STAT 581 - Problem Set 4

Alex Towell (atowell@siue.edu)

Problem 1

The insulating life of protective fluids at an accelerated load is being studied. The experiment has been performed for four types of fluids, with $n = 5$ trials per fluid type. Suppose fluid types 1 and 2 are from manufacturer A, and that fluid types 3 and 4 are from manufacturer B. The data is available on Blackboard as an Excel File.

(a)

(i)

Perform a test of the global null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. Compute the F_0 statistic, and the p-value.

This is a one-factor experiment with $a = 4$ levels of the factor (fluid level) and $n = 6$ replicates for a total of $N = an = 24$ observations.

The test statistic

$$F_0 = \frac{\text{MS}_{\text{tr}}}{\text{MS}_{\text{E}}}$$

under the null hypothesis $\mu_1 = ... = \mu_a$ (or $\tau_1 = ... = \tau_a = 0$ in the effects model) has a reference distribution

$$F_0 \sim F(a - 1, N - a)$$

where $N = na$.

We compute the observed test statistic and its p-value with:

```
library("readxl")
h4.data = read_excel("handout2data.xlsx")
fluid = as.factor(na.omit(h4.data$fluid))
life = na.omit(h4.data$life)
#head(data.frame(fluid=fluid,life=life))

aov.mod = aov(life ~ fluid)
summary(aov.mod)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## fluid        3  30.16   10.05   3.047 0.0525 .
## Residuals   20  65.99    3.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that $F_0 = 3.047$ which has a p-value .0525.

(ii)

Comment on the additional information provided by the p-value, beyond a determination of statistical significance alone.

The p-value quantifies a measure of evidence beyond a determination of statistical significance.

Earlier, we obtained a p-value of .0525. If we were to perform a dichotomous hypothesis test which specifies a significance level $\alpha = 0.05$, we would decide H_0. However, had we obtained a p-value of .049, while the strength of the evidence is nearly the same, we would decide H_A. Presenting the p-value provides more information.

(b)

Consider the orthogonal contrasts

$$\begin{aligned}\Gamma_1 &= \mu_1 - \mu_2 \\ \Gamma_2 &= \mu_3 - \mu_4 \\ \Gamma_3 &= (\mu_1 + \mu_2) - (\mu_3 + \mu_4).\end{aligned}$$

Preliminary analysis

A *constrast* Γ is a linear combination of parameters

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i$$

such that $c_1 + \cdots + c_a = 0$. Assuming a balanced design,

$$\mathrm{V}(\sum_{i=1}^{a} c_i \bar{y}_{i\cdot}) = \frac{\sigma^2}{n} \sum_{i=1}^{a} c_i^2$$

and thus

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i\cdot}}{\left(\mathrm{MS_E} / n \sum_{i=1}^{a} c_i\right)^{1/2}} \sim t_{N-a}$$

under H_0. The F-test may also be used, where $F_0 = t_0^2 = \frac{\mathrm{MS_C}}{\mathrm{MS_E}} = \frac{\mathrm{SS_C}/1}{\mathrm{MS_E}}$ where

$$\mathrm{SS_C} = \frac{(\sum_{i=1}^{a} c_i \bar{y}_{i\cdot})^2}{1/n \sum_{i=1}^{a} c_i^2} \sim t_{N-a}.$$

We have $a = 4$ fluid types, 1, 2, 3, and 4. Fluid types 1 and 2 are from manufacturer A and fluid types 3 and 4 are from manufacturer B.

Γ_1 compares the average effect (on lifetime) of fluid 1 with the average effect of fluid 2, Γ_2 compares the average effect of fluid 3 with the average effect of fluid 4, and Γ_3 compares the average effect of manufacturer A with the average effect of manufacturer B.

(i)

Compute SSC for each contrast. Describe a general property for the sums of squares of orthogonal contrasts. Why is this property desirable?

Orthogonal contrasts

Orthogonality is desirable since the treatment sum of squares can be decomposed into specific effects.

Additional remarks:

Two contrasts (assuming a balanced design) with coefficients $\{c_i\}$ and $\{d_i\}$ are orthogonal if $\sum_{i=1}^{a} c_i d_i = 0$.

For a factor levels (or treatments), a set of $(a-1)$ orthogonal contrasts $\Gamma_1, \ldots, \Gamma_{a-1}$ are independent with $\mathrm{df} = 1$ and thus tests performed on them are independent.

We see that Γ_1, Γ_2, and Γ_3 are orthogonal contrasts.

Computing $\mathrm{SS_C}$

```
library("multcomp")
contrasts(fluid) = cbind(c(1,-1, 0, 0),
                         c(0, 0, 1,-1),
                         c(1, 1,-1,-1))

# we will re-fit the ANOVA model, now with our own contrasts defined as above
contr.mod = aov(life~fluid)

# we can get a decomposition of sum squares into specific effects
# the command split is used to specify the decompositon.
summary(contr.mod,
        split=list(fluid=list("gamma.1"=1,"gamma.2"=2,"gamma.3"=3)))
```

```
##                  Df Sum Sq Mean Sq F value Pr(>F)
## fluid             3  30.16   10.05   3.047 0.0525 .
##   fluid: gamma.1  1   1.47    1.47   0.445 0.5121
##   fluid: gamma.2  1  13.65   13.65   4.138 0.0554 .
##   fluid: gamma.3  1  15.04   15.04   4.559 0.0453 *
## Residuals        20  65.99    3.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that $\mathrm{SS}_{C_1} = 1.47$, $\mathrm{SS}_{C_2} = 13.65$, and $\mathrm{SS}_{C_3} = 15.04$. By orthogonality, $\mathrm{SS}_{\mathrm{tr}} = \mathrm{SS}_{C_1} + \mathrm{SS}_{C_2} + \mathrm{SS}_{C_3} = 30.16$.

(ii)

Perform a test of $H_0 : \gamma = 0$ for each of the contrasts. Compute the F_0 statistics, and the p-values.

null hypothesis	F_0	p-value
$\Gamma_1 = 0$	.445	.512
$\Gamma_2 = 0$	4.138	.0554
$\Gamma_3 = 0$	4.559	.0453

(iii)

Provide an interpretation, stated in the context of the problem. Again, note the additional information provided by the p-value, beyond a determination of statistical significance.

The data is compatible with $H_0 : \Gamma_1 = 0$. Both $H_0 : \Gamma_2 = 0$ and $H_0 : \Gamma_3$ have p-values near significance level $\alpha = .05$, respectively .0554 and .0453, and thus both have approximately the same strength of evidence for an effect. We do not arbitrarily decide compatibility based on whether the p-value falls on one side of $\alpha = .05$ or the other, but rather use the extra information in the p-value.

We summarize our analysis by saying the experiment finds no evidence of (fluid 1 - fluid 2) effects, some evidence of (fluid 3 - fluid 4) effects, and slightly stronger evidence of manufacturer (A - manufacturer B) effects.

Problem 2

A product developer is investigating the tensile strength of a new synthetic fiber. A completely randomized design with five levels of cotton content is performed, with $n = 5$ specimens per level. The data is available on Blackboard as an Excel File.

(a)

Perform pairwise comparisons using the Fisher LSD method, and the Tukey method. Provide grouping information for each method.

Preliminary computations

```
percent = as.factor(na.omit(h4.data$percent))
strength = na.omit(h4.data$strength)
aov.mod = aov(strength ~ percent)
summary(aov.mod)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## percent      4  475.8  118.94   14.76 9.13e-06 ***
## Residuals   20  161.2    8.06                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mse = 8.06
a = 5
n = 5
df = a*(n-1)
# investigate some pairwise multiple comparisons.
# remember that Tukey is used to define pairwise comparisons, as the Tukey
# method is best suited for such comparions
comparisons.mod = glht(aov.mod, linfct=mcp(percent="Tukey"))
```

Fisher LSD method

Fisher LSD tests, error probability controlled for each comparison:

```
summary(comparisons.mod, test=univariate())
```

```
## 
##   Simultaneous Tests for General Linear Hypotheses
## 
## Multiple Comparisons of Means: Tukey Contrasts
## 
## 
## Fit: aov(formula = strength ~ percent)
## 
## Linear Hypotheses:
##               Estimate Std. Error t value Pr(>|t|)    
## 20 - 15 == 0     5.600      1.796   3.119 0.005409 ** 
## 25 - 15 == 0     7.800      1.796   4.344 0.000315 ***
## 30 - 15 == 0    11.800      1.796   6.572 2.11e-06 ***
## 35 - 15 == 0     1.000      1.796   0.557 0.583753    
## 25 - 20 == 0     2.200      1.796   1.225 0.234715    
## 30 - 20 == 0     6.200      1.796   3.453 0.002514 ** 
## 35 - 20 == 0    -4.600      1.796  -2.562 0.018595 *  
## 30 - 25 == 0     4.000      1.796   2.228 0.037541 *  
## 35 - 25 == 0    -6.800      1.796  -3.787 0.001157 ** 
## 35 - 30 == 0   -10.800      1.796  -6.015 7.01e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## (Univariate p values reported)
```

Tukey method

Tukey pairwise tests, error probability controlled across all comparisons:

```r
summary(comparisons.mod)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = strength ~ percent)
##
## Linear Hypotheses:
##              Estimate Std. Error t value Pr(>|t|)
## 20 - 15 == 0    5.600      1.796   3.119  0.03845 *
## 25 - 15 == 0    7.800      1.796   4.344  0.00254 **
## 30 - 15 == 0   11.800      1.796   6.572  < 0.001 ***
## 35 - 15 == 0    1.000      1.796   0.557  0.97977
## 25 - 20 == 0    2.200      1.796   1.225  0.73722
## 30 - 20 == 0    6.200      1.796   3.453  0.01900 *
## 35 - 20 == 0   -4.600      1.796  -2.562  0.11635
## 30 - 25 == 0    4.000      1.796   2.228  0.21009
## 35 - 25 == 0   -6.800      1.796  -3.787  0.00904 **
## 35 - 30 == 0  -10.800      1.796  -6.015  < 0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Grouping information: Fisher method

We summarize the results for Fisher with a compact letter display:

```r
cld(summary(comparisons.mod,test=univariate()))
```

```
##  15  20  25  30  35
## "a" "b" "b" "c" "a"
```

According to the Fisher method, the experiment finds that 30% cotton has the greatest effect on tensile strength, 25% and 20% have the second greatest effect (but no difference between them), and 15% and 35% have the weakest effect (but no difference between them).

Grouping information: Tukey method

We summarize the results for Tukey with a compact letter display:

```r
cld(summary(comparisons.mod))
```

```
##   15   20   25   30   35
##  "a" "bc" "cd"  "d" "ab"
```

Turkey groupings:

Note that the column headers 1, 2, 3, 4 denote the strength of the effect, 1 being weakest and 4 being strongest.

Cotton	1	2	3	4
15%	A			
20%		B	C	
25%			C	D
30%				D
35%	A	B		

In this case, there is no single way to order them as we could with the Fisher grouping, since the sets are not disjoint. Therefore, I tend to prefer the above table to show the groupings where strength of effect is in ascending order of group identifiers A, B, C, and D.

We may then refer to the p-values for the groupings to shed more insight on specific comparisons, e.g., $H_0 : \mu_1 = \mu_5$ has a p-value of .98, and thus we say that data is *extremely* compatible with the hypothesis. Additionally, we can say that 15% and 35% cotton levels have the weakest effect on strength.

Similar claims may e constructed for other comparisons.

(b)

(i)

Describe the defining characteristics for each of the above pairwise comparison methods.

Fisher LSD controls the probability of a type I error for each pairwise comparison.

Tukey HSD controls the overall probability of type I error across all pairwise comparisons, which is sometimes denoted the experimentation type I error,

$$\Pr\{\text{decide } H_A^{(i,j)} \text{ for some pair } (i,j) | \mu_1 = \cdots = \mu_5\} = \alpha.$$

Remark: As the number of factors goes up, the type II error goes up.

(ii)

Compute the margin of error and the comparison-wise error rate for the Tukey method in this problem.

```
# the least significant differences (margin of errors) for Tukey
m.tukey = qtukey(.05,a,df,lower.tail=F)*sqrt(mse/n)
m.tukey
```

```
## [1] 5.372958
```

```
# calculation for Tukey comparison error rate
2*pt(qtukey(.05,a,df,lower.tail=F)/sqrt(2),df,lower.tail=F)
```

```
## [1] 0.007198365
```

(iii)

Compute the margin of error and the family-wise error rate for the Fisher LSD method in this problem

```
# the least significant differences (margin of errors) for Fisher
m.lsd = qt(.025,df,lower.tail=F)*sqrt(2*mse/n)
m.lsd
```

```
## [1] 3.745452
```

```r
# calculation for Fisher family error rate
ptukey(qt(.025,df,lower.tail=F)*sqrt(2),a,df,lower.tail=F)
```

```
## [1] 0.2643089
```

(c)

Comment on the seemingly contradictory nature of a pairwise comparisons analysis.

When multiple decisions are made in the presence of uncertainty, a measure of belief/evidence is necessary to avoid contradiction.