H. Spliid, October 20, 2010

Solution to exercise 55

Question 1
One plate is in this experiment a block.

The outcome of one single measurement is assumed to depend additively of the type of penicillin and the effect from the plate (block) on which it is tested.

As the text is formulated we may consider the experiment as an experiment with 12 blocks, especially if all 12 plates are prepared and grown within a short period of time.

For one plate the measured value Y_{ij} can be expressed as:

$$Y_{ij} = \mu + \tau_i + B_j + E_{ij}; \quad \sum \tau_j = 0; \quad B_j \in NID(0, \sigma_B^2); \quad E_{ij} \in NID(0, \sigma_E^2)$$

where τ_i is the effect from the i'th type (of penicillin), and B_j is the effect from the j'th plate. The design used is

Type of penicillin

	A	B	C	D	E	F	G	H	I	J	S
Plate	1	2	3	4	5	6	7	8	9	10	11
1	y	y	y	y	y						y
3	y	y	y	y	y						y
5	y	y	y	y	y						y
7	y	y	y	y	y						y
9	y	y	y	y	y						y
11	y	y	y	y	y						y
2						y	y	y	y	y	y
4						y	y	y	y	y	y
6						y	y	y	y	y	y
8						y	y	y	y	y	y
10						y	y	y	y	y	y
12						y	y	y	y	y	y

Question 2
When a type, fx A, of penicillin is compared with the standard S, it is done by comparing the measurements for A and S on the same plate.

If we consider plate no. 1 and the measurements for A and S we can compute their difference. The model gives for plate no. 1 the difference $Y_{A1} - Y_{S1}$ as

$$D_{A-S,1} = (\mu + \tau_A + B_1 + E_{A1}) - (\mu + \tau_S + B_1 + E_{S1}) = \tau_A - \tau_S + E_{A1} - E_{S1}$$

where E_{A1} and E_{S1} are the measurement errors from the A-measurement and the S-measurement, respectively. Note that μ and B_1 disappear in this difference.

We can compute the mean and the variance of $D_{A-S,1}$. We get

$$\text{Mean}(D_{A-S,1}) = \text{Mean}(\tau_A - \tau_S + E_{A1} - E_{S1}) = \tau_A - \tau_S$$

because E_{A1} and E_{S1} both have zero expectation. Next

$$\text{Var}(D_{A-S,1}) = \text{Var}(\tau_A - \tau_S + E_{A1} - E_{S1}) = \text{Var}(E_{A1}) + \text{Var}(E_{S1}) = 2\sigma_E^2$$

because τ_A and τ_S are constants and have zero variance.

We can do this for the plates where A and S are together, that is for the plates 1, 3, 5, 7, 9 and 11, and then take the average difference. We then get our result, which (of course) is the average difference of the measurements for types A and S over these plates:

$$\widehat{\tau_A - \tau_S} = (D_{A-S,1} + D_{A-S,3} + ... + D_{A-S,11})/6$$

and it has the variance

$$\text{Var}(\widehat{\tau_A - \tau_S}) = (2\sigma_E^2 + 2\sigma_E^2 + ... + 2\sigma_E^2)/6^2 = \sigma_E^2/3$$

Question 3
The types grown on the same plates can all be compared with the above precision.

Types that are grown on different plates cannot be paired wherefore the variation between plates cannot be eliminated directly.

Take for example types A and F from plates no. 1 and 2 with measurements Y_{A1} and Y_{F2}, respectively. Their difference is

$$Y_{A1} - Y_{F2} = (\mu + \tau_A + B_1 + E_{A1}) - (\mu + \tau_F + B_2 + E_{F2}) = \tau_A - \tau_F + B_1 - B_2 + E_{A1} - E_{F2}$$

and it has the variance

$$\text{Var}(Y_{A1} - Y_{F2}) = 2\sigma_B^2 + 2\sigma_E^2$$

Taking the average of the measurements for types A and F we get

$$\widehat{\tau_A - \tau_F} = \overline{Y_A} - \overline{Y_F}$$

and it has the variance

$$\text{Var}(\widehat{\tau_A - \tau_F}) = \sigma_B^2/3 + \sigma_E^2/3$$

Another possibility is to 'go over the standard'. Using types A and F again as example we could compute

$$(Y_{A1} - Y_{S1}) - (Y_{F2} - Y_{S2}) = \tau_A - \tau_F + (E_{A1} - E_{S1}) - (E_{F2} - E_{S2})$$

by which method the B's (the plates) are eliminated. This difference has the variance

$$\text{Var}((E_{A1} - E_{S1}) - (E_{F2} - E_{S2})) = 4\sigma_E^2$$

Taking its average over all 6 + 6 plates gives the result

$$\text{Var}(\widehat{\tau_A - \tau_F})_{over\ standard} = 4\sigma_E^2/6 = 2\sigma_E^2/3$$

If $2\sigma_E^2/3 < \sigma_B^2/3 + \sigma_E^2/3$, that is if $\sigma_E^2 < \sigma_B^2$, the latter method is preferable. This will most often be the case in practise.

By growing the standard on all plates it is thus possible to eliminate the variation between plates, but the precision of comparisons between types depend on whether the types are grown together or not!

Question 4

The way to avoid the above difficulties is to use an incomplete balanced block design.

There are 11 "treatments", but the block size is only 6.

From the tables of incomplete block designs we can look up a design with blocksize 6 and and 11 treatments. It has 11 blocks and is

	A	B	C	D	E	F	G	H	I	J	S
1		y	y	y				y		y	y
2	y		y	y	y				y		y
3	y	y		y	y	y				y	
4		y	y		y	y	y				y
5	y		y	y		y	y	y			
6		y		y	y		y	y	y		
7			y		y	y		y	y	y	
8				y		y	y		y	y	y
9	y				y		y	y		y	y
10	y	y				y		y	y		y
11	y	y	y				y		y	y	

In this design we base the estimation on Q_A and Q_F, for example, giving (with k=6, λ=3 and t=11):

$$\begin{aligned}
\hat{\tau}_A &= k \cdot Q_A/(\lambda \cdot t) \\
\hat{\tau}_F &= k \cdot Q_F/(\lambda \cdot t) \\
\widehat{\tau_A - \tau_F} &= (Q_A - Q_F)\frac{6}{3 \cdot 11} = (Q_A - Q_F)\frac{2}{11} \\
\text{Var}(\widehat{\tau_A - \tau_F}) &= 2 \cdot \sigma_E^2 \frac{k}{\lambda t} = \sigma_E^2 \frac{2 \cdot 6}{3 \cdot 11} = \sigma_E^2 \frac{4}{11}
\end{aligned}$$

Taking into account that this design only uses 11 plates instead of 12 as in the first design we se that <u>all</u> comparisons can now be made and with the same precision as before for types grown together.

The estimate for σ_E^2, i.e. $\widehat{\sigma}_E^2$, is computed from the residual sum of squares in the ANOVA table.

Question 5

The experiment is analysed by the standard technique for the balanced incomplete block design. The design is symmetrically balanced so it is possible also to assess the variation between blocks.

Question 6

Because of the symmetrical balance the design can be layed out as a Youden square (see the table):

	Types of penicillin										
Blocks	A	B	C	D	E	F	G	H	I	J	S
1		α	β	γ				δ		ε	ζ
2	ζ		α	β	γ				δ		ε
3	ε	ζ		α	β	γ				δ	
4		ε	ζ		α	β	γ				δ
5	δ		ε	ζ		α	β	γ			
6		δ		ε	ζ		α	β	γ		
7			δ		ε	ζ		α	β	γ	
8				δ		ε	ζ		α	β	γ
9	γ				δ		ε	ζ		α	β
10	β	γ				δ		ε	ζ		α
11	α	β	γ				δ		ε	ζ	

Question 7

This design is also analysed by the standard technique. The only difference in comparison with question 5 is that also the last source of variation can be eliminated and (possibly) the precision of the design can be improved giving a smaller residual variance. For example:

$$\begin{aligned} Y_{ijk} &= \mu + \tau_i + B_j + Pos_k + E^*_{ijk} \\ B_j &\in N(0, \sigma_B^2) \\ E^*_{ijk} &\in N(0, \sigma_{E^*}^2) \\ Pos_k &\in \{Pos_1, Pos_2, Pos_3, Pos_4, Pos_5, Pos_6\}; \sum Pos_i = 0 \end{aligned}$$

corresponding to 6 specific positions each with its own parameter.

The order of magnitude for the residual variance in this model, $\sigma_{E^*}^2$, will be:

$$\sigma_{E^*}^2 \approx \sigma_E^2 - \frac{1}{6}\sum_{k=1}^{6} Pos_k^2$$

but its estimate will have 5 degrees of freedom less.

There are 66 measurements, giving a total of 65 degrees of freedom.

The τ's take 11-1 = 10 degrees of freedom.

The B's take 11-1 = 10 degrees of freedom.

The Pos's take 6-1 = 5 degrees of freedom.

Thus the residual sum of squares in the Youden square design will have 65-10-10-5 = 40 degrees of freedom, which will be more than enough for all practical cases!

Since the positions are balanced versus the treatments (and blocks) the estimates for the treatments are exactly as before (in question 4), but we now also compute variation between positions. If the totals for the 6 positions are $T_\alpha, T_\beta, \ldots, T_\zeta$ each based on 11 observations the sum of squares for positions is

$$SSQ_{pos} = (T_\alpha^2 + T_\beta^2 + \ldots + T_\zeta^2)/11 - T_{total}^2/66 \qquad \text{d.f.} = \text{6-1}$$

and, to sum up, as usual:

$$SSQ_{Penicillin} = \frac{k}{\lambda t} \sum_{j=1}^{11} Q_j^2, \qquad \text{d.f.} = \text{11-1}$$

$$SSQ_{Blocks} = \sum_{j=1}^{11} T_{i,.}^2/6 - T_{total}^2/66 \qquad \text{d.f.} = \text{11-1}$$

$$SSQ_{Total} = \sum_{i,j} Y_{ij}^2 - T_{total}^2/66 \qquad \text{d.f.} = \text{66-1}$$

$$SSQ_{Residual} = SSQ_{Total} - SSQ_{Blocks} - SSQ_{pos} - SQ_{Penicillin}$$

with d.f. = (65-1)-(6-1)-(11-1)-(11-1) = 40 ,

and again, for example:

$$\begin{aligned}
\hat{\tau}_A &= k \cdot Q_A/(\lambda \cdot t) \\
\hat{\tau}_F &= k \cdot Q_F/(\lambda \cdot t) \\
\widehat{\tau_A - \tau_F} &= (Q_A - Q_F)\frac{6}{3 \cdot 11} = (Q_A - Q_F)\frac{2}{11} \\
\text{Var}(\widehat{\tau_A - \tau_F}) &= 2 \cdot \sigma_E^2 \frac{k}{\lambda t} = \sigma_E^2 \frac{2 \cdot 6}{3 \cdot 11} = \sigma_E^2 \frac{4}{11}
\end{aligned}$$

So a systematic variation between positions can be removed from the data by the balanced design shown.

[End-of-solution]