STAT 3119

Week 7: 10/10/2019 @GWU

Outline

- `Block design`
- `Model for Randomized Complete Block Design (RCBD)`
- `ANOVA analysis and model diagnostics`
- `Analysis of treatment effects`
- `Quiz#3 (today)`
- `Mid-term exam: Review 10/17 (next Thurday); Test: 10/24 (in 2 weeks after the fall break)`

Motivating Example: Blocking

- Consider the problem of investigating whether or not different brands of tires have different amount of tread loss after 20,000 miles of driving. You have been given the task to consider four brands that are available and to make a decision about which brand might show the least amount of tread wear. (*Foundermental Concepts in DOE. Hicks 1993*)
- Say, instead of simulating in the lab, you can try these four brands under actual driving conditions, and the response variable to be measured is the difference in maximum tread thickness on a tire between the time it is mounted on the wheel and after it has completed 20,000 miles on this car.
- The only factor of interest is the tire brands, say A, B, C,D.
    - `Let's design an experiment  to study this problem !`

Motivating Example (2)

- `Design 1 for Tire Brand Test :`
- Since the tires must be tried on cars and there are some measurement errors, more than one tire of each brand must be used. We can take 16 tires , 4 of each brand, and use four car. You might put the four tires of Brand A on Car I, Brand B on car II, …, as follows.

	Car			
	I	II	III	IV
Brand distribution	A	B	C	D
	A	B	C	D
	A	B	C	D
	A	B	C	D

- Problem of this design: the averages for brands are also averages for cars. If the cars travel over different terrains with different drivers, any differences in tire brands are also car differences. So the cars and tires are *completely confounded* in the design and we can't separate their effects in the analysis.

Motivating Example (3)

- `Design 2 for Tire Brand Test :`
- This is a one-factor experiment , we can try a *completely randomized design* (CRD) as we learned in Chapter 16. We can assign the 16 tires to the four cards in a completely random manner, which might give the design as follows. The objective of this CRD is to average out any car differences might affect the results.

	Car			
	I	II	III	IV
Brand distribution	C	A	C	A
	A	A	D	D
	D	B	B	B
	D	C	B	C

- The one-way ANOVA model would be

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

with tire effects $j = 1, 2, 3, 4$. The Car effects are not entered here.

  - `Problem:`
  - If we look carefully in Design 2: Brand A is never used in Car III; B not in Car I; D not in Car II. The variations for each brand might not be due to experimental error only but may also include variation between cars.
  - The objective of DOE is to reduce experimental error, a better design is to remove car variation from the error variation. CRD design averaged out car effects but did not remove the variance among cars.

Motivating Example (4)

- `Design 3 for Tire Brand Test :`

We put an restriction on the randomization, requires that each brand be used only once one each car, which is called the *randomized complete block design* (RCBD). Now four brands are placed on a car at random and each car gets one tire of each brand. This provides a more homogeneous condition to test the four brands of tires.

	Car			
	I	II	III	IV
Brand distribution	B	D	A	C
	C	C	B	D
	A	B	D	A
	D	A	C	B

- The cars are the "blocks", and randomization is restricted within blocks. This design allows the car (block) variation to be independently assessed and removed from the error term in a model.

$$Y_{ij} = \mu_{..} + \rho_i(\text{block effect}) + \tau_j(\text{tire effect}) + \epsilon_{ij}$$

Introduction of Blocking

- **Blocking** provide the mechanism for explaining and controlling variation among the experimental units from sources that are not of interest to you. Block designs help maintain internal validity, by reducing the possibility that the observed treatment effects are due to a confounding or nuisance factor.
- A **nuisance** factor is a factor that has some effect on the response, but is of no interest to the experimenter; however, the variability it transmits to the response needs to be minimized or explained. We will have to account for these nuisance factors in our design and analysis.
  - *Treatment factors*, which we are interested (in studying the effects/differences in the experiments or clinical trials)
  - *Blocking factors*, which are often not our interest in the experiment but have to be controlled for.
- Typical nuisance factors include **batches** of raw material if you are in a production situation, different **operators** of the study units, when studying a *process*, and **time (shifts, days, etc.)** where the time of the day or the shift can be a factor that influences the response. In studies involving human subjects, we often use **gender and age classes** as the blocking factors. Often, in medical studies, the blocking factor used is the **type of institution** (size of the institution, types of patient populations, hospitals versus clinics, etc).
- The design we looked at so far is the **completely randomized design (CRD)** where we only have one factor or two factors. In the CRD setting we simply randomly assign the treatments to the available experimental units in our experiment.
- When we have a single blocking factor, we can utilize a randomized complete **block** design (**RCBD**): quite efficient in terms of power and reducing the error variation, with very simple data structure and analysis.
  - To *randomize* for a RCBD design, within each block, a random permutation is used to assign all the treatments to the experiment units, just like in a CRD design.
  - Independent randomizations (randomized treatment assignments) are used for different blocks.

Data Example

The data from a RCBD can be summarized as:

	Treatment (j)			
Block (i)	1	2	$\cdots$	r
1	y_{11}	y_{21}	$\cdots$	y_{r1}
2	y_{12}	y_{22}	$\cdots$	y_{r2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_b	y_{1b}	y_{2b}	$\cdots$	y_{r,n_b}

Example: In an experiment on decision making, executives were exposed to one of *three methods* : utility method, the worry method, and the comparison method. After using the assigned method, the subjects were asked to state their degree of confidence in the method of quantifying the risk on a scale from 0 (no confidence) to 20 (highest confidence).

Fifteen subjects were used in the study. They were grouped into *five blocks* of three executives, according to **age.** Block 1 contained the three oldest executives, and so on.

C : Comparison method
W : Worry method
U : Utility method

	Experimental Unit		
	1	2	3
Block 1 (oldest executives)	C	W	U
2	C	U	W
3	U	W	C
4	W	U	C
5 (youngest executives)	W	C	U

TABLE 21.1 Data on Confidence Ratings (ratings on scale from 0 to 20)—Risk Premium Example.

Block	Method (j)			
i	Utility	Worry	Comparison	Average
1 (oldest)	1	5	8	4.7
2	2	8	14	8.0
3	7	9	16	10.7
4	6	13	18	12.3
5 (youngest)	12	14	17	14.3
Average	5.6	9.8	14.6	10.0

R analysis: 1. Read the data

```r
# read data from week5 folder online
Ex21  =read.table(
        url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week7/CH21TA01.txt"))



names(Ex21) =  c("Rating", "AgeBlock", "Method")
```

```r
# make  categorical variables for factor A and B
Ex21$AgeBlock    =  as.factor(Ex21$AgeBlock)
Ex21$Method =  as.factor(Ex21$Method)

levels(Ex21$Method)= c("Utility", "Worry" , "Comparison")

str(Ex21)
```

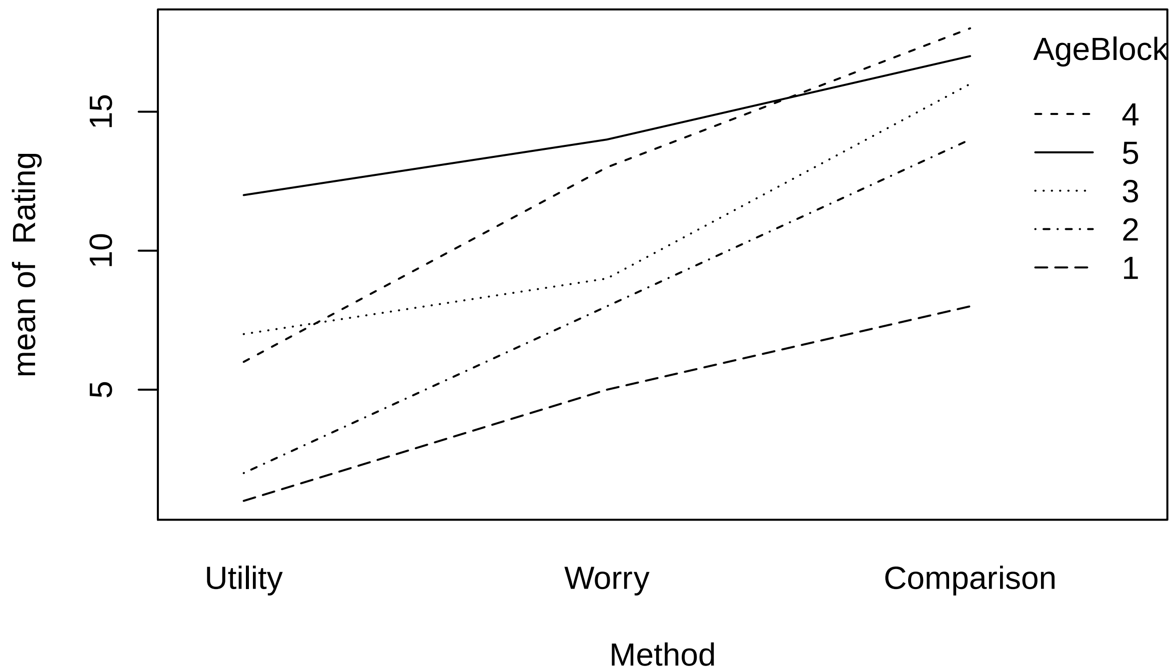
```
## 'data.frame':    15 obs. of  3 variables:
##  $ Rating  : int  1 5 8 2 8 14 7 9 16 6 ...
##  $ AgeBlock: Factor w/ 5 levels "1","2","3","4",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ Method  : Factor w/ 3 levels "Utility","Worry",..: 1 2 3 1 2 3 1 2 3 1 ...
```

```r
Ex21
```

```
##    Rating AgeBlock     Method
## 1       1        1    Utility
## 2       5        1      Worry
## 3       8        1 Comparison
## 4       2        2    Utility
## 5       8        2      Worry
## 6      14        2 Comparison
## 7       7        3    Utility
## 8       9        3      Worry
## 9      16        3 Comparison
## 10      6        4    Utility
## 11     13        4      Worry
## 12     18        4 Comparison
## 13     12        5    Utility
## 14     14        5      Worry
## 15     17        5 Comparison
```

2. Make a treatment plot by blocks

```r
with(Ex21, interaction.plot(x.factor = Method, trace.factor = AgeBlock, response = Rating))
```



Finding: It appears

1. there are some variations between blocks.
2. Within each block, ratings differ by the three methods.
3. It also appears that there are no important interaction effects between blocks and treatments on the responses; the response curves do not seem to deviate too much from being parallel.

Model for RCBD (Ch 21.2)

- In the most basic form, we assume that we do not have replicates within a block. This means that we only see every treatment **once** in each block.
- The analysis of a randomized complete block design is straightforward. We treat the block factor as "another" factor in our model. As we have no replicates within blocks, we can only fit a main effects model, exactly the same as we discussed in Chapter 20 (one case per treatment).

Factor Effects Model

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

- $\mu_{..}$ 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).
- Y_{ij} is the response for the jth treatment in the ith block, with

$$E(Y_{ij}) = \mu_{..} + \rho_i + \tau_j$$

Because we have only one case for each treatment with a block, we do not have enough information to estimate the interaction of the treatment and block. We assume no interaction.

Parameter estimation (Ch 21.3)

The model parameter estimator are similar to what we discussed in Chapter 20, we can rewrite the observations as follows

$$Y_{ij} = \overline{Y}_{..} + (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})$$

Compared with the model, we can estimate the *overall mean*, the ith block effect, jth treatment effect, the fitted value of Y_{ij} and the residual as follows:

Fitting of Randomized Complete Block Model

The least squares and maximum likelihood estimators of the parameters in randomized block model (21.1) are obtained in the customary fashion and again are the same. Employing our usual notation, they are:

Parameter	Estimator	
$\mu_{..}$	$\hat{\mu}_{..} = \overline{Y}_{..}$	**(21.3a)**
ρ_i	$\hat{\rho}_i = \overline{Y}_{i.} - \overline{Y}_{..}$	**(21.3b)**
τ_j	$\hat{\tau}_j = \overline{Y}_{.j} - \overline{Y}_{..}$	**(21.3c)**

The fitted values therefore are:

$$\hat{Y}_{ij} = \overline{Y}_{..} + (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) = \overline{Y}_{i.} + \overline{Y}_{.j} - \overline{Y}_{..} \quad \textbf{(21.4)}$$

and the residuals are:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..} \quad \textbf{(21.5)}$$

SS partition and ANOVA table (Ch 21.3)

Because we can partition the total variation :

$$Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})$$

Then we sum over all the observations,

$$SSTO = SSBL + SSTR + SSBL.TR$$

where

Sum of squares for blocks $$SSBL = r\sum_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2 \quad \textbf{(21.6a)}$$

Treatment sum of squares $$SSTR = n_b\sum_j(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})^2 \quad \textbf{(21.6b)}$$

Sum of square of residuals (Interaction sum of squares between blocks & treatments) $$SSBL.TR = \sum_i\sum_i(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot})^2 = \sum_i\sum_i e_{ij}^2 \quad \textbf{(21.6c)}$$

Then we can obtain the ANOVA table for RCBD design and derive $E(MS)$ to see why different MS are used to measure the estimate different source of variation due to blocks or treatment, and MSBL.TR is used to estimate the variance σ^2.

TABLE 21.2 ANOVA Table for Randomized Complete Block Design, Block Effects Fixed.

Source of Variation	SS	df	MS	E{MS}
Blocks	$SSBL$	$n_b - 1$	$MSBL$	$\sigma^2 + r\frac{\sum \rho_i^2}{n_b - 1}$
Treatments	$SSTR$	$r - 1$	$MSTR$	$\sigma^2 + n_b\frac{\sum \tau_j^2}{r - 1}$
Error	$SSBL.TR$	$(n_b - 1)(r - 1)$	$MSBL.TR$	σ^2
Total	$SSTO$	$n_b r - 1$		

Hypothesis testing:

- Testing treatment main effect:

$$F^* = MSTR/MSBL.TR \sim F(r-1, (n_b-1)(r-1))$$

.

- Testing block effect (sometimes not of interest):

$$F^* = MSBL/MSBL.TR \sim F(n_b-1, (n_b-1)(r-1))$$

.

- We reject the null hypothesis of no treatment effect or no block effect when the corresponding F^* is large.

Checking Assumptions (Diagnostics) (Ch 21.4)

Assumptions

– Model is additive (no interaction between treatment effects and block effects) (additivity assumption)

– Errors are independent and normally distributed

– Constant variance (across blocks, across treatments)

Checking normality:

- Histogram, QQ plot of residuals, Shapiro-Wilk Test.

Checking constant variance

– Residual Plot: Residuals vs. fitted

– Residuals vs blocks

– Residuals vs treatments

-Run statistical tests (Chapter 18)

Additivity

- treatment by block plot to check interaction: treatments effects can be very different across blocks
- Tukey's One-df test of additivity
- If interaction exists, usually try to use transformation to eliminate interaction

Example: Analysis of RCBD

1. Fit two factor main effects model:

```
fit = aov(Rating~ AgeBlock +  Method, data= Ex21 )
summary(fit)
```

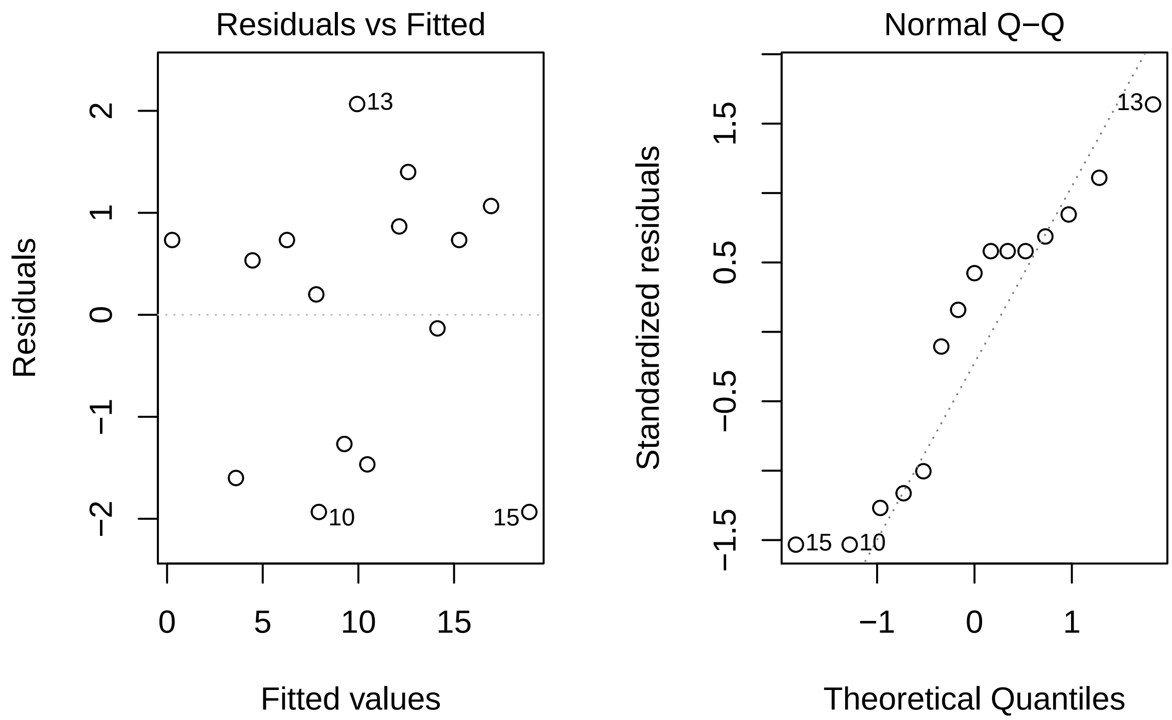
```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## AgeBlock     4 171.33   42.83   14.36 0.001008 ** 
## Method       2 202.80  101.40   33.99 0.000123 ***
## Residuals    8  23.87    2.98                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results:

1. For the treatment effect, F-statistic = 33.99, $P = 0.000123$, we reject the null and conclude that mean confidence ratings for the three methods differ.
2. For the block effects (age groups), F-statistic = 14.36, $P = 0.001008$, we reject the null and conclude that there were significant block effects (although this is the factor that they try to control for, not the main objective of the study).

2. Mode diagnostics to check constant variance and normal error.

```
par(mfrow=c(1,2))
plot(fit,1, add.smooth = F)
plot(fit,2)
```



Overall, there are no substantial departure from model assumptions.

3. Run Tukey's test for Addivity.

```
library(additivityTests)

## make the data into matrix A*B Form (dim a * b)
(Ex21m =  matrix( Ex21$Rating, nrow=5, ncol=3, byrow=T ))
```

```
##      [,1] [,2] [,3]
## [1,]    1    5    8
## [2,]    2    8   14
## [3,]    7    9   16
## [4,]    6   13   18
## [5,]   12   14   17
```

```
##
tukey.test( Ex21m, alpha = 0.05)
```

```
##
## Tukey test on 5% alpha-level:
##
## Test statistic: 0.0779
## Critival value: 5.591
## The additivity hypothesis cannot be rejected.
```

Therefore, we conclude that the interaction effects are not present and the above model seems appropriate.

4. Nonparametric tests (page 900)

Friedman's Test is a rank-based test. It can be applied to data from Randomized Complete Block Designs when the parametric approach is inappropriate. For example, if the normality assumption is violated or we only have ordinal data (such as ranks).

- It assumes that the variable of interest is continuous, and there is no interaction between blocks and treatments. The null hypothesis of the test is that the treatment effects are identical.
- *R* has a buid-in function *friedman.test()* to run nonparametric version of the ANOVA model with a block factor.

```
friedman.test(Rating ~ Method |AgeBlock , data= Ex21)
```

```
##
##  Friedman rank sum test
##
## data:  Rating and Method and AgeBlock
## Friedman chi-squared = 10, df = 2, p-value = 0.006738
```

Results: Here the Friedman's test shows similar findings: we reject null and there is a significant treatment difference: mean ratings differed for the three methods.

Analysis of Treatment Effects (Ch 21.5)

This is similar to the inference procedure we studied before. Since we are not interested in studying the block effects, we estimate the treatment means the same way as we estimate the factor B jth level mean as in a two-factor study (no interaction model): $\hat{\mu}_{.j} = \overline{Y}_{.j}$, where $\mu_{.j}$ is the mean response for treatment j averaged over all blocks.

The formulas in Chapter 17 or 19 for estimating contrasts of the treatment means apply here. The appropriate mean square term to be used in the estimated variance of the contrast is $MSBL.TR$ with the corresponding $df = (n_b - 1)(r - 1)$.

The multiples for the estimated standard deviation of the contrast and used to construct CI are now as follows:

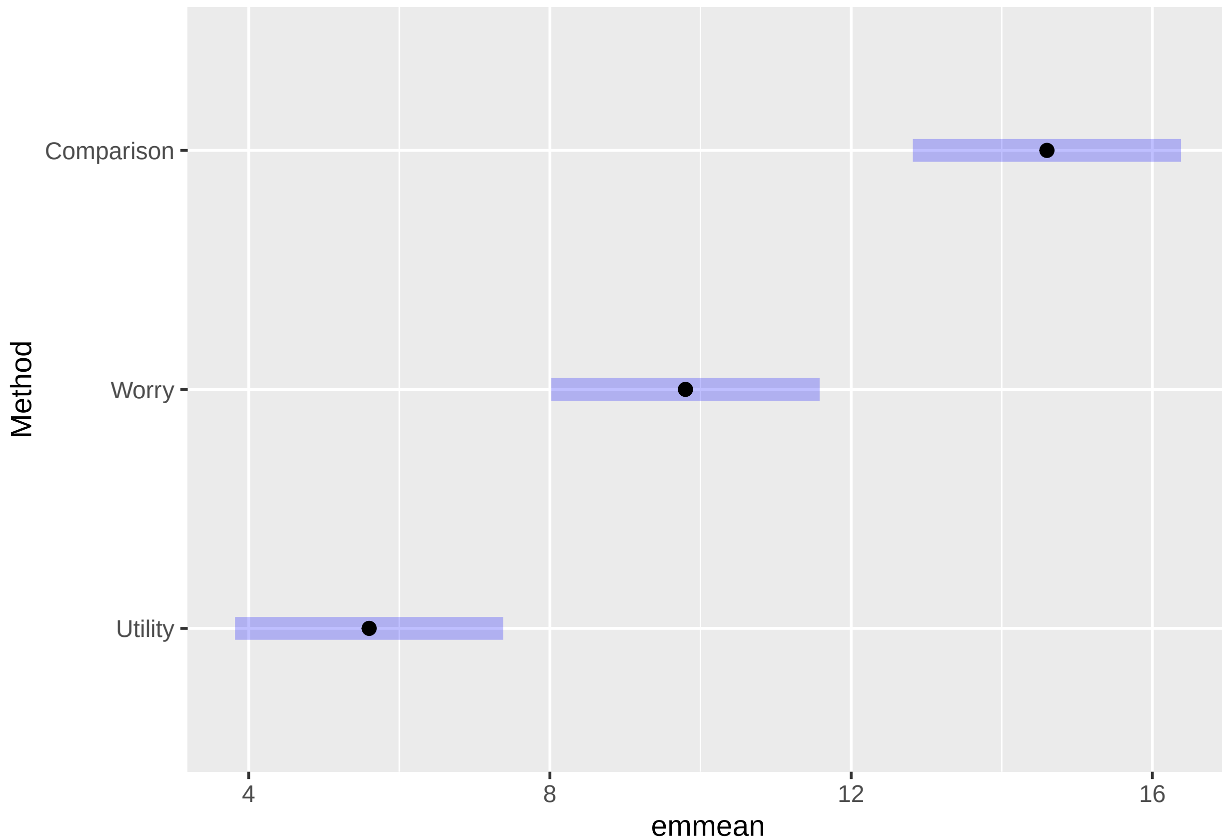
Single comparison	$t[1-\alpha/2; (n_b-1)(r-1)]$	**(21.9a)**
Tukey procedure (for pairwise comparisons)	$T = \frac{1}{\sqrt{2}} q[1-\alpha; r, (n_b-1)(r-1)]$	**(21.9b)**
Scheffé procedure	$S^2 = (r-1)F[1-\alpha; r-1, (n_b-1)(r-1)]$	**(21.9c)**
Bonferroni procedure	$B = t[1-\alpha/2g; (n_b-1)(r-1)]$	**(21.9d)**

Example:

The researcher who conducted the risk premium study was satisfied, on the basis of the residual analyses and tests, that randomized complete block model (21.1) is appropriate for the experiment. To analyze the treatment effects formally, the researcher wished to obtain all pairwise comparisons with a 95 percent family confidence coefficient, utilizing the Tukey procedure.

Here, we can run this analysis in R using functions from *emmeans* package.

```
library(emmeans)
# fit = aov(Rating~ AgeBlock +  Method, data= Ex21 )
fit.emm <- emmeans(fit, ~ Method)
plot(fit.emm)
```




```
pairs(fit.emm, adjust = "Tukey")
```

```
##  contrast              estimate   SE df t.ratio p.value
##  Utility - Worry           -4.2 1.09  8 -3.845  0.0121
##  Utility - Comparison      -9.0 1.09  8 -8.239  0.0001
##  Worry - Comparison        -4.8 1.09  8 -4.394  0.0058
##
## Results are averaged over the levels of: AgeBlock
## P value adjustment: tukey method for comparing a family of 3 estimates
```

```
confint(pairs(fit.emm))
```

```
##  contrast              estimate   SE df lower.CL upper.CL
##  Utility - Worry           -4.2 1.09  8    -7.32    -1.08
##  Utility - Comparison      -9.0 1.09  8   -12.12    -5.88
```

```
##  Worry - Comparison       -4.8 1.09  8    -7.92    -1.68
##
## Results are averaged over the levels of: AgeBlock
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

Results: (The order of paired comparison in the textbook are in the opposite direction, A-B vs. B-A. Other than that, the results are exactly the same.)

From the Tukey's test, all the adjusted p-value <0.05, suggesting that the paired differences <0. Therefore, we conclude Utility < Worry , Utility < Comparison, and Worry < Comparison were all significant. Also all the CIs did not include zero.

Therefore, with family confidence coefficient of .95, we conclude that that the comparison method has a higher mean confidence rating than the worry method, which in turn has a higher mean confidence rating than the utility method, after we have adjusted for the age group (block) effect.

Summary this week

- Reading: Chapter 20, Chapter 21.1- 21.5
- HW: 20.2, 20.3, 20.4 (Due 10/17 by 6 pm before the class)