stat631finals

Jessica Grover

5/6/2022

# Problem 1

# 1. What are sources of variations in the data?

Ans. The sources of variations are training programs, FBI agents and field offices.

# 2. Which factors are random and which factors are fixed?

Ans. Fixed factor is training programs and random factors are field offices and FBI agents.

# 3. Are factors crossed/nested with each others?

Ans. Yes, the factors are nested with each others. The factor office is nested into training programs.

# 4. Apply linear mixed model approach or ANOVA method to evaluate the research hypothesis that there is a difference among the mean test scores of the three types of training programs. State H0, H1, p-value and conclusion.

Null hypothesis (H0): The mean test scores of all the training programs are same. Alternative hypothesis (H1): There is a difference in mean test scores.

library(pacman)  
p\_load(lme4, lmerTest)  
data1<-c(59,81,62,48,50,53,51,54,45,52,46,65,47,48,42,44,51,58,61,52,66,43,51,49,52,55,49, 63,55,59,41,55,53,42,53,62,64,45,66,41,46,68,50,50,48,51,79,59,44,46,26,47,60,56,50,54,65, 40,44,65,42,49,45,60,60,67,40,49,56,50,50,60,66,61,61,44,49,52,51,49,46,69,45,67,46, 42,38,42,49,60)  
management <- rep(c(1,1,1,2,2,2,3,3,3),10);management <- as.factor(management)  
fbi <- rep(c(1:3), 30);fbi <- as.factor(fbi)  
df <- data.frame(data1,management,fbi)  
df

## data1 management fbi  
## 1 59 1 1  
## 2 81 1 2  
## 3 62 1 3  
## 4 48 2 1  
## 5 50 2 2  
## 6 53 2 3  
## 7 51 3 1  
## 8 54 3 2  
## 9 45 3 3  
## 10 52 1 1  
## 11 46 1 2  
## 12 65 1 3  
## 13 47 2 1  
## 14 48 2 2  
## 15 42 2 3  
## 16 44 3 1  
## 17 51 3 2  
## 18 58 3 3  
## 19 61 1 1  
## 20 52 1 2  
## 21 66 1 3  
## 22 43 2 1  
## 23 51 2 2  
## 24 49 2 3  
## 25 52 3 1  
## 26 55 3 2  
## 27 49 3 3  
## 28 63 1 1  
## 29 55 1 2  
## 30 59 1 3  
## 31 41 2 1  
## 32 55 2 2  
## 33 53 2 3  
## 34 42 3 1  
## 35 53 3 2  
## 36 62 3 3  
## 37 64 1 1  
## 38 45 1 2  
## 39 66 1 3  
## 40 41 2 1  
## 41 46 2 2  
## 42 68 2 3  
## 43 50 3 1  
## 44 50 3 2  
## 45 48 3 3  
## 46 51 1 1  
## 47 79 1 2  
## 48 59 1 3  
## 49 44 2 1  
## 50 46 2 2  
## 51 26 2 3  
## 52 47 3 1  
## 53 60 3 2  
## 54 56 3 3  
## 55 50 1 1  
## 56 54 1 2  
## 57 65 1 3  
## 58 40 2 1  
## 59 44 2 2  
## 60 65 2 3  
## 61 42 3 1  
## 62 49 3 2  
## 63 45 3 3  
## 64 60 1 1  
## 65 60 1 2  
## 66 67 1 3  
## 67 40 2 1  
## 68 49 2 2  
## 69 56 2 3  
## 70 50 3 1  
## 71 50 3 2  
## 72 60 3 3  
## 73 66 1 1  
## 74 61 1 2  
## 75 61 1 3  
## 76 44 2 1  
## 77 49 2 2  
## 78 52 2 3  
## 79 51 3 1  
## 80 49 3 2  
## 81 46 3 3  
## 82 69 1 1  
## 83 45 1 2  
## 84 67 1 3  
## 85 46 2 1  
## 86 42 2 2  
## 87 38 2 3  
## 88 42 3 1  
## 89 49 3 2  
## 90 60 3 3

Using linear mixed model approach,

model1 <- lmer(data1~management+(1|fbi:management),data = df)  
summary(model1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: data1 ~ management + (1 | fbi:management)  
## Data: df  
##   
## REML criterion at convergence: 606.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.1108 -0.5887 0.0283 0.4658 3.0064   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## fbi:management (Intercept) 5.029 2.243   
## Residual 53.062 7.284   
## Number of obs: 90, groups: fbi:management, 9  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 60.333 1.856 6.000 32.505 5.64e-08 \*\*\*  
## management2 -13.133 2.625 6.000 -5.003 0.00244 \*\*   
## management3 -9.667 2.625 6.000 -3.683 0.01030 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) mngmn2  
## management2 -0.707   
## management3 -0.707 0.500

Using ANOVA,

anova(model1)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## management 1427 713.48 2 6 13.446 0.00607 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Since the p-value is significant at 95% confidence level, we are rejecting null hypothesis. This also proves that at least one mean test scores of training program is different.

# 5. If you reject H0 in 4, use Dunnett’s method to find the best type of training programs.

library(emmeans)  
lsm\_mod <- lsmeans(model1,~management)  
dunet <- summary(contrast(lsm\_mod,list("conflictmanagement vs psychology"=c(1,-1,0),"conflict managementdunet vs negotiation"=c(1,0,-1),"psychology v negotiation"=c(0,1,-1)),adjust="dunet"),infer=c(T,F),level=0.95,side="two-sided");  
dunet

## contrast estimate SE df lower.CL upper.CL  
## conflictmanagement vs psychology 13.13 2.62 6 4.50 21.76  
## conflict managementdunet vs negotiation 9.67 2.62 6 1.04 18.30  
## psychology v negotiation -3.47 2.62 6 -12.10 5.16  
##   
## Degrees-of-freedom method: kenward-roger   
## Confidence level used: 0.95   
## Conf-level adjustment: bonferroni method for 3 estimates

Conclusion: The best type of training program is conflict management because 0 is not in the interval.

# Probelm 2

# 1. Write down the model and Hasse diagram for this data analysis.

data1 <- c(192,195,292,249,190,203,218,210,214,139,245,163,221,152,204,134)  
cows <- rep(c(1:4),4);cows <- as.factor(cows)  
period <- rep(c(1:4),each=4);period <- as.factor(period)  
t<- c(4,1,3,2,1,4,2,3,3,2,1,4,2,3,4,1)  
t <- as.factor(t)  
model2 <- data.frame(data1,cows,period,t)  
model2

## data1 cows period t  
## 1 192 1 1 4  
## 2 195 2 1 1  
## 3 292 3 1 3  
## 4 249 4 1 2  
## 5 190 1 2 1  
## 6 203 2 2 4  
## 7 218 3 2 2  
## 8 210 4 2 3  
## 9 214 1 3 3  
## 10 139 2 3 2  
## 11 245 3 3 1  
## 12 163 4 3 4  
## 13 221 1 4 2  
## 14 152 2 4 3  
## 15 204 3 4 4  
## 16 134 4 4 1

The factors are: cows, time period, diets. Fixed effects are: cows, diets Random effects are: Time period Crossed factors are: time period, cow

# 2. Which type of designs is this?

As the treatments, the row factor and the column factor affects response independently, we can conclude that it is latin square design.

# 3. Use linear mixed model approach or ANOVA method to analyze the effects of four diets.

effects1<- lmer(data1~cows+(1|period)+t,data=model2)  
summary(effects1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: data1 ~ cows + (1 | period) + t  
## Data: model2  
##   
## REML criterion at convergence: 99.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -0.98081 -0.49750 -0.09658 0.58664 1.17397   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## period (Intercept) 235.6 15.35   
## Residual 1237.2 35.17   
## Number of obs: 16, groups: period, 4  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 193.938 24.499 8.564 7.916 3.21e-05 \*\*\*  
## cows2 -32.000 24.872 6.000 -1.287 0.246   
## cows3 35.500 24.872 6.000 1.427 0.203   
## cows4 -15.250 24.872 6.000 -0.613 0.562   
## t2 15.750 24.872 6.000 0.633 0.550   
## t3 26.000 24.872 6.000 1.045 0.336   
## t4 -0.500 24.872 6.000 -0.020 0.985   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) cows2 cows3 cows4 t2 t3   
## cows2 -0.508   
## cows3 -0.508 0.500   
## cows4 -0.508 0.500 0.500   
## t2 -0.508 0.000 0.000 0.000   
## t3 -0.508 0.000 0.000 0.000 0.500   
## t4 -0.508 0.000 0.000 0.000 0.500 0.500

anova(effects1)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)  
## cows 9929.2 3309.7 3 6 2.6751 0.1409  
## t 1995.7 665.2 3 6 0.5377 0.6736

Since the p-value of treatment is 0.6736 which is greater than 0.05, we can say that there is no significant effect because of the diet. Hence we fail to reject null hypothesis.

library(ggplot2)  
ggplot(model2,aes(x=period,y=cows))+ geom\_point(aes(x=period, y=cows))



geom\_line(aes(x=period,y=cows))

## mapping: x = ~period, y = ~cows   
## geom\_line: na.rm = FALSE, orientation = NA  
## stat\_identity: na.rm = FALSE  
## position\_identity

ggtitle("xyplot with parallel regression lines (ANCOVA)")

## $title  
## [1] "xyplot with parallel regression lines (ANCOVA)"  
##   
## attr(,"class")  
## [1] "labels"

# 4. Construct 95% Tukey’s pairwise comparisons of the diets and report your findings.

lsm\_mod <- lsmeans(effects1,~cows)  
tukey <- summary(contrast(lsm\_mod,method = "pairwise",adjust="tukey"),infer=c(T,F),level=0.95,side="two-sided");  
tukey

## contrast estimate SE df lower.CL upper.CL  
## 1 - 2 32.0 24.9 6 -54.1 118.1  
## 1 - 3 -35.5 24.9 6 -121.6 50.6  
## 1 - 4 15.2 24.9 6 -70.8 101.3  
## 2 - 3 -67.5 24.9 6 -153.6 18.6  
## 2 - 4 -16.8 24.9 6 -102.8 69.3  
## 3 - 4 50.8 24.9 6 -35.3 136.8  
##   
## Results are averaged over the levels of: t   
## Degrees-of-freedom method: kenward-roger   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 4 estimates

With the tukey’s method we can say that all the intervals include zero with 95% confidence intervals. The effects of treatments are same.

# Problem 3

# 1. What are sources of variations in the data?

The sources of variations are : cows, diets, proteins

# 2. Which factors are random and which factors are fixed?

Fixed factors are: diets Random factors are: cows, proteins

# 5. Based on the answers of the previous questions, apply linear mixed model approach to evaluate the research hypothesis that there is a difference between the three diets. State H0, H1, p-value and conclusion.

prob3<-read.csv("Prob3.csv")  
head(prob3)

## Diet Cow week protein  
## 1 Barley 1 1 3.63  
## 2 Barley 2 1 3.24  
## 3 Barley 3 1 3.98  
## 4 Barley 4 1 3.66  
## 5 Barley 5 1 4.34  
## 6 Barley 6 1 4.36

model4<-lmer(protein~1 + Diet+(1|week)+(1|Cow),data=prob3)  
model4

## Linear mixed model fit by REML ['lmerModLmerTest']  
## Formula: protein ~ 1 + Diet + (1 | week) + (1 | Cow)  
## Data: prob3  
## REML criterion at convergence: 74.8357  
## Random effects:  
## Groups Name Std.Dev.  
## Cow (Intercept) 0.1288   
## week (Intercept) 0.2500   
## Residual 0.2914   
## Number of obs: 120, groups: Cow, 10; week, 4  
## Fixed Effects:  
## (Intercept) DietLupins DietMixed   
## 3.5825 -0.1550 -0.0905

anova(model4)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## Diet 0.48501 0.2425 2 105 2.856 0.06197 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1