Statistics Lab #9
One-way Within-Subjects ANOVA

9 One-way Within-Subjects ANOVA

Section 9.1 reviews the basic commands you need to perform a one-way, within-subject ANOVA and to evaluate a linear trend/contrast. Lab questions are listed in Section 9.2.

9.1 Review of example from Class Notes

In this section we will work through the example given in the course notes for Chapter 11. First, 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"))
library(car)
mw115 <- read.csv(file=url("http://psycserv.mcmaster.ca/bennett/psy710/datasets/mw115.csv"))
```

The last command reads a csv file that contains the data from Table 11.5 in your textbook, and stores everything in the data frame mw115. The data are from a fictitious experiment that measured cognitive ability in 12 children at 30, 36, 42, and 48 months of age.

The within-subjects ANOVA consists of the following steps:

1. The dependent variables are extracted from the data frame and stored in a matrix;
2. The lm function is used to create a multivariate linear model (i.e., mlm) object that specifies the between-subjects aspect of the experimental design;
3. The Anova function in the car package is used to i) specify the within-subjects aspects of the design; and ii) convert the mlm object to an aov object;
4. Finally, the summary function prints the ANOVA table, Mauchly Test, and the corrected p values for the within-subject variables.

9.1.1 extracting the data

Note the wide format of the data frame. Each row contains all of the data from a single subject, and each column contains all of the data in a single condition.

```r
mw115[1:4,] # print 1st 4 rows
## subj age.30 age.36 age.42 age.48
## 1 s1 108 96 110 122
## 2 s2 103 117 127 133
## 3 s3 96 107 106 107
## 4 s4 84 85 92 99
```
To do the analysis, we first extract the dependent variables and create a matrix. Note how a data frame and a matrix are different classes of variables:

```r
(y.mat <- as.matrix(mw115[1:12,2:5]) )
## age.30 age.36 age.42 age.48
## 1 108 96 110 122
## 2 103 117 127 133
## 3 96 107 106 107
## 4 84 85 92 99
## 5 118 125 125 116
## 6 110 107 96 91
## 7 129 128 123 128
## 8 90 84 101 113
## 9 84 104 100 88
## 10 105 114 105 112
## 11 113 117 132 130
class(y.mat)
## [1] "matrix"
class(mw115)
## [1] "data.frame"
```

We can calculate summary statistics for our dependent variables using the `apply` function:

```r
which.dim <- 2 # 1=rows; 2=columns
apply(y.mat,2,mean) # column means
## age.30 age.36 age.42 age.48
## 103 107 110 112
apply(y.mat,2,sd) # column standard dev
## age.30 age.36 age.42 age.48
apply(y.mat,2,length)
## age.30 age.36 age.42 age.48
## 12 12 12 12
# summary(y.mat) # try this command
```

### 9.1.2 multivariate linear models

Next, we use `lm` to create a *multivariate* linear model:
Bennett, PJ

Lab #9

```r
mw115.mlm <- lm(y.mat ~ 1) # multivariate linear model
```

Note that the model contains multiple dependent variables and only an intercept (1) on the side of the formula that contains the independent (or predictor) variables. If our experiment had contained between-subject factors, they would have been included, too. However, there are no between-subject factors and so that side of the formula contains only the intercept.

### 9.1.3 specifying the within-subjects part of the design

Next, we construct a data frame that specifies the factors used in the within-subject part of the design. In this case there is only one factor, `age`. The following code creates the factor and stores it in a data frame:

```r
(age <- factor(x=c("a30","a36","a42","a48")))
## [1] a30 a36 a42 a48
## Levels: a30 a36 a42 a48

(mw115.idata <- data.frame(age))
## age
## 1 a30
## 2 a36
## 3 a42
## 4 a48
```

It is important to see that the order of the levels of `age` (from top to bottom) match the order of levels in `y.mat` (from left to right). In other words, the factor `age` can be used as labels for each column in the dependent variable matrix.

Finally, we use the `Anova` command in the `car` package to construct an `aov` object.

```r
library(car)

mw115.aov <- Anova(mw115.mlm, idata=mw115.idata, idesign=~age, type="III")
```

In the `Anova` command, the `idata` parameter is the data frame that contains all of the within-subject factors in our experiment. The `idesign` parameter is a one-sided formula that specifies how these factors are used in the within-subject model. In this case, the model contains only one factor, `age`, and therefore the model is very simple. Finally, Type III sums of squares are specified only to suppress a warning message produce by `Anova` when type is not specified: it does not make a difference in this case because the design is balanced.

### 9.1.4 printing the ANOVA table

The `summary` function prints the ANOVA table. The default is to print various multivariate statistics; setting `multivariate` to FALSE suppresses that behaviour:

```r
summary(mw115.aov, multivariate=FALSE)
```

```r
##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
```
The print out includes the Geisser-Greenhouse and Huynh-Feldt adjustments, as well as the unadjusted $F$ test. The Mauchly Test of sphericity can be used to see if the deviations from sphericity are significant. If the test is not significant, then the unadjusted $F$ test may be appropriate. Some have argued that the Mauchly Test lacks power, and so you might want to use a liberal criterion (e.g., $p < 0.1$ or $p < 0.2$) for rejecting the null hypothesis (of having spherical data).

### 9.1.5 Geisser-Greenhouse conservative $F$ test

You can calculate a $p$ value using the lower-bound adjustment with the `pf` command. The numerator and denominator degrees of freedom are, respectively, 1 and one less than the number of subjects.

```r
dim(y.mat)
## [1] 12 4
n <- 12;
F <- 3.027  # from ANOVA
1-pf(F,df1=1,df2=n-1)  # not significant, $F(1,11)=3.027$, $p=0.11$
## [1] 0.10975
```

### 9.1.6 linear contrasts/trends

The adjusted $F$ tests are not significant. However, the omnibus test lacks power and specificity. We probably are not interested in any difference among our dependent measures. Instead, we probably would be looking for a linear and/or quadratic changes, or trends, in the dependent variable across age. The `contr.poly` command
can be used to get the weights that we need for our trend analysis: \( n \) specifies the number of levels in the factor and \texttt{scores} specifies the values of the levels. Note that \texttt{scores} is necessary only when the levels are not spaced evenly, however it does not hurt to use them here:

\[
\text{theWeights} \leftarrow \text{contr.poly}(n=4, \text{scores}=c(30, 36, 42, 48))
\]

\[
\begin{align*}
\text{lin.weights} & \leftarrow \text{theWeights[,1]} \quad \# \text{ save 1st column} \\
\text{quad.weights} & \leftarrow -1* \text{theWeights[,2]} \quad \# \text{ save 2nd column}
\end{align*}
\]

Notice that I multiplied the quadratic weights by -1. Why? Because I wanted the weights to match the direction of the trend that I think I will find in my data (i.e., increasing and then decreasing).

Next, we use the trend weights to convert the four measures for each subject into a single composite score: each measure is multiplied by the corresponding weight, and the four products are summed to create a single score:

\[
\begin{align*}
\text{lin.scores} & \leftarrow \text{y.mat} \%\% \text{lin.weights} \\
\text{quad.scores} & \leftarrow \text{y.mat} \%\% \text{quad.weights}
\end{align*}
\]

Finally, for each trend we use a \texttt{t} tests to evaluate the null hypothesis that the trend is zero. The linear trend is significantly different from zero, but the quadratic trend is not:

\[
\begin{align*}
\text{t.test(lin.scores)} \\
\text{t.test(quad.scores)}
\end{align*}
\]
## 95 percent confidence interval:
## -3.7081  5.7081
## sample estimates:
## mean of x
##  1

### 9.2  Lab Questions

#### 9.2.1  ANOVA on trackbox data

The data for this section can be loaded into R with the following command:

```r
load(url("http://psycserv.mcmaster.ca/bennett/psy710/datasets/trackbox09.Rdata"))
```

“Discovery Day” is a day set aside by the United States Naval Postgraduate School in Monterey, California, to invite the general public into its laboratories. On Discovery Day, 21 October 1995, data on reaction time and hand-eye coordination were collected on 118 members of the public who visited the Human Systems Integration Laboratory. The age and sex of each subject were also recorded. Visitors were mostly in family groups. A rotary pursuit tracking experiment was done to examine motor learning and hand-eye coordination. The equipment was a rotating disk with a 3/4” target spot. In the “Circle” condition, the target spot moved at a constant speed in a circular path. In the “Box” condition, the target spot moved at various speeds as it moved along a box-shaped path. The subject’s task was to maintain contact with the target spot with a metal wand. Four trials were recorded for each of 108 subjects. Each trial lasted 15 s, and the total contact duration during each trial was recorded. The data from the Box condition (n=70) are stored in `track.data`. The variables are `sex`, `age`, `subject`, `time1`, `time2`, `time3`, `time4`. The dependent variable, the amount of time the subject maintained contact with the target on each trial, is in the variables `time1-4`.

In the following analyses, we will ignore `sex` and `age`.

1. Extract the four dependent variables (i.e., contact duration on each trial) and store the result in a matrix.
2. Calculate the mean and standard deviation of contact time on each trial.
3. Use a within-subjects (repeated-measures) ANOVA to determine if contact time differed across trials.
   
   (a) Adjust the $p$ value for the effect of trial with the conservative, or lower-bound, F test.
   
   (b) Adjust the $p$ value for the effect of trial with the Huynh-Feldt corrections for sphericity.
4. Use a linear contrast to test for a positive (i.e., increasing) linear trend in contact time across trials.

#### 9.2.2  chick weight data

In this section we will conduct use ANOVA and trend analysis to analyze a subset of R’s `ChickWeight` data set which contains body weights of four groups of chicks that were measured at birth and every second day until day 20. The data frame `chicks` contains weights measured from days 0 though 10 from one group of chicks:

```r
load(file=url("http://psycserv.mcmaster.ca/bennett/psy710/datasets/chicks.Rdata") )
closeAllConnections()
```

The data frame `chicks` contains a subset of the data in R’s `ChickWeight` data set, which contains body weights of four groups of chicks that were measured at birth and every second day until day 20. The data frame `chicks`, which contains weights measured from days 0 though 10 from one group of chicks, can be loaded with the following command:
1. Without looking at the data, do you think it is likely that the sphericity assumption is valid for this design? Explain.

2. Create weights that can be used to test for linear and quadratic trends of weight across days. Verify that the two sets of weights are orthogonal.

3. Evaluate the null hypothesis that the linear trend of weight across days is $\leq 0$.

4. Use the weights to evaluate the hypothesis that the quadratic trend of weight across days is zero.