620-270 Applied Statistics

Computer Lab 11

Analysis of covariance; random effects models for one-way data

In this lab session you will:

- perform analysis of covariance using the `glm` command.
- fit a random effects model to one-way data using the `anova` command.

1. The data: feeding young ducklings

The following data were extracted from a larger study concerned with determining the optimum amount of fish meal to be used as a supplement in rations for young growing ducklings. In the table, $x =$ age in weeks, $y =$ weight in grams; one would expect age to affect weight, and so $x$ could be expected to be a significant covariate.

<table>
<thead>
<tr>
<th>Amount of fish meal</th>
<th>5%</th>
<th>10%</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>$y$</td>
<td>$x$</td>
<td>$y$</td>
</tr>
<tr>
<td>1</td>
<td>46.5</td>
<td>2</td>
<td>67.8</td>
</tr>
<tr>
<td>5</td>
<td>151.6</td>
<td>4</td>
<td>142.8</td>
</tr>
<tr>
<td>8</td>
<td>287.1</td>
<td>7</td>
<td>365.2</td>
</tr>
<tr>
<td>11</td>
<td>545.9</td>
<td>10</td>
<td>637.7</td>
</tr>
<tr>
<td>12</td>
<td>634.2</td>
<td>12</td>
<td>855.9</td>
</tr>
</tbody>
</table>

Enter the data, putting $x$ into c1 and $y$ into c2. Put the fish meal indicator in c3 (1 for 5%, 2 for 10% and 3 for 15%). Name the columns $x$, $y$ and fish respectively.

Before analysing the data, we should explore it. Plot $y$ against $x$, using different symbols for the three groups. To do this, run the menu `Graph → Plot` and enter the variable names. Click on the triangle next to the `For each` button, and select `Group`. Now click on the rectangle under `Group variables` and enter `fish` (either by typing or by double clicking on the variable name). Click OK and the plot should appear.

From the plot, it appears that we might need to transform the data to create a more linear relationship between $y$ and $x$. This is not strictly essential, but it would make the analysis easier, particularly in comparing relationships for the three groups. Take the square root of $y$ (call it `sqy`) and plot it against $x$. This looks more linear, and so we will use $\sqrt{y}$ as the response variable and $x$ as the covariate.

From the above plot, do you think that we should use a model that has parallel lines for the three groups or a model with different slopes? We will now test this formally.

2. Analysis of covariance: different slopes

We want to fit the model $E(\sqrt{y} | x) = \mu + \alpha_i + \beta x + \beta_i x$, and test the hypothesis $\beta_i = 0$ for $i = 1, 2, 3$. Such models are best fitted using the `glm` command (`glm` = general linear model), or its menu equivalent. We used the `glm` command in Computer Lab 9, to perform multiple comparisons for a two-way ANOVA. In the session window, type
MTB > brief 3

MTB > glm sqrty = fish|x;

SUBC> covariate x;

SUBC> means fish.

Comments:
don’t forget this command, or the output
will not have everything you need.
this is shorthand for sqrty = fish + x + fish*x
(the model with interaction).
if you don’t specify x as a covariate,
age will be fitted as a factor with 10 levels!

Examine the \( P \)-value for the interaction; it is very small, and so we reject the hypothesis that
\( \beta_i = 0 \), and we need to fit different slopes for each group. From the output, work out the
equation of the line for the 5% group. You should get

\[
\sqrt{y} = 4.446 + 1.6866x.
\]

(for the 10% group you should get \( \sqrt{y} = 3.832 + 2.133x \); and for 15%, \( \sqrt{y} = 4.495 + 2.516x \); verify these using the output if you have time).

Notice how the slope increases as the amount of fish meal increases, a feature which is evident
from the scatter plot. Because of the unequal slopes, there is not much point in testing whether
\( \alpha_i = 0 \), i.e. whether the group means are significantly different when corrected for age, because
the differences depend on age.

Now perform the analysis using the menu
Stat \( \rightarrow \) Anova \( \rightarrow \) General Linear Model.

In the Responses box, enter sqrty, and in the Model box, type fish|x. Click the Covariates
button and enter x.

Examine the commands and output generated by this menu. Notice that there are not as many
coefficients listed under the analysis of variance table. This is because the menu has used the
subcommand SUBC> brief 2, overwriting our earlier command SUBC> brief 3. If you know
how to stop MINITAB doing this, please let me know! In the meantime, use the commands rather
than the menu to perform analysis of covariance.

3. Analysis of covariance: parallel lines

From both the scatter plot and the analysis, we would conclude that it is not sensible to fit
parallel lines to the data. For learning purposes, however, we are going to do that, so that you
know how it works in MINITAB. Enter the commands

MTB > brief 3;
MTB > glm sqrty = fish;
SUBC> covariate x;
SUBC> means fish.

(Note that instead of typing glm sqrty = fish you can type glm sqrty = fish + x; they
are equivalent, since both models include x because it is listed as a covariate.)

From the output, work out the equation of the line for the 5% group. You should get

\[
\sqrt{y} = 1.696 + 2.058x.
\]

Calculate the “least squares” (i.e. adjusted) mean for the 5% group. Check that you obtain
the value given at the bottom of the output (15.55). The expected weight of a duckling at an
average age (6.73 weeks), fed with 5% fish meal, would therefore be \( 15.55^2 = 242 \).
Finally, to fit a single line through all the data (i.e. assuming no differences between groups), type the command

MTB > glm sqnty = x;
SUBC> covariate x.

Of course, this is just a long-winded way of running a simple linear regression! Now fit the equation using regression (either by a command or by the menu), and compare the output with that resulting from glm.

4. Random effects models

You have learnt the following commands in MINITAB to fit models:

- **regress** for fitting linear regression models.
- **aovoneway** for one-way models on data of balanced or unbalanced design, in unstacked format.
- **oneway** as above, with data in stacked format.
- **twoway** for two-way models, on data with balanced design.
- **glm** for fitting general linear models with factors and/or covariates on data which may be of balanced or unbalanced design.

Now you will learn the last one:

- **anova** for multi-way models, on data with balanced design, in which factors may be fixed or random.

Here we will consider one-way data. Using MINITAB for analysing two-way data with random effects or mixed effects is described in Problem Set 12.

The following table shows the potency measurements taken from vats of liquid medication produced by a pharmaceutical company. A random sample of five vats from a month’s production was obtained, and four separate samples were selected from each vat.

<table>
<thead>
<tr>
<th>Vat 1</th>
<th>Vat 2</th>
<th>Vat 3</th>
<th>Vat 4</th>
<th>Vat 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>2.6</td>
<td>3.4</td>
<td>4.2</td>
<td>1.8</td>
</tr>
<tr>
<td>3.8</td>
<td>2.9</td>
<td>3.9</td>
<td>4.4</td>
<td>2.3</td>
</tr>
<tr>
<td>3.5</td>
<td>2.8</td>
<td>3.3</td>
<td>4.3</td>
<td>1.9</td>
</tr>
<tr>
<td>3.0</td>
<td>2.0</td>
<td>3.1</td>
<td>4.2</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Enter the potency measurements in c1 and the vat indicators in c2. Explore the data with a scatter plot: plot c1*c2. Does the effect of vat look significant?

We will fit the random effects model $Y_{ij} = \mu + A_i + E_{ij}$, where $A_i$ is the effect due to vat $i$. In the session window, type the commands

MTB > anova c1 = c2;
SUBC> random c2;
SUBC> ems.
The analysis of variance looks the same as for a one-way ANOVA with fixed effects, but there is a table of variance components and expected mean squares as well.

Use the mean squares in the ANOVA table and the expected mean squares in the table below it to estimate the variance component for vat. Check that it agrees with the value given in the Minitab output. The variation between vats is clearly large compared to the variation within vats. The highly significant $F$ ratio in the ANOVA table confirms that there is significant variation between vats, i.e. we reject the hypothesis that $\sigma_A^2 = 0$.

Now perform the analysis using the menu, by following the path

Stat → ANOVA → Balanced ANOVA

entering c1 under Responses, c2 under Model, and c2 under Random factors. Click on Results and tick the box Display expected mean squares and variance components. Check that the output is the same as that produced by the commands.

Random effects models can also be fitted in Minitab using the glm command. We won’t bother with that now, but if the data are unbalanced and are two-way (or more), anova will not work, and glm must be used.