Introduction to R

Answers to Session 7 exercises

Statistical Consulting Centre

2 March, 2017

1 *t*-tests

Carry out a two-sample *t*-test to determine whether:

(i) males and females have different mean nerdy scores.

```
with(sports.df, t.test(nerdy.sc~gender))
```

```
Welch Two Sample t-test

data:  nerdy.sc by gender
t = -3.8204, df = 965.27, p-value = 0.0001418
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.18630715 -0.05985835
sample estimates:
mean in group Female   mean in group Male
            2.949468             3.072551
```

(ii) the mean nerdy score of respondents living with their partners differs from that of respondents who do not live with their partners.

```
with(sports.df, t.test(nerdy.sc~partner))
```

```
Welch Two Sample t-test

data:  nerdy.sc by partner
t = -1.0193, df = 211.86, p-value = 0.3092
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.18490690  0.05886231
sample estimates:
 mean in group No mean in group Yes
         2.884848          2.947871
```

2 ANOVA

(i) Perform a one-way anova to test mean nerdy score differs between the three age groups we have been considering.

```
oneway <- with(sports.df, aov(nerdy.sc~age.group))
summary(oneway)

             Df Sum Sq Mean Sq F value Pr(>F)
age.group     2   2.22  1.1081    4.31 0.0137 *
Residuals   986 253.52  0.2571
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
7 observations deleted due to missingness
```

(ii) What are your conclusions from the one-way anova?

At least one age group's mean nerdy score differs from that of the others.

(iii) Find the estimated mean nerdy score over all age groups and for individual age groups.

```
model.tables(oneway, "means")

Tables of means
Grand mean

3.006218

 age.group
   Under 40 41 to 60 Over 61
      3.009    3.059   2.946
rep  293.000  364.000 332.000
```

(iv) Perform pair-wise comparisons of mean nerdy scores between all age groups using Tukey's Honest Significance Difference method to compute p-values adjusted for multiple comparisons.

```
TukeyHSD(oneway)

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = nerdy.sc ~ age.group)

$age.group
                          diff         lwr         upr     p adj
41 to 60-Under 40  0.04925709 -0.04415998  0.14267415 0.4312230
Over 61-Under 40  -0.06360922 -0.15901291  0.03179448 0.2613921
Over 61-41 to 60  -0.11286630 -0.20319233 -0.02254027 0.0096125
```

(v) Which pairs of age groups differ in mean nerdy score?

"41–60" and "Over 61"

(vi) Perform a two-way anova of nerdy score on age group and gender.

```
twoway <- with(sports.df, aov(nerdy.sc~age.group*gender))
summary(twoway)

                  Df Sum Sq Mean Sq F value   Pr(>F)
age.group          2   2.22   1.108   4.384   0.0127 *
gender             1   4.19   4.193  16.588 5.02e-05 ***
age.group:gender   2   0.86   0.428   1.693   0.1845
Residuals        983 248.47   0.253
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
7 observations deleted due to missingness
```

(vii) Which rows of the two-way ANOVA table are statistically significant?

Those corresponding to `age.group` *and* `gender`. *The interaction between* `age.group` *and* `gender` *is not statistically significant at the 5% level since* Pr(>F)>0.05.

(viii) Calculate the estimated means for each `age.group`, `gender` and `age.group-gender` combination. Perform *appropriate* pair-wise comparisons of means.

```
model.tables(twoway, "means")

Tables of means
Grand mean

3.006218

 age.group
    Under 40 41 to 60 Over 61
       3.009    3.059   2.946
rep  293.000  364.000 332.000

 gender
     Female    Male
      2.946   3.076
rep 533.000 456.000

 age.group:gender
           gender
age.group  Female Male
  Under 40   2.99   3.03
  rep      171.00 122.00
  41 to 60   2.98   3.16
  rep      204.00 160.00
```

```
  Over 61    2.86   3.02
  rep      158.00 174.00
```

```
TukeyHSD(twoway)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = nerdy.sc ~ age.group * gender)

$age.group
                          diff         lwr         upr     p adj
41 to 60-Under 40  0.04925709 -0.04336658  0.14188075 0.4250499
Over 61-Under 40  -0.06360922 -0.15820264  0.03098421 0.2554626
Over 61-41 to 60  -0.11286630 -0.20242518 -0.02330742 0.0088880

$gender
                 diff        lwr       upr   p adj
Male-Female 0.1300681 0.06713236 0.1930038 5.4e-05

$`age.group:gender`
                                        diff          lwr        upr
41 to 60:Female-Under 40:Female -0.01193670 -0.160777938 0.13690453
Over 61:Female-Under 40:Female  -0.13137908 -0.289792553 0.02703439
Under 40:Male-Under 40:Female    0.03953121 -0.130597111 0.20965952
41 to 60:Male-Under 40:Female    0.16472588  0.006827846 0.32262391
Over 61:Male-Under 40:Female     0.02933555 -0.125245990 0.18391709
Over 61:Female-41 to 60:Female  -0.11944238 -0.271578176 0.03269342
Under 40:Male-41 to 60:Female    0.05146791 -0.112830956 0.21576678
41 to 60:Male-41 to 60:Female    0.17666258  0.025063566 0.32826160
Over 61:Male-41 to 60:Female     0.04127226 -0.106869319 0.18941383
Under 40:Male-Over 61:Female     0.17091029 -0.002107728 0.34392830
41 to 60:Male-Over 61:Female     0.29610496  0.135097575 0.45711234
Over 61:Male-Over 61:Female      0.16071463  0.002958359 0.31847091
41 to 60:Male-Under 40:Male      0.12519467 -0.047351533 0.29774088
Over 61:Male-Under 40:Male      -0.01019565 -0.179712194 0.15932089
Over 61:Male-41 to 60:Male      -0.13539033 -0.292629003 0.02184835
                                    p adj
41 to 60:Female-Under 40:Female 0.9999140
Over 61:Female-Under 40:Female  0.1685811
Under 40:Male-Under 40:Female   0.9858132
41 to 60:Male-Under 40:Female   0.0350612
Over 61:Male-Under 40:Female    0.9944170
Over 61:Female-41 to 60:Female  0.2196437
Under 40:Male-41 to 60:Female   0.9478634
41 to 60:Male-41 to 60:Female   0.0116853
Over 61:Male-41 to 60:Female    0.9683015
Under 40:Male-Over 61:Female    0.0550741
41 to 60:Male-Over 61:Female    0.0000028
Over 61:Male-Over 61:Female     0.0429638
```

```
41 to 60:Male-Under 40:Male     0.3030179
Over 61:Male-Under 40:Male      0.9999793
Over 61:Male-41 to 60:Male      0.1376475
```

3 Tests of Independence

(i) Produce a two-way frequency table of counts between `income` and `gender`.

```
with(sports.df, table(income, gender))
```

```
                  gender
income             Female Male
  5 000$               59   27
  10 000$-15 000$      61   59
  15 000$-20 000$      57   41
  20 000$-25 000$      38   38
  25 000$-30 000$      40   33
  30 000$-40 000$      67   63
  40 000$-50 000$      55   51
  50 000$-70 000$      66   62
  70 000$-100 000$     31   34
  > 100 000$           30   33
```

(ii) Do you think that income level depends on gender? Perform an appropriate test to find out.

```
chisq.test(with(sports.df, table(income, gender)))
```

```
Pearson's Chi-squared test

data:  with(sports.df, table(income, gender))
X-squared = 11.756, df = 9, p-value = 0.2274
```