STAT 581 - Exam 2: Due Dec 14, 2021

Alex Towell (atowell@siue.edu)

Problem 1

A randomized complete block design is used to study the effect of caliper on the measured diameters (in mm) of ball bearings. A sample of $b = 10$ ball bearings is randomly selected, and each of $a = 3$ calipers produces a measurement on each of the selected ball bearings. The data is provided as an attachment.

P1: (c)

Test for systematic differences in the measurements provided by the calipers. Compute the F_0 statistic, and the p-value. Provide an interpretation, stated in the context of the problem.

```
library("readxl")
#library("lme4")
library("lmerTest")

data = read_excel("exam2data.xlsx")
A = as.factor(na.omit(data$caliper)) # fixed effect
B = as.factor(na.omit(data$ball.bearing)) # random effect
y = na.omit(data$diameter) # response
head(data.frame(caliper=A,ball.bearing=B,diameter=y))
```

```
##   caliper ball.bearing diameter
## 1       1            1    26.88
## 2       1            2    26.53
## 3       1            3    26.58
## 4       1            4    26.86
## 5       1            5    26.33
## 6       1            6    26.60
```

```
# fixed effect tau1 + tau2 + tau3 = 0 (for calipers)
contrasts(A)=contr.sum

random.mod = lmer(y ~ (1|B) + A)

# the anova command is used to compute the test for fixed effects.
anova(random.mod)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##    Sum Sq  Mean Sq NumDF DenDF F value  Pr(>F)
## A 0.10085 0.050423     2    18  5.5157 0.01354 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P1: (d)

Compute estimates of the variance components. Explain when a block design is better than a completely randomized design.

```
print(VarCorr(random.mod),comp="Variance")
```

```
## Groups   Name        Variance
## B        (Intercept) 0.0411089
## Residual             0.0091419
```

P1: (e)

Compute estimates of the fixed effect parameters. Explain why block effects are modeled differently than treatment effects in this design. Explain what treatment effect is estimable in this design.

```
est = coef(summary(random.mod))[1:3,1]
est.tau.hat = c(est,0-sum(est[2:3]))
names(est.tau.hat) = c("mu","tau1","tau2","tau3")
round(est.tau.hat,digits=3)
```

```
##     mu   tau1   tau2   tau3
## 26.532  0.081 -0.032 -0.050
```

Problem 2

Now, a mixed effects design is used to study the effect of caliper (fixed effect, factor A) on the measured diameters of ball bearings. There are $a = 2$ calipers under investigation. A random sample of $b = 8$ ball bearings is selected (random effect, factor B), and each caliper produces $n = 3$ measurements on each of the selected ball bearings. The data is provided as an attachment.

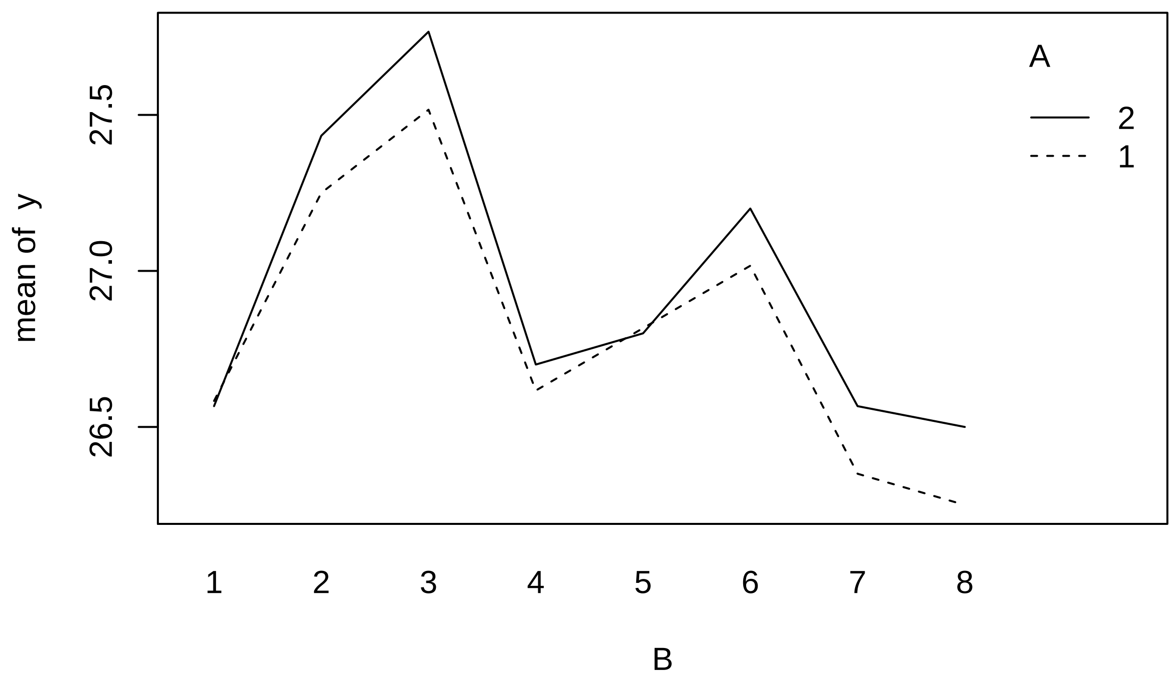
P2: (b)

Create an interaction plot to display the caliper effect on measured diameter. Use a mixed model likelihood approach to test for a systematic difference in the measurements of the two calipers. Compute F_0 and the p-value.

```
A = as.factor(na.omit(data$device)) # fixed effect
B = as.factor(na.omit(data$ball)) # random effect
y = na.omit(data$measurement) # response
contrasts(A)=contr.sum
head(data.frame(caliper=A,ball=B,measurement=y))
```

```
##   caliper ball measurement
## 1       1    1       26.55
## 2       1    1       26.45
## 3       1    1       26.75
## 4       1    2       27.25
## 5       1    2       27.15
## 6       1    2       27.35
```

```
interaction.plot(B,A,y)
```




```
mixed.mod = lmer(y ~ A + (1|B) + (1|A:B))
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(mixed.mod)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##    Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
## A 0.24083 0.24083     1    39  5.1916 0.02825 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P2: (d) and (e)

Perform a test for caliper effects. Compute F_A and the p-value. Compute the unbiased estimates of the random effect parameters.

```
mixed.test(A,B,y)
```

```
##                         SS df         MS
## Fixed Effect A  0.2408333  1 0.24083333
## Random Effect B 8.5491667  7 1.22130952
## Interaction AB  0.1291667  7 0.01845238
## Error           1.6800000 32 0.05250000
##  F-test for fixed effect     p-value
##                 13.05161 0.008593224
##  error.var interaction.var block.var
##     0.0525     -0.01134921 0.2004762
```

Problem 3

A nested design is used to study the number of cases produced from three bottling machines (factor A, fixed effect). Four operators are randomly selected for each of the machines (nested factor B, random effect). Each operator makes $n = 2$ experimental runs. The data is provided as an attachment.

P3: (d)

Test for differences between bottling machines. Write the test statistic F_A statistic as a ratio of mean squares.

```
A = as.factor(na.omit(data$machine))
B = as.factor(na.omit(data$operator))
y = na.omit(data$cases)
nested.test(A,B,y)
```

```
##                           SS df         MS
## Fixed Effect A      628.0833  2 314.041667
## Random Effect B(A)  885.7500  9  98.416667
## Error                80.0000 12   6.666667
##   F-test for fixed effect    p-value
##                    3.19094 0.08964972
##   error.var  B.var
##    6.666667 45.875
```

P3: (e)

Explain why $\mathrm{MS_E}$ is the incorrect error term to use when the nested factor is random. In particular, comment on the pertinent sample size.

Problem 4

A company wishes to study the effect of promotion type $(1, 2, 3)$ on the sales of its crackers. A sample of $N = 15$ grocery stores is selected. Response variable y is the number of cases sold during the promotion period. Factor A is the promotion type. Covariate x is the same store sales prior to the promotion. The data is provided as an attachment.

P4: (a)

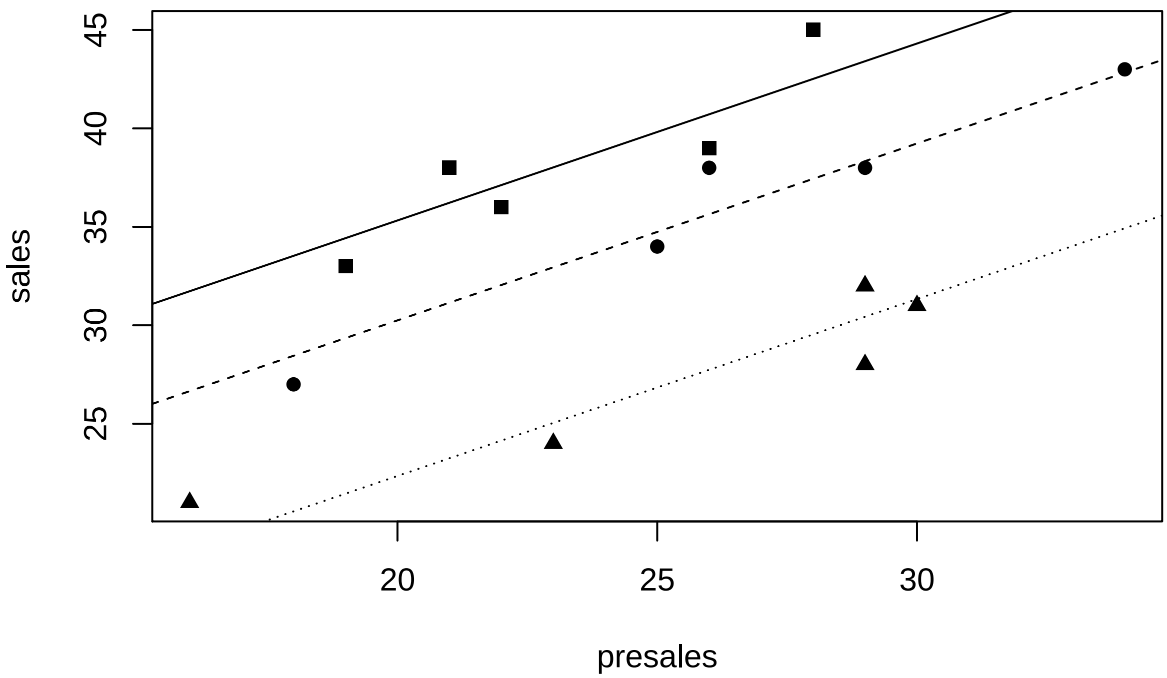
Compute the estimated regression of presales on cases sold for each promotion type. Create a scatterplot of presales versus sales for each promotion type, including the estimated regression lines.

```
prom = as.factor(na.omit(data$promotion)) # factor A
presales = na.omit(data$presales)         # predictor x
sales = na.omit(data$sales)               # response

ancova.mod = lm(sales ~ presales + prom)
est = coef(ancova.mod)
b1 = est[2]
b0 = c(est[1],est[1]+est[3],est[1]+est[4])

reg.funcs = matrix(c(b0,rep(b1,3)),nrow = 3)
dimnames(reg.funcs)=list(c("prom 1","prom 2","prom 3"),c("intercept","slope"))
reg.funcs
```

```
##        intercept     slope
## prom 1 17.353421 0.8985594
```

```
## prom 2 12.278031 0.8985594
## prom 3  4.376591 0.8985594
```

```
# scatterplot
plot(presales[prom==1],sales[prom==1],xlab='presales',ylab='sales', pch=15,
     xlim=c(min(presales),max(presales)),
     ylim=c(min(sales),max(sales)))
points(presales[prom==2],sales[prom==2], pch=16)
points(presales[prom==3],sales[prom==3], pch=17)

abline(b0[1],b1,lty=1)
abline(b0[2],b1,lty=2)
abline(b0[3],b1,lty=3)
```



P4: (b)

Test for a promotion effect. Write the `ANCOVA` $F_{A|x}$ statistic using extra sum of squares notation. Compute $F_{A|x}$ and the p-value. Provide an interpretation, stated in the context of the problem. Note the role the covariate is playing in this analysis.

```
anova(ancova.mod)
```

```
## Analysis of Variance Table
##
## Response: sales
##           Df Sum Sq Mean Sq F value    Pr(>F)
## presales   1 190.68 190.678  54.379 1.405e-05 ***
## prom       2 417.15 208.575  59.483 1.264e-06 ***
## Residuals 11  38.57   3.506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P4: (c)

Compute the sample mean sales and the sample mean presales for each promotion type. Compute the least squares means. Explain how the information from the covariate adjusts the determination of promotion effect.

```
library("lsmeans")
library("car")

xbar.prom = aggregate(presales, by=list(prom), FUN=mean)
ybar.prom = aggregate(sales, by=list(prom), FUN=mean)[2]
means.table = cbind(xbar.prom,ybar.prom)
colnames(means.table) = c("prom","presales.mean","sales.mean")
means.table
```

```
##   prom presales.mean sales.mean
## 1    1          23.2       38.2
## 2    2          26.4       36.0
## 3    3          25.4       27.2
```

```
# overall mean
mean(presales)
```

```
## [1] 25
```

```
# compute least squares means for adjusted means.
lsmeans(ancova.mod,pairwise ~ prom,adjust="none")$lsmeans
```

```
##  prom lsmean    SE df lower.CL upper.CL
##  1      39.8 0.858 11     37.9     41.7
##  2      34.7 0.850 11     32.9     36.6
##  3      26.8 0.838 11     25.0     28.7
##
## Confidence level used: 0.95
```

Code

```
mixed.test = function(A,B,y)
{
  av=anova(lm(y~A*B))
  F.a = av$`Mean Sq`[1]/av$`Mean Sq`[3]
  p.value = pf(F.a,df1=av$Df[1],df2=av$Df[3],lower.tail = FALSE)
  table1 = matrix(c(av$`Sum Sq`[1],av$`Sum Sq`[2],av$`Sum Sq`[3],av$`Sum Sq`[4],
                    av$Df[1],av$Df[2],av$Df[3],av$Df[4],
                    av$`Mean Sq`[1],av$`Mean Sq`[2],av$`Mean Sq`[3],av$`Mean Sq`[4]),nrow = 4)
  dimnames(table1) = list(c("Fixed Effect A","Random Effect B","Interaction AB","Error"),
                          c("SS","df","MS"))
  print(table1)

  table2 = matrix(c(F.a,p.value),nrow = 1)
  dimnames(table2) = list(c(""),c("F-test for fixed effect","p-value"))
  print(table2)

  a=nlevels(A)
  b=nlevels(B)
  n=length(y) / a / b
```

```
  var.hat = av$`Mean Sq`[4]
  var.interaction.hat = (av$`Mean Sq`[3]-av$`Mean Sq`[4])/n
  var.block = (av$`Mean Sq`[2]-av$`Mean Sq`[3])/n/a

  table3 = matrix(c(var.hat,var.interaction.hat,var.block),nrow=1)
  dimnames(table3) = list(c(""),c("error.var","interaction.var","block.var"))
  print(table3)
}

nested.test = function(A,B,y)
{
  av=anova(lm(y~A/B))
  ss.A = av$`Sum Sq`[1]
  ss.B = av$`Sum Sq`[2]
  ss.error = av$`Sum Sq`[3]
  df.A = av$Df[1]
  df.B = av$Df[2]
  df.error = av$Df[3]
  ms.A = ss.A / df.A
  ms.B = ss.B / df.B
  ms.error = ss.error / df.error
  F.a = ms.A / ms.B
  p.value = pf(F.a,df1=df.A,df2=df.B,lower.tail = FALSE)
  table1 = matrix(c(ss.A,ss.B,ss.error,
                    df.A,df.B,df.error,
                    ms.A,ms.B,ms.error),nrow = 3)
  dimnames(table1) = list(c("Fixed Effect A","Random Effect B(A)","Error"),
                          c("SS","df","MS"))
  print(table1)

  table2 = matrix(c(F.a,p.value),nrow = 1)
  dimnames(table2) = list(c(""),c("F-test for fixed effect","p-value"))
  print(table2)

  a=nlevels(A)
  b=nlevels(B)
  n=length(y) / a / b

  var.hat = ms.error
  var.B.hat = (ms.B - ms.error) / n

  table3 = matrix(c(var.hat,var.B.hat),nrow=1)
  dimnames(table3) = list(c(""),c("error.var","B.var"))
  print(table3)
}
```