Initialize R

Initialize R by entering the following commands at the prompt. You must type the commands *exactly* as shown.

```r
options(contrasts=c("contr.sum","contr.poly") ) # set definition of contrasts
load(url("http://psycserv.mcmaster.ca/bennett/psy710/datasets/L5dat1.Rdata") )
load(url("http://psycserv.mcmaster.ca/bennett/psy710/datasets/L5dat2.Rdata") )
```

## Loading required package: MASS
## Loading required package: nnet

ab data

An experiment was done to measure the effects of treatment A and treatment B on a dependent variable, y, using a crossed-factorial design. The data are stored in the dataframe `L5.dat.1`. The following code snippets illustrate how to do several tasks that will help you answer the questions in this lab.

**Example:** The following code shows how to compute the number of observations in each cell of our experimental design:

```r
with(L5.dat.1,tapply(y,list(A,B),length) )
```

**Example:** The following code shows how to define a subset of data, specifically all of the rows of the dataframe `L5.dat.1` that has a value of `b1` on factor `B`:

```r
levels(L5.dat.1$B) # list names of levels
subset(L5.dat.1,B=="b1") # get subset
```

Tasks:

1. Create tables showing the mean, standard deviation, and n for each condition.

```r
with(L5.dat.1,tapply(y,list(A,B),mean) )
```

##b1  b2  b3
##a1 11.41 11.171 17.35
##a2 11.51 7.935 11.46

```r
with(L5.dat.1,tapply(y,list(A,B),sd) )
```
## b1  b2  b3
## a1 2.816 2.214 1.874
## a2 2.457 1.967 3.899
```r
with(L5.dat.1,tapply(y,list(A,B),length) )
```
## b1  b2  b3
## a1 6 6 6
## a2 6 6 6

2. Calculate the marginal means for A and B.

```r
with(L5.dat.1,tapply(y,A,mean) )  # marg means for levels of A
## a1  a2
## 13.31 10.30

with(L5.dat.1,tapply(y,B,mean) )  # marg means for levels of B
## b1  b2  b3

with(L5.dat.1,tapply(y,A,length) )  # n at each level of A
## a1  a2
## 18 18

with(L5.dat.1,tapply(y,B,length) )  # n at each level of B
## b1  b2  b3
## 12 12 12
```

3. List the data from all subjects who received the second level of treatment B.

```r
levels(L5.dat.1$B)
## [1] "b1" "b2" "b3"

subset(L5.dat.1,B=="b2")
##   y  A  B
## 7  9.814 a1 b2
## 8  8.911 a1 b2
## 9 13.598 a1 b2
## 10 13.660 a1 b2
## 11  8.998 a1 b2
```
4. Conduct an ANOVA that evaluates the effects of A and B on y. Explain your results.

```r
ab.aov.01 <- aov(y~A + B + A:B,data=L5.dat.1)
summary(ab.aov.01)
```

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>81.6</td>
<td>81.6</td>
<td>11.81</td>
<td>0.00174 **</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>143.5</td>
<td>71.7</td>
<td>10.38</td>
<td>0.00037 ***</td>
</tr>
<tr>
<td>A:B</td>
<td>2</td>
<td>54.1</td>
<td>27.1</td>
<td>3.92</td>
<td>0.03079 *</td>
</tr>
<tr>
<td>Residuals</td>
<td>30</td>
<td>207.3</td>
<td>6.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Answer:** There was a significant main effect of A ($F(1,30) = 11.81, p = .0017$), so we reject the null hypothesis of no differences between the marginal means of A. There was a main effect of B ($F(2,30) = 10.38, p = .00037$), so the null hypothesis of no difference among the marginal means of B is rejected. Finally, the A × B interaction was significant ($F(2,30) = 3.92, p = .031$), and therefore the effect of A depends on the level of B, and the effect of B depends on the level of A.

5. Interpreting interactions:

**Example:** Before computing the simple main effects, it might be helpful to graph the data. The following command created Figure 1.

```r
with(L5.dat.1,interaction.plot(B,A,y))
```

**Example:** The following code calculates the simple main effect of A at b1:

```r
MS.resid <- 6.909 # from main anova
df.resid <- 30   # from main anova
levels(L5.dat.1$B) # levels of B

## [1] "b1" "b2" "b3"

B.b1 <- subset(L5.dat.1,E=="b1") # get subset of data
aov.A.at.B1 <- aov(y~A,data=B.b1); # one-way anova of A at b1
summary(aov.A.at.B1) # anova table
```

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0.0</td>
<td>0.03</td>
<td>0</td>
<td>0.95</td>
</tr>
<tr>
<td>Residuals</td>
<td>10</td>
<td>69.8</td>
<td>6.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Interaction plot of the AB data.
Task: Calculate the simple main effect of B at each level of A.

**Answer:** Here is how you calculate the simple main effect of B at a1 and a2:

```r
MS.resid <- 6.909  # from main anova
df.resid <- 30  # from main anova
levels(L5.dat.1$A)

## [1] "a1" "a2"

A.a1 <- subset(L5.dat.1, A == "a1")
summary(aov(y ~ B, data = A.a1))

## Df Sum Sq Mean Sq F value  Pr(>F)
## B  2 147.3  73.6  13.5 0.00044 ***
## Residuals 15  81.7  5.4
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(F.B.a1 <- 73.634/MS.resid)

## [1] 10.66

(1-pf(F.B.a1, df1=2, df2=df.resid))

## [1] 0.0003185

A.a2 <- subset(L5.dat.1, A == "a2")
summary(aov(y ~ B, data = A.a2))

## Df Sum Sq Mean Sq F value Pr(>F)
## B  2  50.3 25.17  3.01 0.08 .
## Residuals 15 125.6 8.37
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(F.B.a2 <- 25.168/MS.resid)

## [1] 3.643

(1-pf(F.B.a2, df1=2, df2=df.resid))

## [1] 0.03834
```
The simple main effect of B at a1 is significant \((F(2, 30) = 10.65, p = .0003)\), as is the simple main effect of B at a2 \((F(2, 30) = 3.64, p = .038)\). Note that the second simple main effect would not have been significant if we had used the Bonferroni adjustment to maintain a familywise Type I error rate of .05.

6. Pairwise comparisons:

**Example:** The following code shows how to use `TukeyHSD` to do pairwise comparisons of each level on factor B. Note that the functions computes 90% adjusted confidence intervals, and the familywise \(\alpha\) therefore is 0.1:

```r
ab.aov.01 <- aov(y ~ A + B + A:B, data=L5.dat.1) # the anova
TukeyHSD(ab.aov.01, which="B", conf.level=0.90)
```

```
## Tukey multiple comparisons of means
## 90% family-wise confidence level

## Fit: aov(formula = y ~ A + B + A:B, data = L5.dat.1)

## $B
## diff  lwr  upr  p adj
## b2-b1 -1.904 -4.1939 0.3851 0.1952
## b3-b1 2.948 0.6588 5.2378 0.0264
## b3-b2 4.853 2.5632 7.1422 0.0003
```

**Task:** Use `TukeyHSD` to evaluate all pairwise comparisons of cell means while maintaining a familywise \(\alpha = 0.10\). Are all of these comparisons equally interesting?

```r
# answer:
TukeyHSD(ab.aov.01, which="A:B", conf.level=0.90)
```

```
## Tukey multiple comparisons of means
## 90% family-wise confidence level

## Fit: aov(formula = y ~ A + B + A:B, data = L5.dat.1)

## $`A:B`
## diff  lwr  upr  p adj
## a2:b1-a1:b1 0.09809 -4.0344 4.2306 1.0000
## a1:b2-a1:b1 -0.23713 -4.3696 3.8954 1.0000
## a2:b2-a1:b1 -3.47360 -7.6061 0.6589 0.2299
## a1:b3-a1:b1 5.94564 1.8131 10.0782 0.0058
## a2:b3-a1:b1 0.04903 -4.0835 4.1815 1.0000
## a2:b2-a2:b1 -0.33522 -4.4677 3.8073 0.9999
## a1:b3-a2:b1 5.84755 1.7150 9.9801 0.0069
## a2:b3-a2:b1 -0.04906 -4.1816 4.0834 1.0000
## a2:b2-a1:b2 -3.23647 -7.3690 0.8960 0.2985
## a1:b3-a1:b2 6.18277 2.0503 10.3153 0.0000
## a2:b3-a1:b2 0.28616 -3.8464 4.4187 1.0000
## a1:b3-a2:b2 9.41924 5.2867 13.5517 0.0000
```
Answer: Comparisons of cells within a single row or column are more interesting than comparisons in different rows and columns because cells in the latter kinds of comparisons differ on two factors and therefore are difficult to interpret.

7. Calculate Cohen’s $f$ for the $A \times B$ interaction. (For more information about calculation Cohen’s $f$ in a factorial design, see Section 7.8 in the notes.)

```
# answer:
F.AxB <- 3.9174  # AxB F taken from anova
df.AxB <- 2;   # from anova
N <- sum( with(L5.dat.1, tapply(y, B, length) ) )
(omega.AxB <- (df.AxB * (F.AxB - 1))/(df.AxB * (F.AxB - 1) + N))

## [1] 0.1395

(cohens.f <- sqrt(omega.AxB / (1-omega.AxB)))

## [1] 0.4026
```

Answer: The association strength and effect size are large.

cd data

This section, in which we analyze data from an unbalanced design, draws on material in Sections 7.12.3-7.12.12 in the course notes.

An experiment was done to measure the effects of treatment C and treatment D on a dependent variable, y, using a crossed-factorial design. Six subjects were assigned randomly to each condition, however the data from two subjects in one of the conditions were lost. The data are stored in the dataframe `L5.dat.2`.

1. Verify that the CD data are unbalanced.

```
with(L5.dat.2,tapply(y,list(C,D),length))
```

## b1 b2 b3
## a1 6 6 6
## a2 6 6 4

2. Verify that the results of the two-way ANOVA depend on the order of the terms in the full linear model.

```
cd.aov.01 <- aov(y~C+D+C:D, data=L5.dat.2)
cd.aov.02 <- aov(y~D+C+D:C, data=L5.dat.2)
summary(cd.aov.01)
```
### Df Sum Sq Mean Sq F value Pr(>F)
## C 1 65.3 65.3 11.40 0.0022 **
## D 2 187.1 93.5 16.32 2e-05 ***
## C:D 2 30.2 15.1 2.64 0.0892 .
## Residuals 28 160.4 5.7
### ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(cd.aov.02)

### Df Sum Sq Mean Sq F value Pr(>F)
## D 2 205.1 102.6 17.90 9.8e-06 ***
## C 1 4.7 4.7 8.25 0.0077 **
## D:C 2 30.2 15.1 2.64 0.0892 .
## Residuals 28 160.4 5.7
### ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

3. Your book defines Type I sums of squares as the sums of squares associated with one main effect when all other variables are ignored. According to this definition, what are the Type I sums of squares for C and D? What null hypotheses about the main effects are being evaluated with these Type I sums of squares?

C.t1.ss <- 65.338
D.t1.ss <- 205.121

4. What are the Type II sums of squares for C and D? Use Type II sums of squares to evaluate the main effects of C and D.

C.t2.ss <- 47.276
D.t2.ss <- 187.059

summary(cd.aov.01)

### Df Sum Sq Mean Sq F value Pr(>F)
## C 1 65.3 65.3 11.40 0.0022 **
## D 2 187.1 93.5 16.32 2e-05 ***
## C:D 2 30.2 15.1 2.64 0.0892 .
## Residuals 28 160.4 5.7
### ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Answer:** Using Type II sums-of-squares, the main effect of D (after controlling for C but ignoring the interaction) is significant ($F(2, 28) = 16.323$, $p < .0001$). Alternatively, we could use a model that sets the interaction to zero:
The SS value is the same, but the F and p values are different because the error term is estimated with a model that assumes that the interaction effects are all zero. Is this a good assumption in this case? No, because the F value suggests that there may be a small interaction effect (i.e., the CxD effect size is small, but non-zero). The following table shows that the Type II SS for C, after controlling for D but ignoring the interaction, is significant ($F(1,28) = 8.25, p = .0076)$:

The **car** package provides a simpler way of doing Type II tests. First, if the **car** package has not been installed previously, install it with the command `install.packages("car")`, and then load it with the command `library(car)`. Then use **car**'s **Anova** command to do the Type II tests. Note the capital A in the **Anova** command:

Anova(cd.aov.01,type="II")

## Anova Table (Type II tests)
##
## Response: y
##
## # Sum Sq Df  F value Pr(>F)
## C 47.3 1    8.25  0.0077 **
## D 187.1 2   16.32 2e-05 ***
## C:D 30.2 2   2.64  0.0892 .
## Residuals 160.4 28
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova(cd.aov.02,type="II")

## Anova Table (Type II tests)
##
## # Sum Sq Df  F value Pr(>F)
## D 205.1 2   17.90 9.8e-06 ***
## C 47.3 1    8.25  0.0077 **
## D:C 30.2 2   2.64  0.0892 .
## Residuals 160.4 28
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
5. Use `drop1` to compute the Type III sums of squares for C and D. Verify that these sums of squares do not depend on the order of terms in the model. What null hypotheses about the main effects are being evaluated with these Type III sums of squares? (See Section 7.12.7 in the course notes for an example of how to use `drop1`.)

**Answer:** The following code shows that the Type III sums of squares, as computed using `drop1`, are independent of the order of terms in the model:

```r
drop1(cd.aov.01, .~., test="F")
## Single term deletions
## Model:
## y ~ C + D + C:D
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 160 64.8
## C 1 52.2 213 72.3 9.11 0.0054 **
## D 2 170.6 331 85.4 14.89 3.9e-05 ***
## C:D 2 30.2 191 66.6 2.64 0.0892 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

drop1(cd.aov.02, .~., test="F")
## Single term deletions
## Model:
## y ~ D + C + D:C
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 160 64.8
## D 2 170.6 331 85.4 14.89 3.9e-05 ***
## C 1 52.2 213 72.3 9.11 0.0054 **
## D:C 2 30.2 191 66.6 2.64 0.0892 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

C.t3.ss <- 52.18
D.t3.ss <- 170.58
```
When effects are defined using the sum-to-zero constraint, Type III sums of squares can be used to test the null hypothesis of no difference among unweighted marginal means. One last thing: the following code shows how to compute Type III SSs with `Anova`:

```
Anova(cd.aov.01,type="III")
```

```
## Anova Table (Type III tests)
##
## Response: y
## Sum Sq Df F value Pr(>F)
## (Intercept) 4832 1 843.25 < 2e-16 ***
## C 52 1 9.11 0.0054 **
## D 171 2 14.89 3.9e-05 ***
## C:D 30 2 2.64 0.0892 .
## Residuals 160 28
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(cd.aov.02,type="III")
```

```
## Anova Table (Type III tests)
##
## Response: y
## Sum Sq Df F value Pr(>F)
## (Intercept) 4832 1 843.25 < 2e-16 ***
## D 171 2 14.89 3.9e-05 ***
## C 52 1 9.11 0.0054 **
## D:C 30 2 2.64 0.0892 .
## Residuals 160 28
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```