust for all the known confounders, we still can conclude the 'association' from observational studies.
- To prove the causal effect of the treatment, we have to rely on the randomization to balance out all the known and unknown confounders, i.e. to use comparative experimental studies. Sometimes, to estimate treatment effects more accurately (and adjust for the residual 'unbalance' in the covariates), we can still pre-specify a number of known covariates and use ANCOVA for the final analysis in a randomized clinical trials.

Summary this week

- Reading: Chapter 22 ANCOVA
- Reminder:

1. Project instruction online, start early, due befoe Final Exam (I need to post grade within a few days after final).

2. Home problems: 22.7, 22.8, 22.15, 22.16, 22.19, 22.21 due by 6 pm next Thursday

    - 22.7, 22.8, 22.21 (use data: CH22PR07.txt) : Productivity improvement, previously analyzed in HW for Ch 16.
    - 22.15, 22.16 (use data:CH22PR15.txt) : Cash offers, previously analyzed in HW for Ch 19
    - 22.19 (use data:CH22PR19.txt): Auditor training, previously analyzed in HW for Ch 21
    - In each data, a new continuous covariate has been introduced to use ANCOVA.

3. A short Quiz next Thursday for ANCOVA (11/7)

    - Goal of ANCOVA
    - Single-factor ANCOVA factor effects model
    - Equivalent regression model: how to set the indicator functions in regression
    - Two additional ANCOVA assumptions
    - How to test and estimate model parameters
    - use ANCOVA in the two-factor studies or block design

How to draw a Jack-o- Lantern in R

```r
source('pumpkin.R')
```