Petter Mostad
Matematisk Statistik
Chalmers

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

1. (a) To use a t-test, one must assume that both groups of numbers are independent random samples from normal distributions. She may or may not assume equal variances.

   (b) For process A we get sample mean 6 and sample variance 10/3, and for process B we get sample mean 9 and sample variance 5/2. One can *either* choose to assume equal population variances, or not to make this assumption. **Assuming equal population variances**, the pooled variance estimate becomes

$$s_p^2 = \frac{\frac{10}{3} \cdot 3 + \frac{5}{2} \cdot 4}{3+4} = \frac{20}{7} = 2.857,$$

   and the test statistic becomes

$$\frac{9-6}{\sqrt{\frac{20}{7}(\frac{1}{4}+\frac{1}{5})}} = 2.646.$$

   Comparing this with $t_{7,0.025} = 2.365$, we conclude that we can reject the null hypothesis at a 5% significance level. The 95% confidence interval for the differences in population means is

$$9 - 3 \pm t_{7,0.025}\sqrt{\frac{20}{7}(\frac{1}{4}+\frac{1}{5})} = 3 \pm 2.68.$$

   Alternatively, **not assuming equal population variances**, the test statistic becomes

$$\frac{9-6}{\sqrt{\frac{10}{3}/4 + \frac{5}{2}/5}} = 2.598$$

   and the degrees of freedom

$$df = \frac{\left(\frac{10/3}{4} + \frac{5/2}{5}\right)^2}{\left(\frac{10/3}{4}\right)^2/3 + \left(\frac{5/2}{5}\right)^2/4} = 6.047 \approx 6$$

   Comparing 2.598 with $t_{6,0.025} = 2.447$ we again get that we can reject the null hypothesis at a 5% significance level. The 95% confidence intervals in population means becomes

$$9 - 3 \pm t_{6,0.025}\sqrt{\frac{10/3}{4} + \frac{5/2}{5}} = 3 \pm 2.83.$$

   (c) The results in b) are based on the assumption that the given values are random samples, so that they are assumed independent. Consecutive runs of this process may not be independent. Also, the order in which all the runs are made has not been randomized, meaning that nuisance factors that change over time could influence the result. Finally, with so few observations, it is difficult to verify the assumption of normal distributions.

(d) A possibility is to use an "external reference distribution": From the data for process A, Alice could compute averages of 5 consecutive runs with various starting points in the data series. The averages for all such possible starting points constitute a sample with which the average of the five yields for process B could be compared. If the last average is "extreme" compared to the population of averages for process A, one could use this to reject the null hypothesis that data from process A and process B are interchangeable. More precisely, if the average for process B was higher than the median of the sample values for process A, the p-value would be twice the proportion of the sample values larger than the average for process B, and corresponding in the opposite case.

2. (a) Estimate for the main effect of factor B:

$$\frac{-439.2 + 219.7 - 331.3 + 192.9}{6 \cdot 2} = -29.8.$$

(b) The pooled variance estimate is

$$s_p^2 = \frac{4295 + 1740 + 3855 + 873}{4} = 2690.75,$$

and as each of the 4 sample variances in this average has 5 degrees of freedom, the pooled variance estimate has 20 degrees of freedom. The standard error becomes

$$\sqrt{s_p^2\left(\frac{1}{12} + \frac{1}{12}\right)} = \sqrt{2690.75/6} = 21.17,$$

and thus the confidence interval becomes

$$-29.8 \pm t_{20,0.025} \cdot 21.17 = -29.8 \pm 44.2.$$

For this confidence interval to be valid, one must assume that the random errors in the model are normally distributed with the same variance for each combination of settings of the factors.

(c) We have

$$\begin{aligned} AB &= C \\ AC &= B \\ BC &= A \end{aligned}$$

meaning that, for each line, the given interaction effect is confounded with the given main effect.

(d) The question is a bit unclear: If Anna should do 24 experiments *instead* of the ones she has done, she should use a full factorial design with three replications, i.e., doing 3 replications with the following design:

A	B	C
-	-	-
-	-	+
-	+	-
-	+	+
+	-	-
+	-	+
+	+	-
+	+	+

If she should do 24 experiments *in addition* to the ones she has performed, she should follow the same experimental plan again, but with the signs reversed.

3. The important thing is that each friend should do exactly one test with each soil type and with each seed type. The resulting experimental design is called a Latin Square. For example, if the 5 friends are called A,B,C,D, and E, then a possible experimental plan would be

	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5
Soil 1	A	B	C	D	E
Soil 2	B	C	D	E	A
Soil 3	C	D	E	A	B
Soil 4	D	E	A	B	C
Soil 5	E	A	B	C	D

The reason why this is a good experimental plan is that any effect of the different ways different people take care of the plants will be evened out on both the effect of the seed types and the effect of the soil type. The technique is called *blocking*.

4. (a) The usual test statistic in this case is

$$\frac{s_B^2}{s_A^2} = \frac{18.8}{5.9} = 3.186$$

and it should be compared with an F distribution with 11 and 9 degrees of freedom. From the provided tables, we can see that $F_{0.05,11,9} < 3.186 < F_{0.025,11,9}$. As we should make a 2-sided test, this means that the p-value is between 0.1 and 0.05.

(b) The test is based on the values being random samples from normal distributions. To investigate whether for example the values from population A come from a normal distribution, Albert can make a normal probability plot: If the points fall approximately on a line, it indicates that the values may come from a normal distribution.

(c) If it seems unreasonable to assume that both data sets come from normal popoulations, Albert could use for example a permutation test: The difference D between the means of the two sets of values could be compared with a sample of differences between the mean of randomly selected 10 values among the 22 values and the mean of the remaining 12 values. If D was quite large or quite small compared to the sample of such differences, one could reject the null hypothesis that the values come from the same distribution. Alternatively, Albert could use a non-parametric test, such as the Mann-Whitney test (also called the Wilcoxon Rank-sum test).

5. (a) The sum of squares for the colors can be computed as

$$\mathrm{SS}_{color} = 4\left((31-34)^2 + (36-34)^2 + (33-34)^2 + (30-34)^2 + (40-34)^2\right) = 264$$

The sum of squares for the fonts can be computed as

$$\mathrm{SS}_{font} = 5\left((32-34)^2 + (37-34)^2 + (33-34)^2 + (34-34)^2\right) = 70$$

The total sum of squares can most easily be computed by comparing its definition to the definition of the sample variance: We then see that

$$\mathrm{SS}_{total} = (5 \cdot 4 - 1)s^2 = 19 \cdot 29.47386 = 560$$

We can then fill out the ANOVA table as follows:

Source var.	Sum of squ.	D.f.	Mean square	F value	p-value
Color	264	4	66	3.50	$0.025 < p < 0.05$
Font	70	3	23.33	1.24	$p < 0.25$
Error	226	12	18.83		
Sum	560	19			

(b) The assumptions are that the effects of the fonts and the colors are additive, and that the random variation at each combination of font and color is normally distributed, with the same variance at each combination. (Note: As the data in this case consists of counts, the data can never be exactly normally distributed. However, the approximation to a normal distribution may be close enough for the results of the analysis to be useful.)

6. (a) With the probability of success p and thus the probability of failure $1 - p$, the probability of observing the exact sequence shown is

$$p \cdot (1-p) \cdot (1-p) \cdot p \cdot p \cdot (1-p) \cdot p \cdot (1-p) \cdot p \cdot p = p^6(1-p)^4.$$

(b) As Alex will stop his experiments when he has reached 6 successful experiments, his last experiment will necessarily be a success. If he has performed 4 unsuccessful experiments before this, and 5 successful ones, it would not matter among these 9 experiments in which order the successes and failures came. Thus the probability for these 9 first experiments is given by the Binomial formula, while the probability for the last experiment is p. In summary, the probability we are asked compute is

$$\binom{9}{5} p^5(1-p)^4 \cdot p = \frac{9!}{5!4!} p^5(1-p)^4 \cdot p = 126p^6(1-p)^4.$$

Because of the unclear language of the question, the answer

$$\binom{9}{5} p^5(1-p)^4$$

was also accepted.

(c) Generalizing the thinking above, we get that the probability that Alex will have to endure k failed experiments before he has finished r successful experiments is

$$\binom{r+k-1}{r-1} p^r(1-p)^k = \frac{(r+k-1)!}{(r-1)!k!} p^r(1-p)^k.$$

7. (a) False. It is the distribution of the *mean* of a sample that becomes more and more normally distributed (for samples from distributions where the variance exists).

(b) True. As the uniform distribution has a variance, the central limit theorem applies to it.

(c) True. As long as the distributions from which the samples are drawn have a variance, the central limit theorem will apply to the means that appear in the t-statistic, making the whole t-statistic approximately normally distributed for large samples. Note also that as the number of degrees of freedom increases, the t distribution approaches the standard normal distribution.

8. The formula for the width of his confidence interval is

$$2t_{\alpha/2,n-1}s\sqrt{1/n}$$

where n is the number of observations, s is the sample standard deviation, and $1-\alpha$ is the confidence level. Looking at the table for the t-distribution, we can see that, approximately, $t_{\alpha/2,49} \approx t_{\alpha/2,199}$ for relevant α. We also have that s is approximately equal to the standard deviation of the population of possible measurements, whether he is using 50 or 200 measurments. Thus the main difference in the computation of the length of the confidence interval will be that in the case of 50 measurements it will contain the factor $\sqrt{1/50}$, while in the case of 200 measurements, it will contain the factor $\sqrt{1/200} = (\sqrt{1/50})/2$. Thus the length will be approximately half the previous length, i.e., $230/2 = 115$ kronor.