Petter Mostad
Matematisk Statistik
Chalmers

Suggested solution for exam in MSA830: Statistical Analysis and Experimental Design, March 2009

1. (a) The merits of the four choices of experimental plan are discussed below:
   - In this plan, there is a danger that the effects of systematic differences across the plot, such as soil quality or water in the ground, could be confounded with the effect of the different fertilizers. Thus it is not a good idea to apply the fertilizers in the same pattern as for example the soil could possibly be varying.
   - This randomization plan is fairly good, except that he might end up using the different fertilizers different number of times, reducing the efficiency with which he can compare their effect.
   - This plan is probably the best among the four plans. Eric gets to test each fertilizer exactly four times, and the effects of such things as soil and water, which may be expected to vary smoothly across the plot, are *blocked* as well as possible.
   - This plan is also good, the difference between the previous plan is that it does not use blocking among rows and columns to block the effects of environmental variables. Thus it may be judged slightly inferiour to the previous plan, to the extent that the environmental variables are varying smoothly across the rows and columns.

   (b) The means in the A and B groups are 38 and 32, and the sample variances are 34 and 37, respectively. The average increase in yield when using fertilizer A instead of B is thus 6. It may be reasonable in this context to assume that the population variances for A and B results are equal, and the data does not contradict this. Using a two-sample t-test where variances are assumed equal, we get the t-statistic 1.5922. Comparing this with a t-distribution with 8 degrees of freedom, we get that a two-sided p-value is between 0.1 and 0.2. Thus the difference in the means can not be considered statistically significant. If one instead chooses a t-test where population variances are not assumed equal, we get the same t-statistic, but we now compare it with a t-distribution with 7.98 degres of freedom; in other words, we get the same result.

2. (a) One way to fill out the experimental plan is

A	B	C	D	E	F	G
-	+	-	+	+	-	-
-	+	-	-	-	+	+
-	-	+	+	-	-	+
-	-	+	-	+	+	-
+	-	-	-	+	-	+
+	-	-	+	-	+	-
+	+	+	-	-	-	-
+	+	+	+	+	+	+

which has been derived using $B = AC$, $D = AF$, $E = CF$, and $G = ACF$.

(b) The main effect of A:

$$\frac{-43 - 56 - 24 - 98 + 34 + 52 + 38 + 81}{4} = -4$$

The main effect of F:

$$\frac{-43 + 56 - 24 + 98 - 34 + 52 - 38 + 81}{4} = 37$$

(c) If there are effects of unmeasured variables that are correlated to the order in which she does her experiments, such effects are confounded with for example the effect of the A factor. It may be a better idea to randomize the order in which the experiments are done.

(d) $2^7 = 128$

(e) The simplest solution in Annas situation is to perform another 8 experiments using exactly the same experimental plan. If she followed a different experimental plan, she would have to assume that some of the interaction effects between the factors A through F are zero, in order to compute a confidence interval.

3. (a) The table becomes

	Sums of squares	Degrees of freedom	Mean square	F ratio	p-value
Between hormones	120	2	60	4	Between 0.025 and 0.01
Between mice	180	6	30	2	Between 0.1 and 0.05
Residuals (errors)	3015	201	15		
Total	3315	209			

(b) The assumptions of this ANOVA analysis are: The expected effects of the hormones and of the mice type are additive, so that there is no interaction effect between the two. For each combination of hormones and mice, the results are normally distributed, and the variances in all of the groups are the same. The assumption that there is no interaction effect could be tested with an ANOVA table including an interaction effect. To check that the assumptions of normality and equal variaces are reasonable, one may study the residuals, i.e., plot them in various ways.

(c) Harri could conclude that there is a significant difference between the effect of the hormones, whereas the differences between the mice is not significant. (However, using mice as a blocking variable still improves the estimation of the effect of the effect of the hormones).

4. At least two different tests could be used. If Jenny makes the assumption that the differences in running times are normally distributed, she could make a t-test for the hypothesis that the mean of the population of such differences is zero. Specifically, if the mean of the differences is denoted $\overline{x}$, and their sample variance is denoted s^2, she could compute $\overline{x}/\sqrt{s^2/7}$ and compare the resulting statistic with a t distribution with 6 degrees of freedom,

to obtain a p-value for the test. The only assumption needed would be that the differences are normally distributed. Another alternative would be to compare $\overline{x}$ with the corresponding statistic obtained by averaging over all possible differences in running times, where for each athlete, one can compute Type A minus Type B, or Type B minus Type A. Taking twice the fraction of these numbers that are "more extreme" than $\overline{x}$ would give an approximate p-value for a two-sided test of the null hypothesis that the shoe types do not influence the running time. No assumptions would be needed. Note that this test would give a fairly rough p-value, as there are only $2^7 = 128$ possible ways of computing the mean of differences with Type A and Type B permuted.

5. (a) The probability is $0.85 \cdot 0.15 \cdot 0.85 \cdot 0.15 \cdot 0.85 \cdot 0.85 \cdot 0.85 = 0.01$.

   (b) The probability is

$$\frac{7!}{2!5!}0.15^2 0.85^5 = 0.2097$$

   (c) Here, we can use a normal approximation: The number of days that she will win is Binomially distributed with expectation $0.15 * 365 = 54.75$ and variance $0.15 * 0.85 * 365 = 46.5375$. To find the probability for a normal distribution with this expectation and variance to produce values of 20 or less, we can compute

$$\frac{20 - 54.75}{\sqrt{46.5375}} = -5.09$$

   Comparing this with the table for the normal distribution (and disregarding continuity corrections) we get that the probability is less than $1 - 0.99997 = 0.00003$.

6. (a) As the daily observed values are not independent, it is inappropriate to use a t-test. Instead, Hans could compute the average over every period of consecutive 30 days for the last 2 years (including also overlapping periods of 30 days in this set). He could then compare the average of the values for the 30 days the new system has been running with the previous computed set of averages. Twice the fraction of such averages that are "more extreme" than the new average could be considered the p-value for a test of whether the new system produces a different amount of pollution than the old system.

   (b) Hw would only need to assume, under the null hypothesis, that any 30 day period is exchangeable with any other such period.