Introduction to R and RStudio

Part 4: Extended Inferential Statistics in R

Rob Cribbie
Department of Psychology
York University

http://www.psych.yorku.ca/cribbie/r_course_trent.html

One-way Independent Groups ANOVA

- Hypothesis #6: Is there a difference between the three treatment conditions on posttest-perfectionism?
  - Option 1:
    - > mod4<- lm(perf3 ~ group, data=dat)
    - > anova (mod4)
  - Option 2:
    - > mod4<-aov(perf3 ~ group, data=dat)
    - > summary(mod4)
  - Option 3:
    - >oneway.test(perf3 ~ group, var.equal=TRUE, data=dat)

One–way Independent Groups ANOVA

```
> oneway.test(perf3 ~ group, var.equal=TRUE,data=dat)

 One-way analysis of means

data:  perf3 and group
F = 0.2913, num df = 2, denom df = 87, p-value =
0.748

> mod1<- lm(perf3 ~ group, data=dat)
> anova (mod1)
Analysis of Variance Table

Response: perf3
          Df  Sum Sq Mean Sq F value Pr(>F)
group      2   131.2  65.575  0.2913  0.748
Residuals 87 19586.0 225.126
```

Multiple Comparisons for One-way Independent Groups ANOVA

- Tukey's Honestly Significant Difference (HSD) Familywise Error Controlling Procedure for Pairwise Comparisons

```
> mod2<-aov(perf3 ~ group, data=dat)
> TukeyHSD(mod2)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = perf3 ~ group, data = dat)

$group
                     diff        lwr      upr     p adj
control-cbt     0.1048851  -9.132747 9.342518 0.9995960
stress-cbt     -2.5067089 -11.744341 6.730924 0.7945257
stress-control -2.6115940 -11.849227 6.626039 0.7791236
```

Multiple Comparisons for One-way Independent Groups ANOVA

- Flexible procedure for all pairwise comparisons

```
> pairwise.t.test(perf3, group, p.adj="none")

 Pairwise comparisons using t tests with pooled
SD

data:  perf3 and group

        cbt  control
control 0.98 -
stress  0.52 0.50

P value adjustment method: none
```

This option can be changed to any post hoc test you prefer, e.g., ‘bonf’, ‘holm’, ‘fdr’

How can we check for assumption violation?

- Variance Homogeneity Assumption

```
> library(car)
> leveneTest(dat$perf3,dat$group)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.5558 0.5756
      87
```

Note that by default it uses the median, rather than the mean, to compute deviations

How can we check for assumption violation?

- Normality Assumption

```
> tapply(dat$perf3,dat$group,shapiro.test)
$cbt

 Shapiro-Wilk normality test

data:  X[[1L]]
W = 0.9445, p-value = 0.1203


$control

 Shapiro-Wilk normality test

data:  X[[2L]]
W = 0.9631, p-value = 0.3712


$stress

 Shapiro-Wilk normality test

data:  X[[3L]]
W = 0.9281, p-value = 0.04379
```

Plots are better, but I just wanted to show a different method that can be used along with plots

One-way Independent Groups ANOVA under Variance Inequality

- Welch's Independent Groups ANOVA
  - As with the t.test function, the default for the oneway.test function is to use Welch's heteroscedastic ANOVA

```
> oneway.test (perf3 ~ group)

 One-way analysis of means (not assuming
 equal variances)

data:  perf3 and group
F = 1.8752, num df = 2.000, denom df =
55.448, p-value = 0.1629
```

Good hint that the Welch test is being reported

Multiple Comparisons for Welch's Independent Groups ANOVA

- Multiplicity control with pairwise.t.test

```
> pairwise.t.test(perf3, group, p.adjust.method = "holm",
pool.sd=FALSE, data=dat)

 Pairwise comparisons using t tests with non-pooled SD

data:  perf3 and group

        cbt  control
control 0.57 -
stress  0.20 0.56

P value adjustment method: holm
```

pool.sd = FALSE indicates that you would like to use Welch's t-test for conducting the analyses

One-way Independent Groups ANOVA under Nonnormality

- Kruskal-Wallis Nonparametric Test

```
> kruskal.test (perf3 ~ group, data=dat)

 Kruskal-Wallis rank sum test

data:  perf3 by group
Kruskal-Wallis chi-squared = 4.5791, df = 2,
p-value = 0.1013
```

One-way Independent Groups ANOVA under Nonnormality and Variance Heterogeneity

- As in the two independent groups situation, we can use one of Rand Wilcox's functions (in this case *t1way*) for computing a Welch omnibus test on trimmed means
  - This test is much more reliable than a standard one-way ANOVA when the normality and variance homogeneity assumptions are violated

One-way Independent Groups ANOVA under Nonnormality and Variance Heterogeneity

```
> library(WRS2)
> t1way(perf3 ~ group, data=dat, tr=.2)
Call:
t1way(formula = perf3 ~ group, data = dat, tr = 0.2)

Test statistic: 2.7615
Degrees of Freedom 1: 2
Degrees of Freedom 2: 33.18
p-value: 0.07774
```

One-way Repeated Measures ANOVA

- Hypothesis #7: Is there a significant difference in perfectionism scores from pretest to one-month to posttest?
  - Problem: Simple methods for conducting repeated measures ANOVAs ignore the important sphericity assumption that is regularly violated with repeated measures data and inflates Type I error rates
  - Example:
    - mod5<- aov(perf ~ week + error (subject / week))
  - However, other functions are available in R that use adjusted df or multivariate solutions to solve the sphericity issue

One-way Repeated Measures ANOVA with the "car" package

This is the old method which lost popularity with newer functions and the emergence of mixed-models for repeated measures

- library(car)
- time<-c(1,2,3)
- time<-as.factor(time)
- idat<-data.frame(time)
- mod6<-lm(cbind(perf1,perf2,perf3)~1)
- aov1<-Anova(mod6, idata=idat, idesign=~time)
- summary(aov2)
  - Multivariate Tests: time

	Df	test stat	approx F	numDf	denDf	Pr(>F)
Pillai	1	0.290019	17.973521	2	88	2.85e-07 ***
Wilks	1	0.709981	17.973521	2	88	2.85e-07 ***
Roy	1	0.408489	17.973521	2	88	2.85e-07 ***

  - Greenhouse-Geisser Correction for Departure from Sphericity

	GG eps	Pr(>F[GG])
time	0.68104	1.728e-07 ***

One-way Repeated Measures ANOVA with the "ez" package

- As the name implies, the *ez* package makes repeated measures ANOVA easier
  - However, one catch is that the data must be in long-form rather than wide-form
  - To do this we can use the 'reshape' function

Columns in the data set that specify the repeated measures

```
> head(dat,n=2)
  sex group     dep1    perf1    perf2    perf3
1   m   cbt 87.59479 68.70716 62.76111 78.43853
2   m   cbt 96.39773 87.44450 75.06312 76.95246
> longdata<-reshape(dat,direction="long",varying=4:6,sep="")
> head(longdata,n=2)
    sex group     dep1 time     perf id
1.1   m   cbt 87.59479    1 68.70716  1
2.1   m   cbt 96.39773    1 87.44450  2
```

A new variable that represents the levels of the perfectionism variable

One-way Repeated Measures ANOVA with the "ez" package

Within Subject Variable

```
> longdata$time<-factor(longdata$time)
> library(ez)
> ezANOVA(data=longdata,dv=perf,wid=id,within=time)
Warning: Converting "id" to factor for ANOVA.
$ANOVA
  Effect DFn DFd        F            p p<.05       ges
2   time   2 178 40.02523 4.422495e-15     * 0.07638647

$`Mauchly's Test for Sphericity`
  Effect         W           p p<.05
2   time 0.6921641 9.31536e-08     *

$`Sphericity Corrections`
  Effect       GGe       p[GG] p[GG]<.05       HFe
2   time 0.7646219 4.239785e-12         * 0.7752988
         p[HF] p[HF]<.05
2 3.103753e-12         *
```

ID variable (automatically assigned by 'reshape')

Factorial Independent Groups ANOVA

- Hypothesis 8: Is there a significant relationship between posttest perfectionism scores and the predictors group and sex?
- Factorial ANOVA is computed using the linear model (lm) function, along with a function for computing the anova summary table
  - anova
    - Function in R for computing, by default, Type I SS
  - Anova
    - Function in R for computing, by default, Type II SS

Factorial Independent Groups ANOVA with an Interaction

With no interaction term

```
> anova(lm(perf3 ~ group + sex, data=dat))
Analysis of Variance Table

Response: perf3
          Df  Sum Sq Mean Sq F value Pr(>F)
group      2   138.7  69.327  0.3359 0.7156
sex        1     2.2   2.220  0.0108 0.9176
Residuals 86 17747.1 206.362
> anova(lm(perf3 ~ group*sex, data=dat))
Analysis of Variance Table

Response: perf3
          Df  Sum Sq Mean Sq F value Pr(>F)
group      2   138.7  69.327  0.3374 0.7146
sex        1     2.2   2.220  0.0108 0.9175
group:sex  2   485.3 242.652  1.1808 0.3121
Residuals 84 17261.8 205.498
```

Factorial Independent Groups ANOVA – Specifying the Model

```
> anova(lm(perf3 ~ group*sex, data=dat))
Analysis of Variance Table

Response: perf3
          Df  Sum Sq Mean Sq F value Pr(>F)
group      2   138.7  69.327  0.3374 0.7146
sex        1     2.2   2.220  0.0108 0.9175
group:sex  2   485.3 242.652  1.1808 0.3121
Residuals 84 17261.8 205.498
> anova(lm(perf3 ~ group + sex + group:sex, data=dat))
Analysis of Variance Table

Response: perf3
          Df  Sum Sq Mean Sq F value Pr(>F)
group      2   138.7  69.327  0.3374 0.7146
sex        1     2.2   2.220  0.0108 0.9175
group:sex  2   485.3 242.652  1.1808 0.3121
Residuals 84 17261.8 205.498
```

These are equivalent specifications of the model

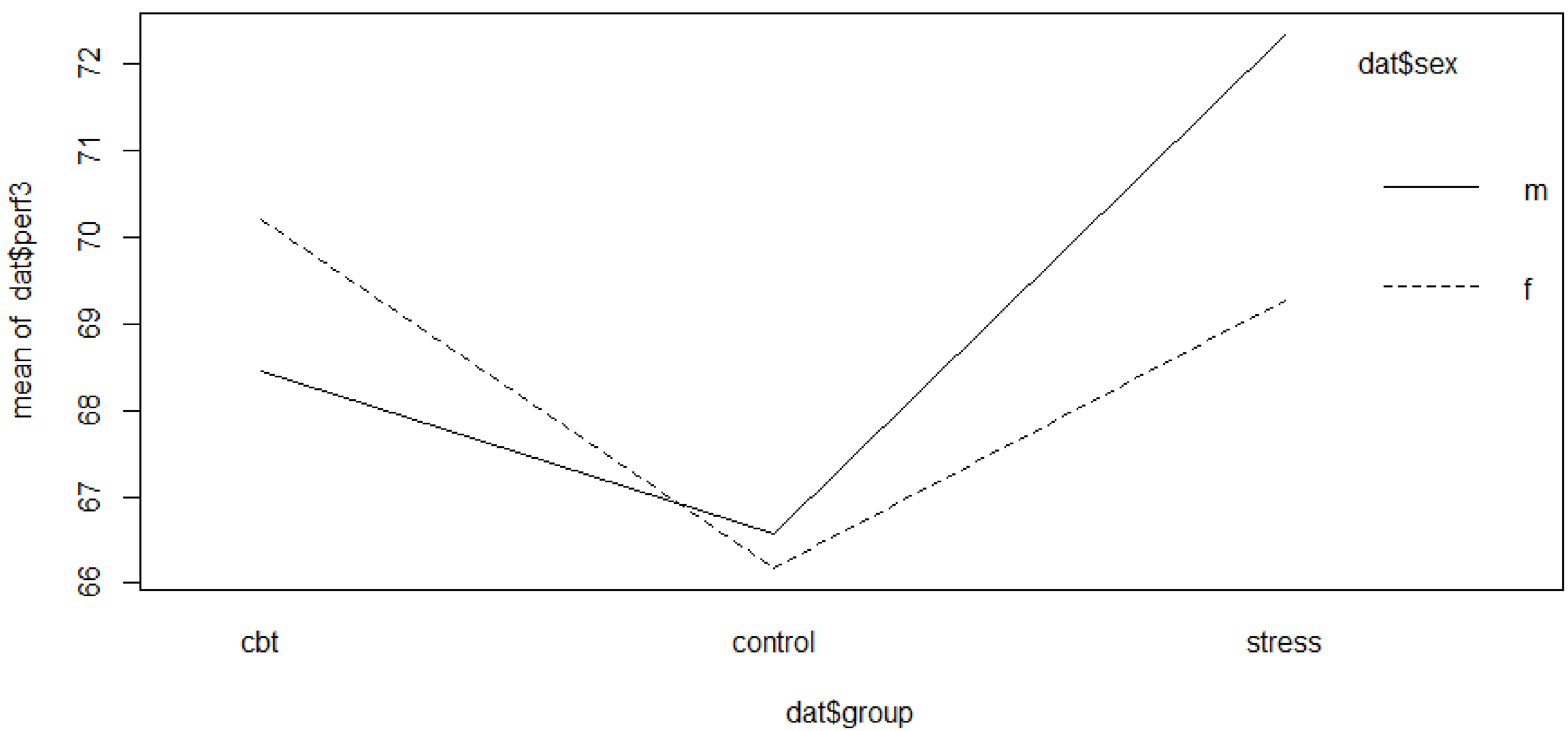
Factorial Independent Groups ANOVA with Type II SS

```
> library(car)
> Anova(lm(perf3 ~ group*sex, data=dat))
Anova Table (Type II tests)

Response: perf3
           Sum Sq Df F value Pr(>F)
group       136.2  2  0.3314 0.7188
sex           2.2  1  0.0108 0.9175
group:sex   485.3  2  1.1808 0.3121
Residuals 17261.8 84
```

Factorial ANOVA: Plotting a Potential Interaction

- >interaction.plot(dat$group, dat$sex, dat$perf3)



Mixed ANOVA

- Hypothesis 9: Are perfectionism scores affected by time, group, or the interaction of time & group?

- We will again use the *ez* package since it makes computing repeated measures analyses very straightforward
- The only difference is that we will add a between subject variable

Mixed ANOVA

```
> ezANOVA(data=longdata,dv=perf,wid=id,within=time,between=group)
Warning: Converting "id" to factor for ANOVA.
$ANOVA
      Effect DFn DFd           F            p
2      group   2  87  0.9253750 4.002485e-01
3       time   2 174 79.0046855 3.869946e-25
4 group:time   4 174  0.1125054 9.780010e-01
  p<.05         ges
2       0.017189378
3     * 0.139034787
4       0.000459715

$`Mauchly's Test for Sphericity`
      Effect         W            p p<.05
3       time 0.7135874 4.991465e-07     *
4 group:time 0.7135874 4.991465e-07     *

$`Sphericity Corrections`
      Effect       GGe        p[GG] p[GG]<.05
3       time 0.7773556 4.068395e-20         *
4 group:time 0.7773556 9.565367e-01
```

Within Subject Variable

Between Subject Variable

Mixed Model for Repeated Measures

- One modern approach to analyze repeated measures is to utilize a hierarchical/mixed-model approach
  - A mixed model approach has the following advantages:
    - No need to assume sphericity
    - Flexible treatment of missing data (uses all available data)
    - Flexible treatment of time
      - Not every individual needs to be measured at the exact same time
- Like the *ezANOVA* function, the data must be in longform

Mixed Model Analysis

Non-linear Mixed Effects package, also conducts linear analyses with the *lme* function

```
> library(nlme)
> mixmod<-lme(perf ~ time, random = ~ 1 | id,data=longdata)
> anova(mixmod)
            numDF denDF  F-value p-value
(Intercept)     1   178 4083.730  <.0001
time            2   178   80.612  <.0001
```

Specifies that ids are random, and links the ids to the repeated measures