PhD course: ANOVA Practicals, Sketch Solutions

Possible answer to ANOVA Practical question 1

(This is an annotated Minitab session: parts added outside Minitab have been written in italics: some output from some commands has been suppressed for clarity)

Melanin concentration and OAG dose are in columns 1 and 2

MTB > prin c1 c2

ROW melanin   dose

```
1       7.8       0
2      15.3       1
3      18.5       3
4       6.2       0
5      11.3       1
6      24.2       3
7      11.2       0
8      17.4       1
9      29.8       3
10      1.8       0
11      3.3       1
12      7.1       3
```

MTB > note
MTB > note Perform one way ANOVA
MTB > onew 'melanin' 'dose'

```
ANALYSIS OF VARIANCE ON melanin
SOURCE      DF        SS          MS         F         p
dose         2       351.8      175.9      3.57      0.072
ERROR        9       444.1       49.3
TOTAL       11       796.0
INDIVIDUAL 95 PCT CI'S FOR MEAN
BASED ON POOLED STDEV
LEVEL       N       MEAN      STDEV   --+---------+---------+---------+----
0      4       6.750      3.903   (--------*---------)
1      4      11.825      6.221            (---------*---------)
3      4      19.900      9.701                      (---------*---------)
                                   --+---------+---------+---------+----
POOLED STDEV =     7.025                0.0       8.0      16.0      24.0
```

MTB > note
MTB > note Some evidence of difference but note increasing SD as
MTB > note dose increases. Try logs. (here logs to base *e* are used: base 10 gives similar results, although the values of the sums of squares and mean squares are scaled by a factor $(\log10)^2$)
MTB > note
MTB > name c4 'logmel'
MTB > let 'logmel'=log('melanin')

MTB > onew 'logmel' 'dose'

```
ANALYSIS OF VARIANCE ON logmel
SOURCE      DF         SS         MS         F         p
dose         2      2.618      1.309      2.44      0.142
ERROR        9      4.826      0.536
TOTAL       11      7.444
INDIVIDUAL 95 PCT CI'S FOR MEAN
BASED ON POOLED STDEV
LEVEL       N       MEAN      STDEV   ---------+---------+---------+-------
0        4     1.7206     0.7935   (----------*---------)
1        4     2.3008     0.7598              (----------*---------)
3        4     2.8647     0.6338                  (----------*---------)
---------+---------+---------+-------
POOLED STDEV =    0.7323                    1.60      2.40      3.20
```

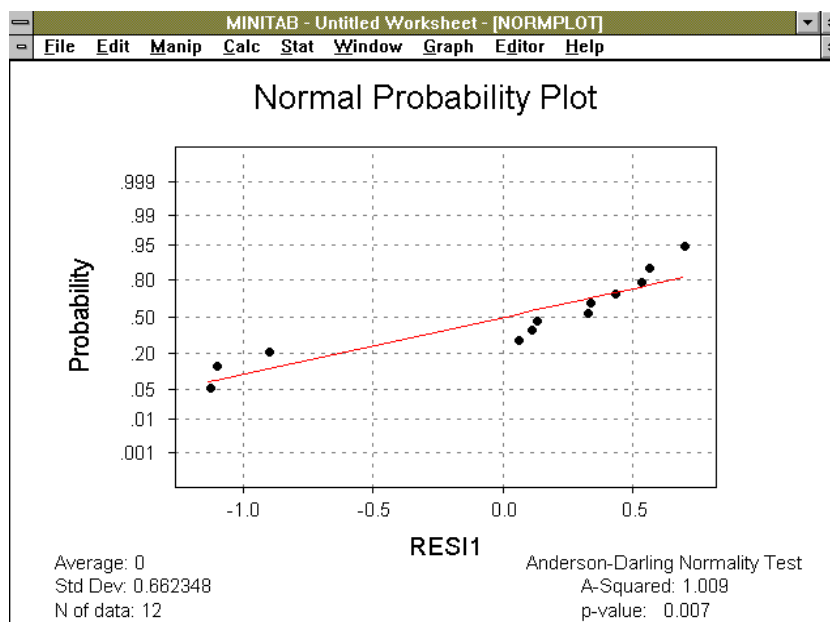SDs now much more similar; still no evidence of difference.

Sample is a bit small for testing normality but we can go through the motions.

(remember the residuals are = (observation) - (group mean).

```
MTB > onew 'logmel' 'dose' 'RESI1'
```

Output suppressed

Now perform a Normal Plot (from **Graph** menu)

There seem to be some problems, with three residuals located some way from the others. Print data to have a look

```
MTB > prin c1 c2 'RESI1'

  ROW   melanin    dose        RESI1

 1       7.8       0    0.33353
 2      15.3       1    0.42709
 3      18.5       3    0.05309
 4       6.2       0    0.10396
 5      11.3       1    0.12404
 6      24.2       3    0.32167
 7      11.2       0    0.69532
 8      17.4       1    0.55571
 9      29.8       3    0.52983
10       1.8       0   -1.13281
11       3.3       1   -1.10684
12       7.1       3   -0.90459
```

The problem surrounds three small melanin responses all on the same foreskin

In order to get round this problem a more sophisticated analysis of variance should be used. The present analysis takes no account of the fact that each of four foreskins has been divided into three and the three doses applied to the same foreskin. This is an important aspect of the *structure* of the data which has not been specified in the analysis: it is usually a good guide to the correctness and appropriateness of an analysis that it takes full account of the structure of the data. An outline of an approach which takes account of this feature of the data is overleaf.

The experimenter ensured that each foreskin received each dose, which is clearly a sensible approach (suppose most dose 3s were given to foreskin 4, it is then unlikely that the effect of dose 3 could have been properly discerned). However, the above analysis only takes account of the variation from a single source, namely doses, and the role of foreskins is ignored. What is needed is a two-way ANOVA, shown below, that takes account of the variation between foreskins.
In Minitab the analysis is accomplished through the TWOWAY command or, with more effort but with more informative output, from the ANOVA command, as

```
MTB > ANOVA 'log mel' = Dose foreskin.
```

ANOVA: log mel versus Dose, foreskin

```
Some output suppressed

Analysis of Variance for log mel

Source        DF          SS          MS        F       P
Dose           2      2.6181      1.3090    51.75   0.000
foreskin       3      4.6740      1.5580    61.59   0.000
Error          6      0.1518      0.0253
Total         11      7.4438
```

This is very different to the one-way analysis; the line for OAG doses is the same, but now it can be seen that mos of the variation previously ascribed to random error can be seen to be due to differences between foreskins
        (the Error MS is now much smaller). Consequently the effect of dose is now seen to be very clear, an impression that is apparent from inspecting the trend with dose within each group of the raw data.