MATH 204 - MID-TERM

The three factors are

- Factor A: Age Group
- Factor B: Therapy
- Factor C: City
- Response Y: change in cholesterol level.

(i) *Summarize the design by finding the numbers of levels of the three factors and the number of replicates. Is this a balanced complete design ? Justify your answer.*

The number of factor levels are as follows:

- Factor A: Age Group - $a = 3$ levels
- Factor B: Therapy - $b = 4$ levels
- Factor C: City - $c = 2$ levels

and there are $r = 5$ replicates for each of the 24 factor-level recombinations. Hence this is a balanced complete design, as all possible factor levels have the same non-zero number of replicates.

5 Marks

(ii) *Analyze the data from the two cities **separately**. Report the results of ANOVA-F tests of the hypotheses you deem appropriate, and comment on the validity of the analysis for these data.*

Page 3 contains a relevant summary of the SPSS output for the analyses of the two cities. The important output for the Bristol analysis is labelled $\boxed{1}$ and $\boxed{2}$, the output for Leeds is labelled $\boxed{3}$ and $\boxed{4}$.

- **Bristol Analysis:** The full factorial model A+B+A.B is fitted; Levene's Test (labelled $\boxed{1}$) indicates that the equal variances assumption is met, and the ANOVA table ANOVA-F test results (labelled $\boxed{2}$) indicate that **both main effects** and **interaction** are significant.

  Therefore, for the Bristol data, it seems that there is a significant effect of age group, and of therapy, and that there is a different effect of changing therapies in the different age groups.

  4 Marks

- **Leeds Analysis:** The full factorial model A+B+A.B is fitted; Levene's Test (labelled $\boxed{3}$) indicates that the equal variances assumption is met, and the ANOVA table ANOVA-F test results (labelled $\boxed{4}$) indicate that **both main effects** are significant, but that there is **no significant interaction** (p-value 0.247).

  Therefore, for the Leeds data, it seems that there is a significant effect of age group, and of therapy, but that the effect of changing therapies is the same in the different age groups.

  A further analysis of the Leeds data that omits the interaction factor could be carried out; such an analysis confirms that the both factors are significant (p-values both 0.000 to three decimal places - analysis not shown).

  4 Marks

The equal variances assumption is just one of the three assumptions that needs to be assessed. The independence assumption cannot be assessed without further information, but it is likely that it is met given the description of the study design. The normality of the data (and the random errors) could be checked using **boxplots** (see page 5 for examples), but these plots are difficult to interpret for such small number of replicates. The boxplots give no categorical evidence that the normality assumption is not met.

2 Marks

(iii) *Analyze the **pooled** data in a similar way, assuming a **three factor** factorial design (that is, use Factor A, Factor B and Factor C). Report the conclusions of the three factor analysis.*

Page 4 contains a relevant summary of the SPSS output for the pooled analysis. The important output is labelled $\boxed{5}$, $\boxed{6}$ and $\boxed{7}$.

The key points of the analysis are as follows:

- Levene's Test (labelled $\boxed{5}$) indicates that the equal variances assumption is met, but that the test statistic is almost significant at the $\alpha = 0.05$ level ($p = 0.053$). This result is a cause for concern, and implies that we need to treat the results of this pooled analysis with some care. However, given a strict interpretation of the test result as not significant, we may proceed with the ANOVA.

  4 Marks

- The ANOVA-F test results (labelled $\boxed{6}$) need careful consideration. It appears that the two main effects for Age Group and Therapy are significant (each $p = 0.000$ to three decimal places), but that the city main effect is not significant ($p = 0.916$). The two-way interaction between Age Group and Therapy is significant ($p = 0.002$) but the other two-way interactions are not significant. The three-way interaction is significant ($p = 0.001$); this indicates that the **interaction** between Age Group and Therapy is **different** in the two cities. This confirms what we observed in the individual cities analysis; in Bristol there was an interaction, and in Leeds there was not.

  5 Marks

- A further analysis of the pooled data that omits the various interaction factor could be carried out. However, on carrying out these analyses, it is evident that **Levene's test indicates non-constant variances** within each analysis (although here, perhaps a multiple testing correction could be made). We return to this point in part (iv).

The other assumptions underlying the ANOVA analysis, independence and normality, can be assumed to hold, the latter confirmed by inspection of boxplots.

1 Mark

(iv) *Given the results of all your analyses, report a conclusion as to whether the therapies alter cholesterol level, and whether the data from the two cities should be pooled into a single ANOVA analysis, giving a brief justification on each point.*

- In all analyses, it is evident that the different therapies yield different changes in cholesterol level. In the means plots graphs on page 6, it seems that therapy 3 yields the smallest change in cholesterol level in both cities.

  2 Marks

- The analyses above indicate that, on balance, the data from the cities should **NOT** be pooled. The reasons are twofold; first, the pattern of interaction is different in the two cities, and although this is picked up in the pooled analysis, it seems more natural to reflect this difference in separate analysis. Secondly, and more compelling, is the indication that the **variances** in the two cities are **different**. Levene's test in the full factorial model indicated this, although the result was not quite significant at $\alpha = 0.05$, but this was even more strongly indicated by the results of the analyses in $\boxed{7}$. The estimates of the population variances in the two cities can be obtained from the MSEs in the ANOVA table:

  $$\text{Bristol} : \widehat{\sigma}^2 = \text{MSE} = 2.710 \qquad \text{Leeds} : \widehat{\sigma}^2 = \text{MSE} = 0.957$$

  so it seems that the variance in Leeds is lower.

  3 Marks

Note that it may be possible to argue convincingly that the pooled analysis should be carried out; if this argument is made, with supporting evidence, then the 3 marks can be given.

MATH 204 – MidTerm
Winter 2008

Bristol Analysis

Levene's Test

Dependent Variable: Y

F	df1	df2	Sig.
1.128	11	48	.362

1

ANOVA Table

Dependent Variable: Y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	236.645	11	21.513	7.937	.000
Intercept	12.623	1	12.623	4.657	.036
AgeGroup	35.719	2	17.859	6.589	.003
Therapy	121.862	3	40.621	14.987	.000
AgeGroup * Therapy	79.064	6	13.177	4.862	.001
Error	130.103	48	2.710		
Total	379.371	60			
Corrected Total	366.748	59			

2

Leeds Analysis

Levene's Test

Dependent Variable: Y

F	df1	df2	Sig.
.628	11	48	.796

3

ANOVA Table

Dependent Variable: Y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	150.755	11	13.705	14.327	.000
Intercept	11.232	1	11.232	11.742	.001
AgeGroup	51.432	2	25.716	26.884	.000
Therapy	91.482	3	30.494	31.879	.000
AgeGroup * Therapy	7.841	6	1.307	1.366	.247
Error	45.915	48	.957		
Total	207.902	60			
Corrected Total	196.670	59			

4

Pooled Analysis

Levene's Test

Dependent Variable: Y

F	df1	df2	Sig.
1.629	23	96	.053

5

ANOVA Table

Dependent Variable: Y

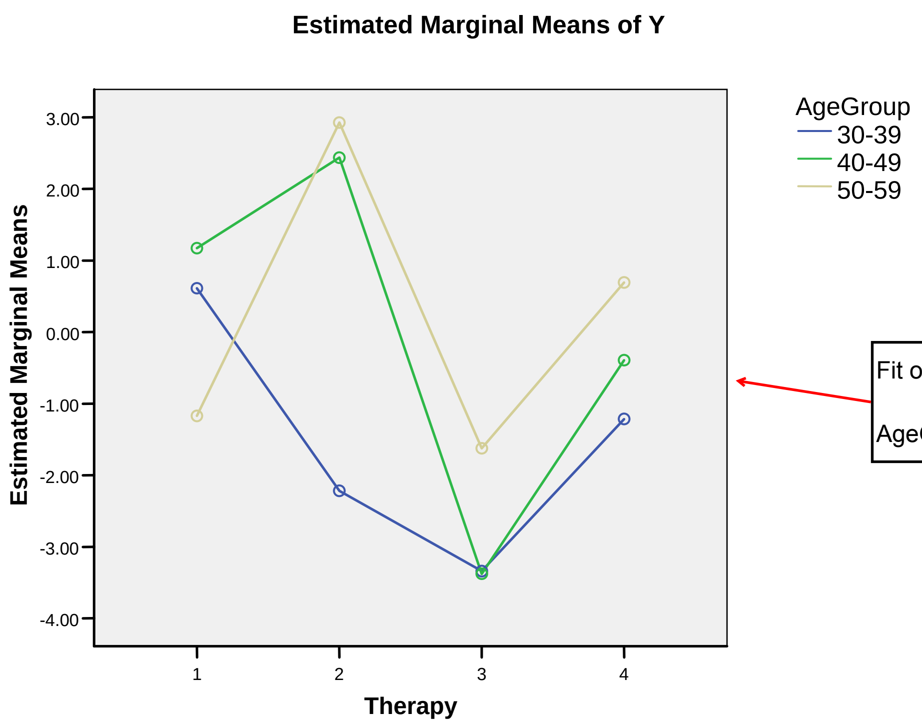
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	387.420	23	16.844	9.187	.000
Intercept	23.834	1	23.834	12.999	.000
AgeGroup	81.520	2	40.760	22.230	.000
Therapy	206.512	3	68.837	37.544	.000
City	.020	1	.020	.011	.916
AgeGroup * Therapy	42.529	6	7.088	3.866	.002
AgeGroup * City	5.631	2	2.815	1.536	.221
Therapy * City	6.832	3	2.277	1.242	.299
AgeGroup * Therapy * City	44.376	6	7.396	4.034	.001
Error	176.018	96	1.834		
Total	587.272	120			
Corrected Total	563.438	119			

6

Model	Levene's Test F	df1	df2	Sig.
AgeGroup + Therapy + AgeGroup*Therapy	2.330	11	108	.013
AgeGroup + Therapy + City + AgeGroup*Therapy + AgeGroup*City +Therapy*City	1.823	23	96	.023
AgeGroup + Therapy + City + AgeGroup*Therapy + AgeGroup*City	1.868	23	96	.019
AgeGroup + Therapy + City + AgeGroup*Therapy	2.176	23	96	.005

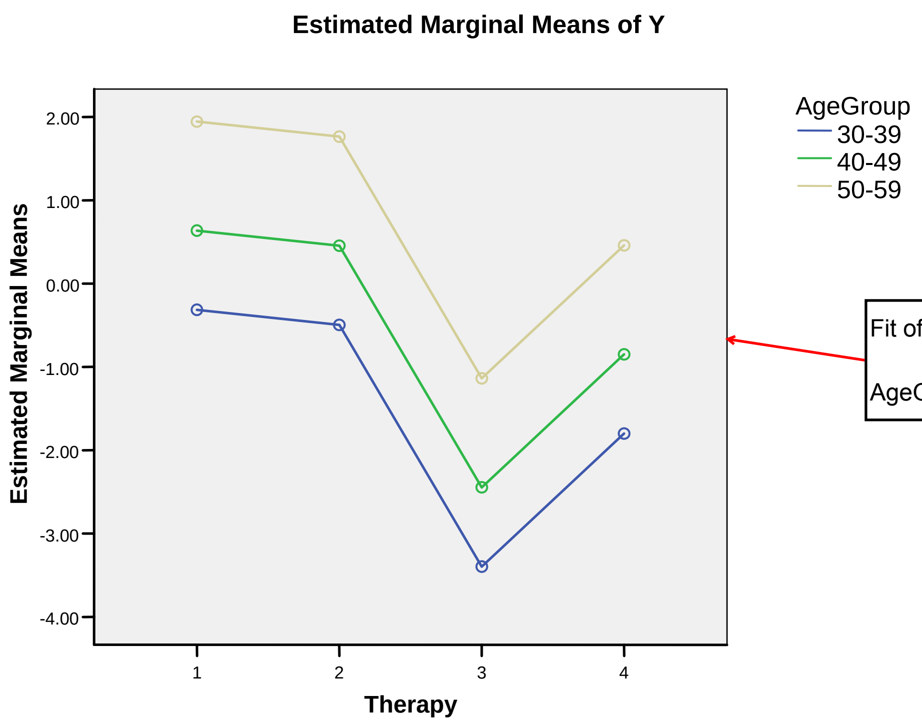
7

Bristol and Leeds Data: Marginal Means Plots



Fit of interaction model

AgeGroup + Therapy + AgeGroup*Therapy



Fit of main effects only model

AgeGroup + Therapy

Bristol and Leeds Data: Boxplots