

Quiz 3

Adon Rosen

Date: 2020-04-20

Question 1

(29 points) The following data consists of a 3 (age) x 2 (sex) natural design in which the proportion of pretend play between parents and infants changes as a function of age. The DV (Y) is the observed proportion of parent-infant play that consists of pretend play. The first independent variable (X1) is age, where 1 = 7-month old infants; 2 = 10-month old infants; and 3 = 13-month-old infants. The second independent variable (X2) is sex, where 1=girls and 2=boys. There are 8 infants in each cell of the design.

a.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	2	0.2321375	0.1160687	12.1978919	0.0000666
X2	1	0.0005333	0.0005333	0.0560490	0.8140041
X1:X2	2	0.0002042	0.0001021	0.0107281	0.9893319
Residuals	42	0.3996500	0.0095155	NA	NA

b.

The unbiased RMSE is: 0.0975473

The biased RMSE is: 0.0912471

c.

The F value is: 12.1978919

The p value is: 6.6570919×10^{-5}

The models being compared are:

model full: $Y_{playtime} = \beta_{intercept} + \beta_{age} \times X_{i,age} + \epsilon_i$

model restricted: $Y_{playtime} = \beta_{intercept} + \epsilon_i$

d.

The F value is: 0.056049

The p value is: 0.8140041

e.

The F value is: 0.0107281

the p value is: 0.9893319

The models being compared are

model full: $Y_{playtime} = \beta_0 + \beta_{sex} \times X_{i,sex} + \beta_{age} \times X_{i,age} + \beta_{sex \times age} \times X_{i,age \times sex} + \epsilon_i$

model restricted: $Y_{playtime} = \beta_0 + \beta_{sex} \times X_{i,sex} + \beta_{age} \times X_{i,age} + \epsilon_i$

f.

While it is often inappropriate to drive model selection via significance in predictors (for instance, sex may be proven to be a significant predictor in the literature), the data suggest here that the most parsimonious model may be: $Y_{playtime} = \beta_0 + \beta_{age} \times X_{i,age} + \epsilon_i$, which is the model I would suggest given the data.

g.

Results from linear model with linear effect:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.12625	0.0134818	9.364462	0.0e+00
linAge	0.08500	0.0165118	5.147836	5.3e-06

Results from linear model with quadratic age effect:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.126250	0.0169129	7.464731	0.0000000
quadAge	0.003125	0.0119592	0.261305	0.7950237

Results from ANOVA model with linear contrasts:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
linAge	1	0.231200	0.2312000	26.50022	5.3e-06
Residuals	46	0.401325	0.0087245	NA	NA

Results from ANOVA model with quadratic contrasts:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
quadAge	1	0.0009375	0.0009375	0.0682803	0.7950237
Residuals	46	0.6315875	0.0137302	NA	NA

Overall the linear model and the ANOVA models provide identical results: The linear effect explains a significant portion of the variance in the outcome; whereas, the quadratic effect does not. However, the linear model provides coefficients which can be used to interpret the magnitude of the effect. Specifically, for every increase in age unit playtime increases by .08 units.

h.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1162500	0.0435516	2.6692482	0.0106074
linAge	0.0775000	0.0533396	1.4529547	0.1533314
X2	0.0066667	0.0275444	0.2420331	0.8098791
linAge:X2	0.0050000	0.0337349	0.1482144	0.8828507

As tested in a linear model, we do not see a significant moderation effect between a linear age trend and sex ($t(3,44)=.148$, $p=.88$). The models being tested include:

model full: $Y_{playtime} = \beta_0 + \beta_{sex} \times X_{i,sex} + \beta_{linearAge} \times X_{i,age} + \beta_{sex \times linearAge} \times X_{i,sex \times linearAge} + \epsilon_i$

model reduced: $Y_{playtime} = \beta_0 + \beta_{sex} \times X_{i,sex} + \beta_{linearAge} \times X_{i,age} + \epsilon_i$

Question 2.

The two-way ANOVA, non-orthogonal case, has been a vexing problem for ANOVA researchers for many years. Please answer the following questions concerning the two-way non-orthogonal ANOVA.

a.

The word orthogonal refers to a geometric phenomenon wherein the dot product of two vectors equals 0, or if $c_i^T c_j = 0$ where c refers to a set of vectors and $i \neq j$. This occurs in a field of real numbers when these two vectors intersect with an angle of 90° . This logic is applied to the design matrix in balanced ANOVA models. Orthogonality exists in balanced ANOVA designs when:

- 1) The sum of the coefficients in each linear contrast sums to 0
- 2) The sum of the products of the corresponding coefficients in any two contrast equals zero

However, when these are not satisfied, a design is said to be unbalanced or non-orthogonal. So the term non-orthogonal when used in context to ANOVA modeling, refers to the phenomenon when a design matrix does not have independence amongst its cells in the design matrix.

b.

Designs are non-orthogonal when:

- 1) The sum of the coefficients in each linear contrast does not sum to 0
- 2) The sum of the products of the corresponding coefficients in any two contrast does not equal zero

One contributing factor is when data are unbalanced, so cells will have unequal representation from various populations being tested or are missing entire contrasts.

c.

Balanced ANOVA models display independence across the comparisons being made. So all forms of decomposing the models (i.e. Types I, II, & III sum of squares) provide identical solutions. For instance type I sum of squares is an ordered decomposition of the sum of squares, so models are tested in a consecutive fashion, but when the two main effects being tested display a large correlation across themselves, the variable fed into the model first will be given a larger weight to the model.

d.

The Type I effects are the order dependent or the sequential sum-of-squares. Assuming the main effect of A was introduced into the model first, the order of tests would follow (all models include intercepts):

$$\text{ME A: } R(A) = R() - R(A)$$

$$\text{ME B: } R(B|A) = R(A) - R(B|A)$$

e.

Type II tests assess the amount of variance a term adds when all terms are included except terms that contain the effect being tested. So the main effects of A and B would be tested as follows (all models include intercepts):

$$\text{ME A: } R(A|B) = R(B) - R(A,B)$$

$$\text{ME B: } R(B|A) = R(A) - R(A,B)$$

f.

Type III tests assess the amount of variance a predictor adds when partialling out all additional covariates. So the main effects of A and B would be tested as follows:

$$\text{ME A: } R(A|B) = R(B) - R(A,B)$$

$$\text{ME B: } R(B|A) = R(A) - R(A,B)$$

g.

Applebaum and Cramer (1972) suggest to use assess Type II effects when no moderation is present. This recommendation stems from the difficulty of interpreting two main effects when in the presence of an interaction.

h.

Applebaum and Cramer (1972) highlight that when assessing Type I effects it is important to test for independence amongst the predictors. This is important because if the predictors are confounded, only one predictor will be significant, wherein both are conveying similar information.

i.

Anova and linear models are in essence identical modeling techniques. However they differ in that ANOVA models typically do not include intercept effects whereas linear models include an intercept and reference group coding.

Question 3

Consider the following 2-way ANOVA Table with the group number listed in the cells of the table.

<i>Factor</i>	<i>B</i> = 1	<i>B</i> = 2	<i>B</i> = 3	<i>B</i> = 4
<i>A</i> = 1	<i>Group</i> = 1	<i>Group</i> = 2	<i>Group</i> = 3	<i>Group</i> = 4
<i>A</i> = 2	<i>Group</i> = 5	<i>Group</i> = 6	<i>Group</i> = 7	<i>Group</i> = 8

a.

With intercept model:

(Intercept)	factorA1	factorB3	factorB2	factorB1	factorA1:factorB3	factorA1:factorB2	factorA1:factorB1
1	1	0	0	1	0	0	1
1	1	0	1	0	0	1	0
1	1	1	0	0	1	0	0
1	1	0	0	0	0	0	0
1	0	0	0	1	0	0	0
1	0	0	1	0	0	0	0
1	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0

Cell mean model:

factorA2	factorA1	factorB3	factorB2	factorB1	factorA1:factorB3	factorA1:factorB2	factorA1:factorB1
0	1	0	0	1	0	0	1
0	1	0	1	0	0	1	0
0	1	1	0	0	1	0	0
0	1	0	0	0	0	0	0
1	0	0	0	1	0	0	0
1	0	0	1	0	0	0	0
1	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0

b.

```
proc glm data = data;
```

```
class a b;
```

```
model y = a b a*b;
```

```
run;
```

c.

```
proc glm data = data;
```

```
class a b;
```

```
model y = a b a*b;
```

```
contrast "ME A" a 1 -1;
```

```
contrast "ME B" b 1 0 0 -1,
```

```

0 1 0 -1,
0 0 1 -1;
run;

```

d.

```

proc glm data = data;
class a b;
model y = a b a*b;
contrast "linear B ME" b  $-\pi$   $-e$   $e$   $\pi$ ;
run;

```