Experimental Design

and

Statistical Analyses of Field Experiments

by

Kåre Ringlund

Agricultural University of Norway

Department of Horticulture and Crop Sciences

Preface

This compendium is based on a Norwegian compendium in Experimental Design by Øivind Nissen and Kåre Ringlund. It contains a summary of important topics of Mathematical Statistics, a discussion of data from experiments and surveys, definitions and explanations of the terms treatments, factors and experimental units, and a discussion of different experimental designs and of series of experiments. A set of student exercises is also included.

The compendium is written with the objective of explaining basic principles for planning and executing field experiments and for analyzing and drawing conclusions from experimental data. For more extensive treatments of the mathematical basis for different designs and analytical methods, the readers are referred to more comprehensive textbooks.

Agricultural University of Norway, Ås, October 1995

Kåre Ringlund

CONTENT

I. INTRODUCTION TO MATHEMATICAL STATISTICS

Experimental design, the science of designing and analyzing experiments, requires knowledge of many subject areas. The analytical component is applied mathematical statistics, and an introduction to statistics will be given by discussing a practical example.

Let us assume that we have two storage bins of potatoes, bin A and bin B. Someone asks which bin has the biggest potatoes. The first problem we are faced with is to define the question. What does biggest mean? Does it mean weight or length or maybe width of the potato tubers? Let us assume that the length of the tubers is the best characteristics of size, and that the question would be best answered by determining the average length of the tubers in the two bins. The term "average" is used in the commonly understood meaning of measuring all the tubers in each bin and dividing the sum of the lengths with the number of tubers. This is called the arithmetic average or the arithmetic mean. There are other ways of calculating "means", for instance the geometric mean, that for certain types of data are "better" than the arithmetic mean.

In our example it is technically possible to determine the mean of tuber lengths in each bin by measuring all tubers. Except for measuring, counting and calculation errors, the means obtained would then be absolutely correct. For most types of measurements it is not technically feasible to measure the whole "bin". We have to draw our conclusions on the basis of a sample.

1. Sampling

By measuring only a sample from each bin an error or uncertainty is introduced. Since all tubers vary in length, measurements of a sample will never give exactly the same results as measurements of the whole bin. Mathematical statistics helps us make predictions about the whole bin from measurements of a sample from the bin.

The first condition for predicting anything from the measurements of a sample is that the sample is drawn randomly from the bins. All tubers in the bin must have an equal chance of being drawn. This is very difficult to achieve, but it is evident that if the sampler selected mainly small or mainly large potatoes, the results from the samples would not be representative of the bin. If we deliberately selected both small and large tubers, our estimate of the means could be reasonable correct, but we would overestimate the variability. If we selected only average size tubers, we could also obtain a reasonable estimate of the mean, but would underestimate the variability.

2. Frequency distributions

Assume that we had measured the length of 20 randomly selected tubers from bin A, and 100 tubers from bin B, and recorded the results as lengths in cm. For the comparison we can count the number of tubers in each length category, and record these numbers in a frequency table (Table I.1).

Table I.1. Frequency distribution; number of tubers for different classes of tuber lengths in bins A and B.

	Tuber lengths in cm												Total		
Bin	5	6	7	8	9	10	11	12	13	14	15	16	Number	Sum	Mean
A	2	6	3	3	2	2	1	1					20	152	7,60
B	4	3	8	20	18	17	15	10	3	0	1	1	100	941	9,41

The example is chosen to show that the samples do not need to be equally large, but it will be shown later that the comparison would have been more accurate by selecting 60 tubers from each bin. The frequency distributions can also be presented as histograms (bar graphs) as shown in Figure I.1.

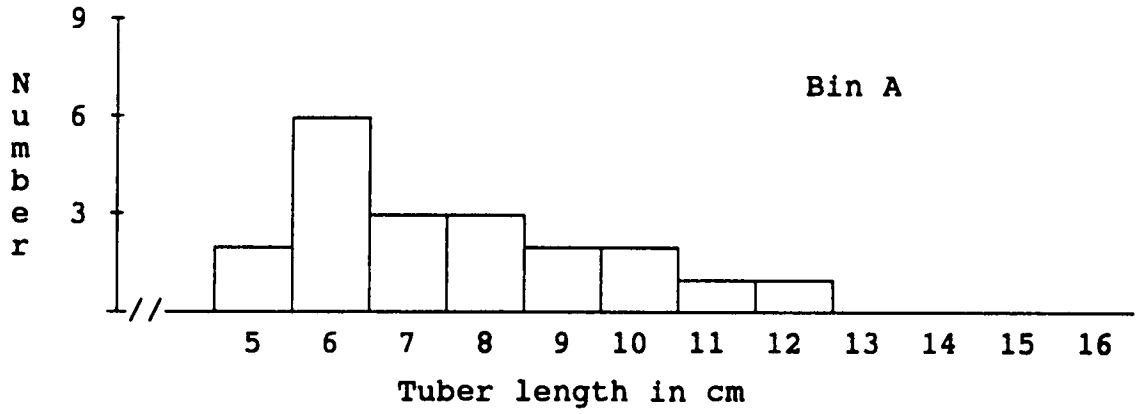
Instead of presenting data as observed, they can be presented as relative figures as in Figure I.1.c. The advantage of relative figures is that they are easier to compare, both in a table and in graphic form. Since the sample from bin B contained exactly 100 tubers, it is necessary to convert only the data from bin A to percentages. The ruggedness of the histograms would be less if the number of observations increased. It would also be possible to reduce the width of each class to reduce the distance between the bars in the histogram. The mathematical limit for this, when the sample size increases towards infinity and the class width decreases towards zero, is a distribution curve.

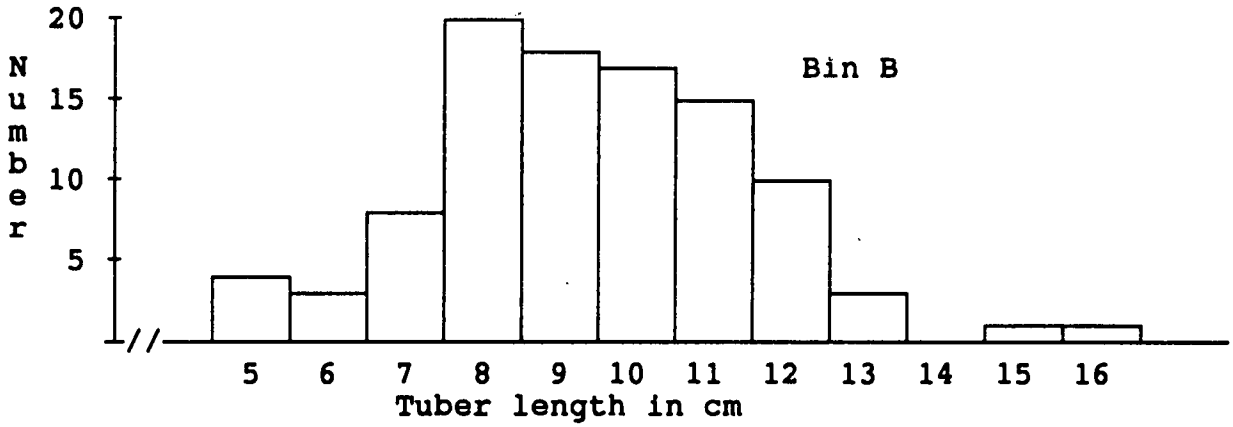
The histograms presented in Figure I.1. are characteristic for most distributions of biological data with one peak near the middle and reduced values on each side. Sometimes we will find distributions with two or more peaks. This normally means that our data contain measurements from two or more different "populations".

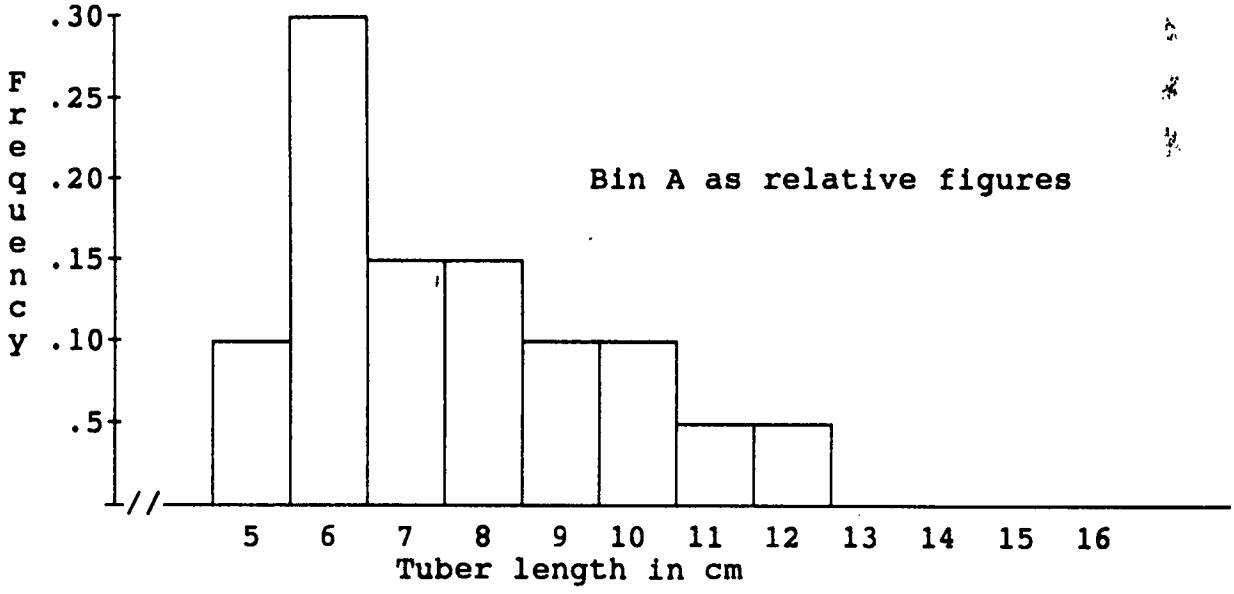
The frequency distributions can be fairly symmetrical, as in our potato example, or more or less skewed. If we had determined the weight of the tubers, the distribution would certainly have been skewed. Most tubers would weigh between 100 and 150 grams. None would weigh below

0 grams, and on the other side some tubers would probably weigh more than 500 grams.

Figure I.1 Histogram for frequency distributions of tuber lengths in potatoes from 2 bins.







A one-topped and fairly symmetrical frequency distribution can be characterized by two parameters, one parameter describing where on the scale most of the observations are found, and one parameter describing the width of the distribution. The arithmetic mean can be used as the first parameter. To describe the width we could use the difference between the largest and the smallest measurement. In our example this would give 7 cm for bin A and 11 cm for bin B. The disadvantages with such a parameter are firstly that it is based only on two measurements, the highest and the lowest, and secondly that it would increase with increased sample size. As we measure more tubers, the chance of including a very big one increases. For several reasons we use a parameter called the Standard Deviation (SD) as a measure for variation or width of a frequency distribution.

3. Calculation of the mean

For bin A the sum is:

2 x 5 + 6 x 6 + 3 x 7 ...= 152, and the mean = 152 / 20 = 7,60

Equally for bin B:

4 x 5 + 3 x 6 + 8 x 7 ...= 941, and the mean = 941 / 100 = 9,41

We describe the individual measurements as x, and the mean as $\bar{x}$ (x-bar) and the formula for the mean becomes:

$$\bar{x} = \Sigma x / N$$, where N is the number of observations.

For a frequency distribution the formula becomes:

$$\bar{x} = \Sigma x_i n_i / N$$

where n_i is the number of observations for each class (x_i) in the frequency table, $\Sigma n_i = N$.

We now have an estimate of the mean tuber length in bin A of 7.60 cm and in bin B of 9.41 cm. Our preliminary answer would therefore be that bin B has the longest tubers. How certain can we be that this result would be close to the result based on all the tubers in the bins? This is the question that mathematical statistics can answer for us.

4. Calculation of the standard deviation

The standard deviation (SD) is calculated as follows:

$$SD = \sqrt{\Sigma(x-\bar{x})^2/(N-1)}$$

The differences between the individual observations and the arithmetic mean are squared and added, and the Sum of Squares (SS) is divided by the number of observations minus 1. N-1 is called degrees of freedom (DF). The quotient SS / DF is called mean square (MS), and SD is the square root of this quotient. SD does not change systematically with the size of the sample.

The formula above is a definition formula. By simple algebra we can deduct a simpler calculation formula for SS.

$$SS = \Sigma(x-\bar{x})^2 = \Sigma(x^2 - 2x\bar{x} + \bar{x}^2)$$

$$= \Sigma x^2 - \frac{2(\Sigma x)(\Sigma x)}{N} + \frac{N(\Sigma x)^2}{N^2}$$

$$= \Sigma x^2 - 2\frac{(\Sigma x)^2}{N} + \frac{(\Sigma x)^2}{N} = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$$

The component $(\Sigma x)^2 / N$ is called the correction term (CT), and as we will see later, the use of this formula greatly simplifies the computation of analyses of variance. In our example we find:

Bin A: Σx^2=1232, CT=1155,20, SS= 76,80, MS=4,04, SD=2,01
Bin B: Σx^2=9281, CT=8854,81, SS= 426,19, MS=4,30, SD=2,07

5. True values for means and standard deviations

If we had measured all the tubers in the bins, we would have found what we could call the true values of the means and the SDs. Such true values are described by the greek letters μ and σ. The values of $\bar{x}$ and SD, calculated from our samples are estimates of these unknown true values. From these estimates we can make predictions about μ.

Mean and standard deviation of a mean

What would the estimates of means and SD have been if we, instead of sampling one tuber at a time, had sampled 16 tubers at a time and calculated our means and SDs on the basis of the means and SDs of these sub-samples? The means for the two bins would not have been systematically

changed, but the estimates of the SDs would have decreased. It can be shown that the new SDs are no longer estimates of σ but of $\sigma/\sqrt{n}$, where n is the number of observations in each sub-sample. With 16 tubers in each sub-sample the SD on the means of the sub-samples would be 1/4 of the SD calculated on the samples of single tubers. The SD of a mean is called Standard Error (SE).

$$SE = SD/\sqrt{n}$$

6. The normal distribution

If we draw histograms of the means of sub-samples, the highest concentration of observations will be on the same point on the scale as for the original observations, but the histograms will be much narrower. Another important aspect of the histograms of the means of sub-samples is that they are more symmetrical. By increasing n the histograms would approach a certain mathematical form called the normal distribution (Figure I.2).

The term "normal" does not mean that these distributions are common in nature. However, the distribution of averages of a number of observations from distributions that deviates strongly from the normal, approach the normal distribution. This is the reason why we can use formulas derived from the normal distribution for evaluations of means and differences between means from empirical distributions, even if these are far from normal.

The normal distribution has the following mathematical form:

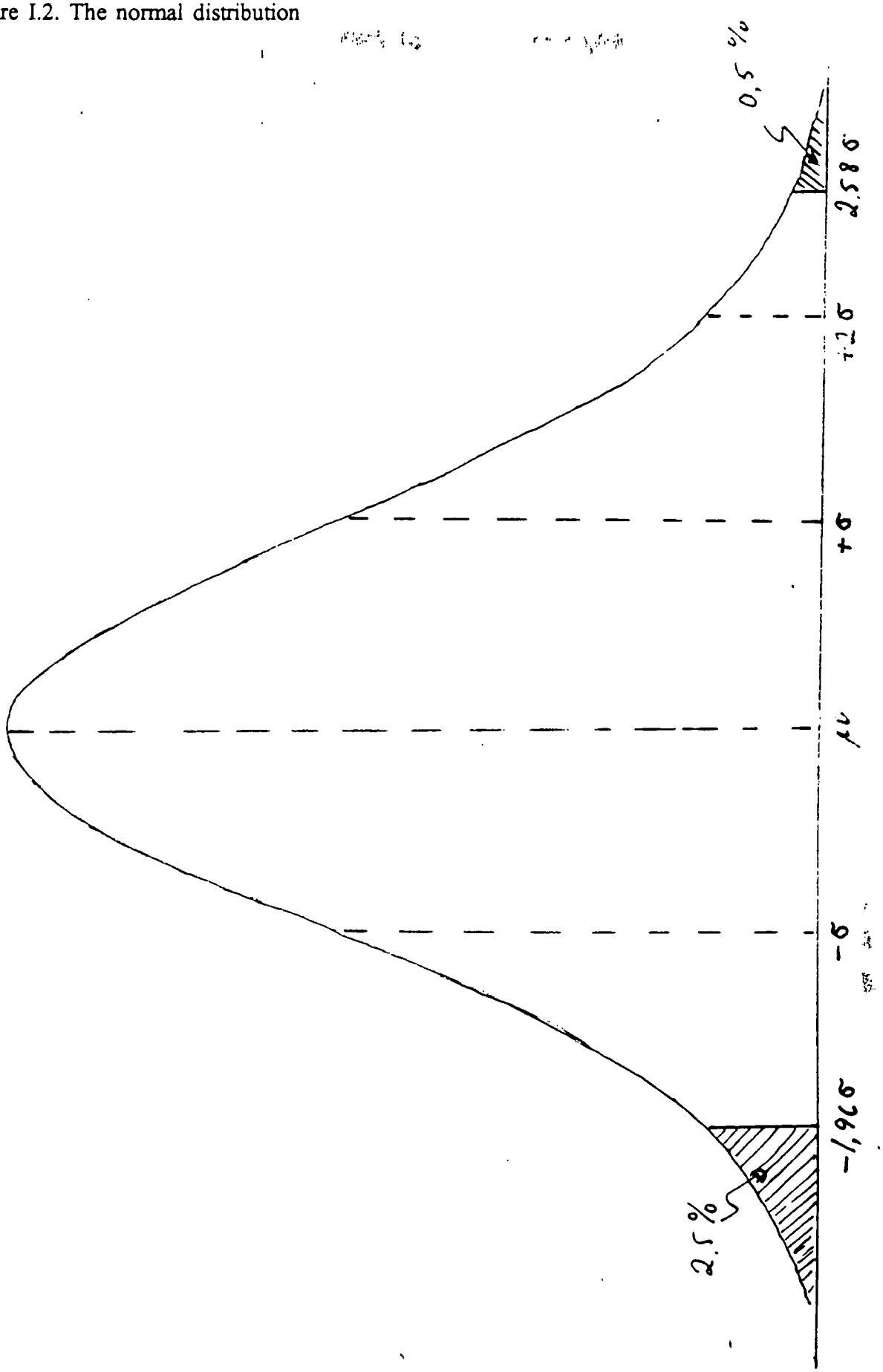
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

If we set $\mu = 0$, the change from concave to convex form occurs at $+\sigma$ and $-\sigma$. Between $+\sigma$ and $-\sigma$ we will find 68% of the area under the curve. Outside $+1.96\sigma$ and -1.96σ we will find 5% (2,5% on each side), and outside $\pm 2.58\sigma$ we will find only 1% of the area under the curve.

7. Samples from a known normal distribution

If we have a known normal distribution, i.e. we know μ and σ, we know the probability for the mean of a sample from this distribution to fall outside certain limits. We know that the probability for the mean of a sample to deviate more than $1,96\sigma$ from μ is only 5%. If we have an unknown sample, we can use this information to evaluate if the sample could have been drawn from the known normal distribution.

Figure I.2. The normal distribution



Numeric example

We have a normal distribution with the parameters $\mu = 100$ and $\sigma = 10$. The question is if an individual observation of 111 could have been drawn from this distribution. Since the deviation from μ is only slightly above σ, the answer is yes. It is about 30% probability that a single observation deviates that much or more from the mean of the distribution.

If the value had been 70, the deviation from the mean would have been 3σ. There is still a slight possibility that the individual was drawn from the known distribution, but the probability is below 1%, and we would conclude that the observation is most probably not from the known distribution.

As a third example, let us assume that we have a sample consisting of 16 individuals with a mean of 108. Could this sample have been randomly drawn from a distribution with $\mu = 100$ and $\sigma = 10$? In this case we can be fairly certain that the answer is no. For a mean of 16 individuals the σ would be $10/4 = 2.5$. The deviation from μ is 3.2σ and the probability of drawing such a sample from the known distribution is less than 0.1%.

Statistically we express this as: "the sample deviates significantly from the given distribution", and to be more precise: "the deviation is significant at the 0.1% level". It is customary to say that the deviation is significant when the probability (P) is between 5% and 1%, and very significant when $P \leq 1\%$.

8. Samples from an unknown normal distribution, t-test

For practical examples we do not know σ, and have to use the estimate SD. This introduces an uncertainty, and, consequently, we have to increase the coefficients 1.96 and 2.58 as limits for the 5% and 1% probabilities respectively. The uncertainty increases with the decrease in number of observations or the degrees of freedom for SD. The distribution function

$$t = (\bar{x}-\mu)/SE$$

is identical with the normal distribution when $DF = \infty$, and the coefficients for the different probability levels increase with decreasing DF (Table I.2).

Table I.2. t-values for different degrees of freedom.

	degrees of freedom (DF)										
P	2	4	8	15	20	25	30	40	60	120	∞
10%	2,92	2,13	1,86	1,75	1,73	1,71	1,70	1,68	1,67	1,66	1,65
5%	4,30	2,78	2,31	2,13	2,09	2,06	2,04	2,02	2,00	1,98	1,96
1%	9,93	4,60	3,35	2,90	2,84	2,79	2,75	2,70	2,66	2,62	2,58
.1%	31,60	8,61	5,04	4,07	3,85	3,73	3,65	3,55	3,46	3,37	3,20

9. Confidence intervals

Let us now turn the argument around. If we, from an unknown distribution, take a sample of n observations (DF = n-1) and determine $\bar{x}$, SD and SE, we can give confidence limits for μ. We start with the t-value for P = 5% which is the numerical value of t where 2.5% of the area under the curve is outside -t and +t. The probability of finding a t-value between $\pm t_{5\%}$ is 95%.

$$-t_{5\%} \leq (\bar{x} - \mu) / SE \leq t_{5\%}$$

By solving these two inequalities we find that:

$$(\bar{x} - \mu) / SE \leq t_{5\%}$$

$$(\bar{x} - \mu) \leq SEt_{5\%}$$

$$(\bar{x} - SEt_{5\%}) \leq \mu$$

and that

$$(\bar{x} - \mu) / SE \geq -t_{5\%}$$

$$(\bar{x} - \mu) \geq SE(-t_{5\%})$$

$$(\bar{x} + SEt_{5\%}) \geq \mu$$

or that

$$(\bar{x} - SEt_{5\%}) \leq \mu \leq (\bar{x} + SEt_{5\%})$$

$(\bar{x} - SEt_{5\%}$ and $(\bar{x} + SEt_{5\%})$ are the confidence limits for μ.

Let us now turn back to the potato example. The results from the measurements and calculations of confidence limits are given in table I.3.

Table I.3. Means, variances and confidence intervals (c.i.) for tuber length in two bins of potatoes.

Bin	n	$\bar{x}$	SS	SD	SE	$t_{5\%}$	95% c.i.	$t_{5\%}$	99% c.i.
A	20	7,60	76,8	2,01	0,450	2,09	6,66-8,54	2,84	6,32-8,88
B	100	9,41	426,2	2,07	0,207	1,99	9,00-9,82	2,63	8,87-9,95

The 95% confidence interval is the interval between the lower 95% confidence limit to the upper 95% confidence limit and similarly for the 99% interval.

It is **not** correct to say that it is 95% probability that μ for bin A is to be found between 6.66 and 8.54. μ is not a random variable, and it has, with 100% probability, a fixed value. It is 95% probability that our statement is correct when we postulate that μ is to be found within the confidence limits. This means that if we calculate a large number of 95% confidence intervals, we will make a correct prediction about μ in 95% of the cases, and we will make a wrong prediction in 5% of the cases.

From Table I.3 we see that the 99% confidence intervals for bins A and B barely overlap. This means that it is **not** reasonable to assume that μ_A and μ_B are equal. Let us assume that this common μ = 8.875. The probability of finding the two means 7.60 or less and 9.41 or larger is then only 1%. To find two such unlikely means simultaneously is even less probable.

10. The SE on a difference

We will now calculate the standard error on a difference (SE_d) between two means and see how this can help us determine if the two means are indeed different.

Let us first assume that we measure the length of two randomly selected tubers from the same bin and record the difference between tuber no. 1 and tuber no. 2. This is then repeated several times, and we can set up a frequency table and calculate $\bar{x}$ and SD for these differences. If the tubers were really randomly selected, the differences will have an equal chance of being positive and negative. The frequency distribution will be symmetrical with an expected mean of 0.

The standard error of the difference will be larger than for the measurements of the individual tubers. On the average the MS for the difference will be twice that of the individual measurements. Subsequently $SE_d = SE\sqrt{2}$.

What will be the result if we draw two samples from the same bin, one sample of n_1 (e.g. 10 tubers) and the other with n_2 (e.g. 20 tubers)? We calculate the means of the two samples and the difference between these two means, and repeat this exercise a number of times. Also in this case the mean of the differences should approach zero, and the MS for the differences approaches $\sigma^2/n_1 + \sigma^2/n_2$ where σ is the SD for the single observations.

The best estimate of a common MS for comparing samples from the two different bins is to add SS from the two samples and divide this sum by the sum of DF from bins A and B.

$$MS = (SS_1 + SS_2) / (DF_1 + DF_2) = (76.8 + 426.2) / (19 + 99) = 4.26$$

The difference between the mean tuber length in the two bins is

$$d = \bar{x}_1 - \bar{x}_2 = 9.41 - 7.60 = 1.81$$

and the standard error of the difference SE_d becomes

$$SE_d = \sqrt{MS/n_1 + MS/n_2} = \sqrt{4.26/20 + 4.26/100} = 0{,}50$$

11. Testing differences

To test if a difference is different from 0, we test the hypothesis that $\mu = 0$. If this hypothesis is correct, the t-value becomes

$$t = (d-\mu) / SE_d = d / SE_d$$

For the potato example $d = 1.81$, and $SE_d = 0.50$, and

$$t = 1.81 / 0.50 = 3.62$$

The degrees of freedom for this test is $DF_1 + DF_2 = 19 + 99 = 118$, and in the t-table we find that the probability of finding a $t \geq 3.62$ is less than 0.1%. We therefore reject the hypothesis $\mu = 0$, and accept the alternative that bin B has longer tubers than bin A.

Using the t-value for DF = 120 and P = 0.1% (3.29), and our estimate of $SE_d = 0.50$, the 99% confidence limits are 0.16 and 3.46. The interval does not contain 0, which is in accordance with the result from the t-test.

Independent samples

In the potato example we determined the difference in tuber lengths by first determining the mean

length in each bin and then calculating the difference between these two means. We could have determined the mean difference by measuring a randomly selected tuber from each bin, calculating the difference between these two tubers, and repeating this a number of times. The expected mean and SE_d are equal for both methods given the same sample size. In the latter case we determine MS on the difference directly, whereas in the first case the MS for the difference is computed from the MS of each of the two bins.

Estimating the mean difference from a random sample from each bin gives the most reliable estimate of the SE_d. This is related to the degrees of freedom. If we measure 50 tubers from each bin, the DF for MS_d is 49 + 49 = 98, whereas if we measure 50 differences, the DF = 49. A disadvantage for the latter method is that we have to take equal size samples from each bin. If we do not have real pairs, there is no advantage of determining the SE_d on individual differences.

It is most efficient to measure the same number of individuals in both samples. We will obtain the most accurate estimate of the SE_d when the two samples are of equal size. This is easy to show with a numeric example. We had measured 20 tubers in bin A and 100 tubers in bin B, and SE_d is:

$$SE_d = \sqrt{MS/20 + MS/100} = \sqrt{0.06\ MS} = 0.24\ SD$$

If we had measured 10 tubers from bin A and 110 tubers from bin B, the result would have been:

$$SE_d = \sqrt{MS/10 + MS/110} = \sqrt{0.11\ MS} = 0.33\ SD$$

which is considerably higher. We would have obtained the least SE_d by having the same number of observations in each sample:

$$SE_d = \sqrt{MS/60 + MS/60} = \sqrt{0.033\ MS} = 0.18\ SD$$

Equal sample size also has other advantages.

Paired samples

Let us turn to another example. People are in general a little taller in the morning than at night. The body is compressed during the day. To determine this difference, it is obvious that a sample of 100 individuals measured in the morning and another 100 individuals measured at night would

not be very accurate. It is obvious that we would obtain a much more accurate estimate of this difference by measuring the same individuals morning and night. What is actually the difference between this example and the potato example?

In the potato example there is no connection between tuber number 1 in bin A and tuber number 1 in bin B. The two measurements are independent of each other. The heights of the same person morning and night, however, are strongly related or correlated as it is called in statistical terms. It is only for un correlated measurements that MS_d is the sum of MS_1 and MS_2. If the measurements are correlated, the SE_d will be smaller. If we have observations that are natural pairs, the mean difference and SE_d must be determined on the measurements of individual differences.

12. Universe, samples and hypotheses

In mathematical statistics we use the term "universe" on a large number of individuals or observations. A universe can be limited, as in the potato example, or unlimited, as for example the yield of a certain crop variety. This is unlimited because we are both interested in the yields that have been found in the past as in future yields. For both types of universe, we are interested in estimating properties of the universe through random samples.

We have seen how we can estimate means and standard deviations of a universe from a random sample, and how we can determine confidence limits for the expected mean (μ). The means of two groups can be compared by a t-test, and by computing confidence limits for the expected difference.

Null hypothesis

The hypotheses that we test by t-tests, and by other significance tests that we will discuss later, are called null hypotheses. In the potato example the null hypothesis is that the tubers in the two bins are of equal length, or to be more exact: that the normal distributions of tuber length are identical in the two bins.

We can never prove a null hypothesis to be correct. A significance test can only give us the probability for the hypothesis to be right, and when this probability is low, for instance below 5%, we reject the hypothesis.

13. One way analysis of variance

By use of a t-test we can compare two different means or test one difference. By the use of an analysis of variance (ANOVA), we can compare several means at the same time and test if there are any differences between these means. We test the hypothesis that all the means are equal.

We will show the calculation of an ANOVA first, and later come back to some of the theory behind the analysis. Assume that we have asked 27 farmers which variety of potatoes they grew last year, and what yields they obtained. The answers are given in table I.4.

Table I.4. Yields of potatoes in tons per hectare.

Variety													Sum	n	$\bar{x}$
A	30	37	21	23	28	35	31	24	27	24	26	23	329	12	27.4
B	36	39	31	33	35	32	40	28	30	31			335	10	33.5
C	32	35	28	29	30								154	5	30.8
Total													818	27	30.3

We begin by calculating the correction term (CT) and the total sum of squares (SS_{total}):

$$CT = 818^2/27 = 24782$$

$$SS_{total} = 30^2+37^2+.....-CT = 25430 - 24782 = 648$$

The SS_{total}, 648, can be divided into two parts, the variation between varieties and the variation within varieties. The variation within varieties is an estimate of the SS_{error}.

$SS_{var} = 329^2 / 12 + 335^2 / 10 + 154^2 / 5 - CT = 9020 + 11222 + 4743 - 24782 = 203.$

$SS_{error} = SS_{total} - SS_{var} = 648 - 203 = 445$

For control we can also calculate the SS_{error} as the sum of the individual within variety SS, $SS_A = 275$, $SS_B = 139$ and $SS_C = 31$. These add up to 445.

The total degrees of freedom, 27-1 = 26, can also be divided into two parts, $DF_{var} = 2$ and $DF_{error} = 24$. The latter DF is the sum of DF for the individual varieties, 11 + 9 + 4 = 24.

Finally we compute MS for each component and the ANOVA becomes:

Analysis of variance:

Source of variation	DF	SS	MS	F
Total	26	648	24.9	
Between varieties	2	203	101.5	5,49
Within varieties = error	24	445	18.5	

14. F-test

F is defined as the ratio between the largest and the smallest mean square. In an ANOVA the F is in general $MS_{treatment} / MS_{error}$. In the above example $F = MS_{var} / MS_{error}$. The F-value can, analogous to the t-value, be used to test hypotheses about the varietal means. It can be shown that if there are no differences between the varietal means (and standard deviations), i.e. all the observations are drawn from the same distribution, the three mean squares are estimates of the standard deviation of this distribution. Given that this distribution is a normal distribution, a table of F-values for different probabilities has been worked out. These F-tables have three entry parameters, namely P, DF for the MS in the numerator and DF for the MS of the denominator.

In the F-table for P = 5% and DF = 2, 24 we find the value 3.40. This means that if we on data from a normal distribution had determined two independent standard deviations, it is 5% probability to find an F-value of 3.40 or above. The corresponding F-value for P = 1% is 5.61, and the F-value for P = 0.1% is 6.95. In our example the F-value was 5.49, and the probability of obtaining such a high value is only slightly above 1% if the yield data were drawn from the same normal distribution.

Our null-hypothesis is that the three varieties give the same yield. When the F-value exceeds the value for P = 5% we reject the hypothesis and accept the alternative that the three varieties gave different yields. The actual F-value shows that the yield differences are significant at the 1% level, or very significant.

We know á priori that the 0-hypothesis is wrong. We know that the varieties are different, and we are not interested in only the question of whether the varieties are different. Our main interest is to find out which variety yields the most, how big the differences are, and to determine confidence limits for the differences.

The null-hypothesis and the F-test are technical aids. If we do not find a significant F-value, the interpretation can be that the differences are small and without practical importance, or that our

experimental error is too large to detect eventual differences.

15. Conditions for testing of differences

Normal distribution

If we find a large F-value, and consequently a low P-value, we know that the probability of obtaining such a result is very low if the samples were drawn at random from the same normal distribution. An obvious question to ask is: should we expect to find such results if the samples were drawn from a non normal distribution, e.g. a very skewed distribution or a distribution with two or more peaks? Several investigations have shown that the significance level of the F-test is not very much influenced by the shape of the original distribution. We must realize that the probability-limits of 5%, 1% and 0.1% are arbitrarily chosen and not any sort of magical limits.

Equal random error

Another condition for testing of differences is that the observations have the same random error. If some of the observations have a much larger error than others, the significance level of the test can be considerably influenced.

In some instances we know á priori that different treatments have different variability. One example is the comparison of for instance 4 herbicides with no treatment at all. If the herbicides are approximately equal, and there are lots of weeds in the field, the variation within the non treated plots will be much larger then within the treated plots. The error MS in the analysis of variance is, therefore, too large for the comparison between the herbicides and too small for the comparison between the non treated plots and the ones with herbicides.

16. Means, and error of differences

The analysis of variance told us that there were differences between the three potato varieties. It did not say anything about which variety was best or which of the three were significantly different. To answer these questions we have to look at the means, the differences between the means and the error on these differences. In our example the two varieties A and B had the highest and the lowest mean. The significant F-value in the analysis of variance shows that at least these two varieties are different. The analysis of variance does not indicate whether variety C is different from A or B or both of them.

We use MS_{error} as an estimate of σ^2, and assume that this is a valid common estimate for all

the three varieties. In the example we can calculate three differences and a standard error and a t-test for each of these three (Table I.5).

Table I.5. Yield differences and standard errors for three potato varieties.

Variety	difference	$m_d = \sqrt{MS_{error}/n_1 + MS_{error}/n_2}$	t
B-A	0,61	0,18	3,4
C-A	0,34	0,23	1,5
B-C	0,27	0,24	1,1

It is only the difference between A and B that is statistically significant. The variety C is neither significant from A nor from B.

17. Two way ANOVA

Let us assume that part of the data discussed above was collected from 5 different farms on which all three varieties were grown. The results are shown in table I.6.

Table I.6. Yield data in tons/ha from three potato varieties grown at five different farms.

	Variety				
Farm	A	B	C	Sum	$\bar{x}$
1	30	36	32	98	32.7
2	37	39	35	111	37.0
3	21	31	28	80	26.7
4	23	33	29	85	28.3
5	28	35	30	93	31.0
Sum	139	174	154	467	
$\bar{x}$	27.8	34.8	30.8		31.1

For a two way analysis of variance we start with calculation of the correction term and the total sum square as in the previous example.

$CT = 467^2 / 15 = 14\ 539$

$SS_{total} = 30^2 + 37^2 + ... - CT = 14\ 889 - 14\ 539 = 350$

This total sum square and the total DF of 14 can be divided into three components, a between farm component with 4 DF, a between varieties component with 2 DF and an error term with 8 DF. The SS between farms is calculated on the basis of the total yields for each farm.

$SS_{farm} = (98^2 + 111^2 + 80^2 + 85^2 + 93^2)/3 - CT = 194$

The SS between varieties is calculated on the basis of the total yields for varieties,

$SS_{var} = (139^2 + 174^2 + 154^2)/5 - CT = 124$

which is equivalent to the calculation of SS between varieties in the one way analysis. Since we have the same number of observations for all three varieties, we can add the squares first and divide by the number of observations at the end. We save work, and we reduce the rounding error. The results are set up in a table and the SS_{error} is calculated by subtraction. The variances are calculated by dividing the sum squares by the degrees of freedom, MS = SS / DF.

Analysis of variance:

Source of variation	DF	SS	MS	F
Total	14	350		
Between farms	4	194	48.5	
Between varieties	2	124	62.0	15,5 **
Error (farm x variety)	8	32	4.0	

Finally the F-value for varieties, $F = MS_{var} / MS_{error}$ is calculated and checked against an F-table. The F-value is larger than the value for P = 0.01 which we indicate by two stars (one star for $0.05 > P > 0.01$, two stars for $0.01 > P > 0.001$, and three stars for $P < 0.001$). Our conclusion is that there are very significant differences between the three varieties.

The error is the interaction between farms and varieties. If all the varietal differences were equal on all the five farms, the error variation would be zero. In real experiments the error

variation is never zero. It contains random variation such as soil variation, weighing errors, writing errors, random attacks of diseases and pests etc. There is also a possibility that there is a real interaction between the varieties and the growing conditions on the five farms, that one of the varieties is best adapted to one of the farms and another variety to another farm. Such effects cannot be separated from the error in a two way analysis.

The advantage of the two way ANOVA compared to the one way ANOVA is that the variation between farms can be separated from the error term. The difference between one- and two way ANOVA is comparable to the difference between a t-test on random samples and a t-test on paired samples.

18. t-test or F-test

If we are comparing only two treatments, we can choose between t-test and ANOVA, and both analyses give the same significance test. In the ANOVA we have 1 DF between treatments and $F = t^2$. If we compare an F-table an a t-table we find that for the same significance level (P) the F-values in the column for 1 DF for the MS in the denominator are the squares of the values in the t-table.

19. The Standard error on differences between two means in a two way analysis.

For data that can be analyzed by a two way ANOVA we have the same number of observations for all the treatments. Hence, the standard error is the same for all differences between treatments. in our example the standard error on a difference between two treatments is:

$$SE_d = \sqrt{0{,}040/5 + 0{,}040/5} = 0{,}13 = \sqrt{2MS_{error}/r}$$

where r is the number of replications (in our example the number of farms).

20. Least significant difference

Based on MS_{error} we can calculate how large a difference must be in order to be significant. We call this the least significant difference (LSD), and it is obtained by multiplying the SE_d by the t-value for the appropriate significance level and DF. In our example the DF for MS_{error} is 8, and the t-values are 2.31 and 3.36 for P = 5% and P = 1% respectively. These t-values multiplied by $SE_d = 0.13$ gives:

$LSD_{5\%} = 0.32$, and $LSD_{1\%} = 0.43$

The three differences (Table I.6) were: B - A = 3.48 - 2.78 = 0.70
B - C = 3.48 - 3.08 = 0.40
C - A = 3.08 - 2.78 = 0.30

Based on this analysis the difference between varieties A and B is very significant, the difference between B and C is significant, whereas the difference between A and C is not significant. The LSD is an á priori test which means that is is valid only for questions defined independent of the experimental results.

21. The Coefficient of variation

The coefficient of variation (CV%) is a measure of the quality of an experiment. The coefficient of variation varies with the type of observation and will be different for instance yield, lodging, resistance data etc. The CV% is only useful together with a certain amount of reference data for comparison. Since this value is calculated as a quotient of the mean, it is evident that for data with means around zero, the CV% will be very large and of little value for evaluation of the quality of an experiment.

$$CV\% = \frac{\sqrt{MS_{error}} \cdot 100}{\bar{x}}$$

22. ANOVA for more than two factors

For experiments with more than one factor we can divide the total variation into components for the main effects of factors, interactions between the factors and estimates of error terms for each of the effects.

Let us first look at the organization of a three way table for factors A, B and C with a, b and c levels respectively. We indicate the specific level of a factor by the subscripts $_{ijk}$. In this case $_i$ would have values from 1 to a, $_j$ would have values from 1 to b and $_k$ would have values from 1 to c. A three way table can be organized in three different ways. If for instance a = 4, b = 3 and c = 2, we could arrange the data in a two way table for factors A and B for each level of factor C as shown in table I.7. We could also arrange two way tables over A and C for each value of B,

or two way tables over B and C for each value of A.

Table I.7. Three way table for the factors A, B and C.

	C_1					C_2					
	A_1	A_2	A_3	A_4	$\sum_{i=1}^{4}$	A_1	A_2	A_3	A_4	$\sum_{i=1}^{4}$	$\sum_{i=1}^{4}\sum_{k=1}^{2}$
B_1	X_{111}	X_{211}	X_{311}	X_{411}	$X_{.11}$	X_{112}	X_{212}	X_{312}	X_{412}	$X_{.12}$	$X_{.1.}$
B_2	X_{121}	X_{221}	X_{321}	X_{421}	$X_{.21}$	X_{122}	X_{222}	X_{322}	X_{422}	$X_{.22}$	$X_{.2.}$
B_3	X_{131}	X_{231}	X_{331}	X_{431}	$X_{.31}$	X_{132}	X_{232}	X_{332}	X_{432}	$X_{.32}$	$X_{.3.}$
$\sum_{j=1}^{3}$	$X_{1.1}$	$X_{2.1}$	$X_{3.1}$	$X_{4.1}$	$X_{..1}$	$X_{1.2}$	$X_{2.2}$	$X_{3.2}$	$X_{4.2}$	$X_{..2}$	$X_{...}$

The table value X_{231}, represent the data for level 2 of factor A, level 3 of factor B and level 1 of factor C. Generally we designate a value in the table as X_{ijk}. For sums over one factor the indexes are replaced by a dot (.) as shown in the table. The sums over A represent the two way table for B and C, and the bottom line represent the two way table for A and C. The sums for the different levels of each factor, the one way tables, are represented by two dots, $X_{.j.}$ and $X_{..k}$ for factors B and C respectively. Finally the two way table for factors A and B is obtained by summation over factor C as shown in table I.8. This table also give the sums for factors A and B. $X_{...}$ is the total sum of the three way table.

Table I.8. Two way table for factors A and B.

	A_1	A_2	A_3	A_4	$\sum_{i=1}^{4}\sum_{k=1}^{2}$
B_1	$X_{11.}$	$X_{21.}$	$X_{31.}$	$X_{41.}$	$X_{.1.}$
B_2	$X_{12.}$	$X_{22.}$	$X_{32.}$	$X_{42.}$	$X_{.2.}$
B_3	$X_{13.}$	$X_{23.}$	$X_{33.}$	$X_{43.}$	$X_{.3.}$
$\sum_{j=1}^{3}\sum_{k=1}^{2}$	$X_{1..}$	$X_{2..}$	$X_{3..}$	$X_{4..}$	$X_{...}$

As shown earlier (page 17) the total sum of squares in a two way table can be subdivided into

three components, two main effects and the two-factor interaction. Similarly the total sum of squares in a three way table can be subdivided into seven components, three main effects of factors A, B and C, three two-factor interactions, AB, AC and BC, and a three-factor interaction, ABC. The different sum squares can be calculated as follows:

$$CT = X^2_{...}/abc$$

$$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} X^2_{ijk} - CT$$

$$SS_A = \sum_{i=1}^{a} X^2_{i..}/bc - CT$$

$$SS_B = \sum_{j=1}^{b} X^2_{.j.}/ac - CT$$

$$SS_C = \sum_{k=1}^{c} X^2_{..k}/ab - CT$$

$$SS_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b} X^2_{ij.}/c - CT - SS_A - SS_B$$

$$SS_{AC} = \sum_{i=1}^{a}\sum_{k=1}^{c} X^2_{i.k}/c - CT - SS_A - SS_C$$

$$SS_{BC} = \sum_{j=1}^{b}\sum_{k=1}^{c} X^2_{.jk}/c - CT - SS_B - SS_C$$

$$SS_{ABC} = SS_{error} = SS_{Total} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

Analysis of variance for a three factor experiment

Source of variation	DF	SS
Total	abc - 1	SS_{Total}
A	a - 1	SS_A
B	b - 1	SS_B
C	c - 1	SS_C
AB	(a-1)(b-1)	SS_{AB}
AC	(a-1)(c-1)	SS_{AC}
BC	(b-1)(c-1)	SS_{BC}
ABC	(a-1)(b-1)(c-1)	$SS_{ABC} = SS_{error}$

For experiments with more than three factors the principles for subdividing the total variance

is analogous to that shown for three factors. A four factor experiment gives four main effects, six two-factor interactions, four three-factor interactions and an error term or four-factor interaction, fifteen components all together. For a five factor experiment the total variation can be subdivided into thirty one components.

If the three factor table presented in Tables I.7 and I.8 represent data from a block experiment with factor A being the block or replication, and factors B and C are experimental treatments, we normally accept the three interaction terms with factor A as estimates of the random variation. These can then be bulked to a common estimate of the error. An ANOVA of this data would be as follows:

Analysis of variance

Source of variation	DF
Total	23
Replication (A)	3
Factor B	2
Factor C	1
BC	2
Error	15

The error contains the three interactions AB, AC and ABC with 6, 3 and 6 DF respectively. In some instances the different experimental factors have different error terms. In such instances we cannot bulk these interactions into one common error.

II. MODELS AND VARIANCE COMPONENTS

In this chapter we will discuss models for different designs and the expectations for variance components of different mean squares. We will also discuss the concepts of fixed and random variables.

1. Random distribution of treatments

For the simplest possible experimental design, complete random distribution of treatments, the model for the individual observations is:

$$X_{ij} = \mu + a_i + E_{ij}$$

X_{ij} is the observed value of treatment $_i$ and replication $_j$. The index $_i$ has values from 1 to t where t is the number of treatments, and the index $_j$ has values from 1 to r_i where r_i is the number of replications for each treatment. The model is additive. Each observation is considered being a sum of the mean value (μ), an effect of treatment (a_i) and a random effect E_{ij}. If we define μ as the mean of the treatment's true values, the sum of a_i will be zero. Given a very large number of observations the sum of E_{ij} will also be zero.

Variance components

In an analysis of variance of a complete random distribution design, it can be shown that the variance components estimated are:

Source of variation	DF	MS	Estimated components
Total	$\sum_{i=1}^{t} r_i - 1$	-	
Treatment	t - 1	MS_1	$\sigma^2 + r_0 K^2$
Error	$\sum_{i=1}^{t} (r_i - 1)$	MS_2	σ^2

MS_2 is an estimate of σ^2, and $(MS_1 - MS_2) / r_0$ is an estimate of K^2. σ^2 and $r_0 K^2$ are components of MS_1. The F-test MS_1 / MS_2 is a test for the hypothesis K^2 = zero. If K^2 is different from zero, the effects of treatments, a_i, are also different from zero.

2. Fixed and random variables

In the model for complete random distribution both μ and a_i are fixed variables, and E_{ij} is a random variable. We will explain the difference between fixed and random variables through some practical examples. For fixed variables we are interested in determining the differences between the levels of the variable, whereas for random variables we are interested in the variation between levels. Experimental treatments are normally fixed variables. We are either interested in differences between the levels of the treatment factors, or in determining an optimal level of the factor. In both cases the experimenter has *fixed* the levels of the factor, they are not drawn at random from a universe.

An exception from this general rule is an experiments with different genotypes drawn from a population with the purpose of estimating genetic parameters for the population. In that case we are not interested in determining differences between the genotypes, but in using the variation between genotypes to characterize the population.

Replications, or blocks, is always a random variable. We are not interested in the differences between replications, but in determining the effects of treatments for an area and the replications are random samples of this area. In a series of experiments on different farms, the farms is also a random variable. We are not interested in determining the treatments effects for each farm, but use the farms as random samples of the area where the experiments will be used to give advice to the farmers.

In the models we use capitals for random variables and lower case letters for fixed variables. Since μ is defined as the grand mean of an experiments, the mean of the other fixed variables is zero. A random variable is drawn from a population with a mean of zero. The expected mean will therefore also be zero, but due to the random variation, the mean is not exactly zero.

The frequency distribution of a random variable has a variance that we are interested in estimating. A fixed variable has not such a variance, but we can define a similar value

$$K^2 = \sum_{i=1}^{t} a_i / (t-1)$$

which is a component of the mean squares in the analyses of variance.

3. Randomized block experiments

The model for a randomized block experiment is:

$$X_{ij} = \mu + a_i + B_j + E_{ij}$$

This is also an additive model. B_j is the effect of block $_j$. The means squares of the analysis of variance are estimates of the variance components as shown:

Analysis of variance

Source of variation	DF	MS	Estimated components
Total	tr - 1	–	
Rep.	r - 1	MS_1	$\sigma^2 + t\sigma^2_B$
Treatment	t - 1	MS_2	$\sigma^2 + rK^2_a$
Error	(t-1)(r-1)	MS_3	σ^2

The F-test, MS_2 / MS_3, is a test of the hypothesis: $K^2_a = 0$. If we reject the hypothesis we accept the alternative that K^2_a is different from zero which implies that one or more of the a_i s are different from zero, or, with other words, that there are differences between treatments.

MS_1 / MS_3 is an F-test for differences between blocks or replications. As mentioned above we are normally not interested in this test, but the variance components illustrates the effect of blocks. If $\sigma_B = 0$, there is nothing to be gained by using block experiments instead of complete random distribution, but when $\sigma_B > 0$, the error sum square and mean square will be reduced compared to a complete random design. MS_1, MS_2 and MS_3 are independent estimates and if there are no block or treatment effects, MS_3 may be larger than the other two due to random variation.

4. Latin squares

The model for Latin square is:

$$X_{ijk} = \mu + a_i + R_j + C_k + E_{ijk}$$

As for the other models μ is the expected mean and a_i is the treatment effect. R_j and C_k are

the effects of rows and columns respectively. Since this design requires the same number of rows and columns as the number of treatments, both $_i$, $_j$ and $_k$ runs from 1 to t. The model has two random variables, R for rows and C for columns. The only interesting F-test in the analysis of variance is MS_3 / MS_4 which is a test of the hypothesis that there are no differences between the treatments.

Analysis of variance

Source of variation	DF	MS	Estimated components
Total	$t^2 - 1$		
Rows	t - 1	MS_1	$\sigma^2 + t\sigma^2_R$
Columns	t - 1	MS_2	$\sigma^2 + t\sigma^2_C$
Treatment	t - 1	MS_3	$\sigma^2 + tK^2_a$
Error	(t-1)(t-2)	MS_4	σ^2

5. Factorial designs

In a factorial design the model and the estimated variance components depend on whether the factors are fixed or random. In an experiment with for instance three levels of fertilizer and four varieties the fertilizer factor is always a fixed variable. We are interested in the differences between these particular levels of fertilizer and not in these three levels as random samples of fertilizer treatments in general. The varieties, however, can be either fixed or random. If we are interested in the differences between these four particular varieties, the variable varieties is fixed. If, on the other hand, we are interested in the effect of fertilizer for the species in question and use the four varieties as random samples of this species, varieties is random. In that case the interaction between varieties and fertilizer becomes part of the error term for fertilizer.

Let us assume that we have an experiment with r blocks, each with mn treatments (m fertilizer levels and n varieties), and that varieties is a fixed variable. The model for such an experiment is:

$$X_{ijk} = \mu + B_i + f_j + v_k + (fv)_{jk} + E_{ijk}$$

B is the block effect with $_i$ varying from 1 to r, f is the effect of fertilizer with $_j$ values from

1 to m, and v is the effect of varieties with $_k$ varying from 1 to n. The interaction term $(fv)_{jk}$ represent the deviation from the additive effect of fertilizer and varieties, with other words allows for different fertilizer effect for different varieties. Finally the random variable E_{ijk} represent the random variation or the error for each plot.

If we have an experiment where varieties is a random variable, the model will be almost identical. The only difference is that v is replaced by V. It is only for the estimated variance components that the difference between the two models becomes evident.

Analysis of variance

			Estimated variance components	
Source of variation	DF	MS	Varieties fixed	Varieties random
Total	rmn - 1			
Block	r - 1	MS_1	$\sigma^2 + nm\sigma^2_B$	$\sigma^2 + nm\sigma^2_B$
Fertilizer	m - 1	MS_2	$\sigma^2 + rnK^2_f$	$\sigma^2 + r\sigma^2_{fV} + rnK^2_f$
Varieties	n - 1	MS_3	$\sigma^2 + rmK^2_v$	$\sigma^2 + rm\sigma^2_V$
Fert. x var.	(m-1)(n-1)	MS_4	$\sigma^2 + rK^2_{fv}$	$\sigma^2 + r\sigma^2_{fV}$
Error	(nm-1)(r-1)	MS_5	σ^2	σ^2

If varieties is a random variable, the effect of fertilizer, MS_2, must be tested against the interaction fertilizer x variety, MS_4. If varieties is a fixed variable, the interaction term disappears from the estimated variance components for fertilizer, and the effect of fertilizer can be tested against the common error, MS_5.

Split plot

A special case of factorial experiments is a split plot design. In a split plot experiments there are two different error terms, one for main plots and one for sub-plots. For the same example as above with fertilizer on main plots and varieties (regarded as a fixed variable) on sub-plots the model is:

$$X_{ijk} = \mu + B_i + f_j + ES_{ij} + v_k + (fv)_{jk} + Es_{ijk}$$

ES_{ij} is the error for main plots and Es_{ijk} is the error for sub-plots. The effect of fertilizer should

be tested against error (a), and the effect of varieties and the interaction between fertilizer and varieties against error (b).

Analysis of variance

Source of variation	DF	MS	Estimated components
Total	rmm - 1		
Blocks	r - 1	MS_1	$\sigma^2_{ES} + mn\sigma^2_B$
Fertilizer	m - 1	MS_2	$\sigma^2_{ES} + rmK^2_f$
Error (a)	(r-1)(m-1)	MS_3	σ^2_{ES}
Varieties	n - 1	MS_4	$\sigma^2_{Es} + rmK^2_v$
Fert. x Var.	(m-1)(n-1)	MS_5	$\sigma^2_{Es} + rK^2_{fv}$
Error (b)	m(r-1)(n-1)	MS_6	σ^2_{Es}

6. Series of experiments

The models for experimental series over sites and years are in principle analogous to the models for factorial experiments. However, special attention has to be placed on the definition of the error terms. For an annual series of experiments of one factor (varieties) over different locations and with replications (blocks) within location, the model is:

$$X_{ijk} = \mu + L_i + (BiL)_{ij} + v_k + (Lv)_{ik} + E_{ijk}$$

In this model L_i is the effect of location, $(BiL)_{ij}$ is the effect of blocks within location, v_k is the effect of varieties, and $(Lv)_{ik}$ is the effect of the interaction between varieties and location. E_{ijk} represent the within location error. The index $_i$ runs from 1 to n, the index $_j$ from 1 to r and the index $_k$ from 1 to t. The F-test for the main effect of varieties is MS_1 / MS_2.

Analysis of variance

Source of variation	DF	MS	Estimated components
Varieties	t - 1	MS_1	$\sigma^2 + r\sigma^2_{Lv} + rnK^2_v$
Loc. x var.	(n-1)(t-1)	MS_2	$\sigma^2 + r\sigma^2_{Lv}$
Error within loc.	n(r-1)(t-1)	MS_3	σ^2

In order to give advice for next year, we need to sample different years for our data to include reactions to changes in climate over years. Extending the previous model to include m years, and assuming that we have new locations each year, results in the following model:

$$X_{ijkl} = \mu + Y_i + (LiY)_{ij} + (BiLiY)_{ijk} + v_l + (LiYv)_{ijl} + E_{ijkl}$$

Since there are new locations each year, there is no main effect of locations or interaction between locations and years. The variation between locations is contained in the term locations within years $(LiY)_{ij}$. As reveled by the estimated components, the interaction between year and variety is the error term for the test of main varietal effects.

Analysis of variance

Source of variation	DF	MS	Estimated components
Variety	t-1	MS_1	$\sigma^2 + r\sigma^2_{LiYv} + nr\sigma^2_{Yv} + rmmK^2_v$
Year x Variety	(m-1)(t-1)	MS_2	$\sigma^2 + r\sigma^2_{LiYv} + nr\sigma^2_{Yv}$
Loc.within Yr x var.	m(n-1)(t-1)	MS_3	$\sigma^2 + r\sigma^2_{LiYv}$
Error within loc.	m(n-1)(r-1)(t-1)	MS_4	σ^2

III EXPERIMENTS OR SURVEYS

There are two different ways to collect data for research,

1. Collection of old data through surveys
2. Establishment of new data through experiments

1. Surveys

The survey method is used to collect existing information and to analyze this information for relationships between different variables, and to find differences between previously defined groups. Certain questions are difficult - or sometimes impossible - to analyze through experiments. The survey method is then the only possibility. A survey is normally cheeper than a series of experiments, but the method has some disadvantages in relation to the interpretation of the results.

Let us go back to the example of different potato varieties grown on different farms. We found that variety B gave higher yield than variety A. Is this an acceptable basis for recommending variety B to the farmers in the area (apart from the problems of not having discussed other characters than yield, and having data for only one year)?

In the first example, where it was assumed that the varieties were grown on different farms, the difference might be due to artifacts such as the variety B being a new variety that only the best farmers had access to. These farmers might plant earlier and fertilize more, and the apparent varietal difference could have been due to differences in cultivation practices.

Even in the second example, where it was assumed that all three varieties were grown on the same farms, there could be other causes for the yield difference than the genetic difference between varieties. It could be known that the variety B responded to higher fertilizer levels than A; or variety A could be earlier maturing than B, and the yield difference could be due to different fertilizer levels or harvest times.

In general it is difficult to analyze cause and effect by the survey method. To solve this problem, at least partly, the survey could include information on several variables, such as in our case fertilizer levels and harvest times. For the experimental method we design each experiment to answer only those main effects and interactions that we are interested in, and keep other factors constant.

Certain questions cannot be answered through experimentation. If we want to test possible advantages for future farmers to have a degree from an agricultural college, we can not select a group of future farmers and allow only a randomly selected group of the individuals to go to college. We can, however, through a survey, analyze the results from farms run by farmers with a college degree and farmers without such training.

Another advantage of the survey method is that we can use data from past times. In forestry we can find out how a tree has been growing during the last 50 to 100, and sometimes 1000 years, through measuring the annual rings or the distances between sets of branches. If we in addition have information about the climate, we can analyze long term relationships between climate and growth by the survey method.

2. Experiments

In an experiment we define the cause and measure the effect. The objective of a good experiment is to obtain as accurate answers as possible to the experimental questions. We can define confidence limits for the answers, and also specify the area, or more general the universe, for which the data are valid. Finally the cost of the experiments should be as low as possible.

There are many factors contributing to the accuracy of an experiment. Reduction of the random variation is probably the most important, and this will be discussed under the chapter on different designs. It is important to avoid large errors such as swapping of treatments, writing errors and weighing and measuring errors. Sometimes accuracy is related to costs, and it is necessary to judge accuracy in relation to price. If two alternative methods cost the same, the most accurate is the best method. If two methods at equally accurate, the cheapest one is the best.

A prerequisite for correct error estimates is that the error variation is random. Much of the variation in a field experiment is not random. Soil structure and nutrition, soil moisture conditions, competition from weeds and attack of insects and diseases often change continuously from one end of a field to the other. We transform such systematic variation to random variation by random distribution of the treatments.

IV EXPERIMENTAL FACTOR / EXPERIMENTAL TREATMENT

When we set out to gather information about a biological, economic or technical phenomenon through experimentation, we must start with the definition of the experimental treatment. This can be more or less difficult depending on the complexity of the problem raised. We will start this chapter by a discussion of the concepts of experimental questions, experimental factors, steps or levels of a factor and experimental treatments.

The concepts of experimental question and treatment are often used as synonym concepts, but the **treatment** is the concrete material that we test in an experiment whereas the **question** is a more abstract concept describing the phenomenon we want to increase our knowledge about. It is easiest to explain the difference through some practical examples. In a varietal experiment the question is which variety is the most suitable for growing in a certain area or under a specific environment or agricultural system. The treatments are the different varieties included in the experiment.

If we want to test different questions in the same experiments we must introduce the concepts of factors and steps / levels. In a combined experiment with varieties and fertilizers the factors are varieties and fertilizer. The different varieties are the steps or levels of the factor varieties, and the different quantities of fertilizer are the steps / levels of the factor fertilizer.

Steps or levels are linguistically most correctly used for factors that vary in quantities or concentration, but steps or levels can also be used for application times or harvest times, different fertilizer types or different pesticides. Treatments and steps or levels are synonyms in a one factor experiment.

The choice of experimental questions must be decided within the different subject areas, but we would like to stress the importance of asking relevant and essential questions. Further it is important to evaluate if the treatments chosen give relevant answers to the question asked.

In a varietal experiment we are interested in varietal differences. We are interested in, for instance, whether the varieties have genetically different yield potential. The treatments are different seed lots or plant materials of the varieties. If, in an experiment with different potato varieties, some varieties are contaminated with virus and others are not, our treatments will not give a correct answer to the questions asked.

If we know the answer we should not ask the question. To conduct an experiment if we know the result is only waste of resources, but there are numerous examples of such experiments. The argument is often that the experiment will be used for demonstration purposes. Demonstration of known effects should never be mixed with experimental questions, it always leads to complex and poor experimental design. Demonstration of known effects can be done on single plots with less resources than what is needed to include an extra factor in an experiment. This does not mean that experiments cannot be used for demonstration.

1. Quantitative and qualitative factors

It is practical to distinguish between quantitative and qualitative experimental factors. A qualitative factor has treatments that are qualitative alternatives such as varieties, methods or different tools. The treatments cannot be arranged in an increasing or decreasing order. For a qualitative factor we are interested in determining differences between the different alternatives and to set confidence limits for these differences.

Quantitative factors are factors with different quantities, temperatures, times or concentrations as alternatives. The steps or levels of such a factor can be arranged in an increasing or decreasing order prior to the experiment. For quantitative factors we are interested in the increase or decrease in relation to the treatment or to determine an optimum value on a response curve. We are not interested in the difference between treatments since that depends on the difference between levels or steps which are determined by the experimenter.

There will always be some border cases. If, for instance, we want to test the effect of a factor on varieties with different maturity, we can select a number of varieties with different maturity and, hence, change the experimental factor variety from a qualitative to a quantitative factor.

There are some important differences between qualitative and quantitative factor regarding how we proceed with an experimental program. For a qualitative question we start with as many alternatives as possible, and reduce the number of treatments on the basis of a set of preliminary trials. It is important not to leave out the best alternatives. For a quantitative factor we start with a few alternatives which cover the whole range of potential variation between treatments. The idea is to define the optimum on a curve, and it is

important to make sure that our initial experiment include this optimum. Secondly we design an experiment with many treatments around this optimum in order to determine the optimum as accurate as possible.

2. Factorial questions

It is often economical to test more than one factor in the same experiment. Two experimental questions can be to compare three different varieties, A, B and C, and to determine the effect of two levels of fertilizer, 1 and 2. It is possible to evaluate these two questions separately by two one-factor experiments as indicated in table IV.1. Both experiments have five replications with random distribution of treatments within replications.

Table IV.1. Two one-factor experiments with varieties and fertilizer levels.

Varietal experiment		Fertilizer experiment	
Block	Variety	Block	Level
I	A B C	I	1 2
II	B A C	II	2 1
III	A C B	III	2 1
IV	B C A	IV	1 2
V	C A B	V	1 2

For simplicity we assume that we use the same size plots for both experiments, and that the cost of the experiment is directly related to the total number of plots. The information obtained is then determined by the number of comparisons we obtain for each of the factors. For the varietal experiment we have five comparisons of varietal differences, and in the fertilizer experiment we have five comparisons of the two fertilizer levels.

We can combine these two questions in a factorial experiment with varieties and fertilizer levels as shown in table IV.2. With three varieties and two fertilizer levels we have six treatments, and in order to use approximately the same resources as in the previous example, we plan the experiment with four replications for a total of twenty four plots. The treatments are randomized within blocks.

In this alternative we can compare varieties two times per replication, one for each

fertilizer level, which give a total of 8 varietal comparisons. The fertilizer levels can be compared for each of the three varieties in each replication, a total of 12 comparisons for the experiment. For the same amount of resources we have more answers for the main effects using a factorial plan than two one-factor experiments. In addition we can also determine eventual interactions between the two factors in a factorial design. An interaction means that the effect of one factor depends on the level of another. In our case the effect of fertilizer could be different for the different varieties, or, with other words, the varieties could react differently to added fertilizer. The interaction is determined ones per replication.

We obtain a total of twenty four answers in the factorial experiment compared to ten answers in the two one-factor experiments.

Table IV.2. Factorial experiment with three varieties and two fertilizer levels.

Block	Variety and fertilizer level					
I	A,1	B,2	B,1	C,2	C,1	A,2
II	B,2	C,2	A,1	C,1	A,2	B,1
III	B,1	C,2	C,1	A,2	A,1	B,2
IV	C,2	C,1	B,2	A,1	A,2	B,1

A final point is that for the one-factor experiments we have to choose a level of the other factor for each of the experiments. If there are interactions between factors, the results from one-factor experiments have limited value.

The advantage of factorial designs is fully exploited only if the plans are complete. If we for instance had omitted treatment C,2 because the variety C was known not to respond to higher levels of fertilizer, the number of answers from the experiment would be drastically reduced. The difference between varieties A and B would still be determined two times per replication, but we would have only one answer per replication on A-C and B-C. The effect of fertilizer would be determined twice instead of three times per replication, and concerning the interaction, we would only determine that for varieties A and B, and not for A-C and B-C. Whenever possible, factorial plans should be complete.

If we have several factors or many levels of each factor, the resulting number of treatments can be very large and the experiment very expensive. Specific designs (split plot,

confounding) have been developed to reduce the size of such experiments. Some of these designs will be discussed later.

EXPERIMENTAL UNIT, REPLICATIONS AND PARALLEL OBSERVATIONS

1. Experimental unit

The experimental unit is the unit on which we apply the experimental treatments and make the basic observations. In field experiments the unit is normally a plot of 1 to 20 m^2. The experimental unit can be a box or a pot in a greenhouse or growth chamber, a single plant or tree, or a leaf or part of a leaf. In animal husbandry experiments the units can be single animals, groups of animals or the production of an animal over a certain period of time. In economic research the units can be a farm, a farming system or any subdivision into meaningful economic components.

The experimental unit is the representative of the universe that we want to make predictions for. It is important that the unit is a real representative for what we want to measure. A section of a leaf can be a proper representative of a variety if the purpose is to determine resistance properties of the variety, whereas agronomic characteristics such as yield, earliness, lodging etc. must be measured on a plot that is big enough to represent a field stand of a crop.

It is only the variation between units within replication that contributes to the error variance of a trial. This error variance will normally decrease with increasing unit size. Weighing and measuring errors will be relatively less important on larger units. Only if there are very variable soil conditions will small plots be less variable than large plots.

It is difficult to determine an optimal size of the experimental unit. The cost will increase with increased unit size, and the economic considerations are the most important ones to keep the units small. The unit size is often based on availability of technical equipment, economic resources and land resources. From a statistical point of view the most important factors are representativity and reduction of the experimental error.

Neighbor effect

In many types of experiments each unit can be influenced by the treatment applied to the neighbor units, if measures are not taken to avoid such effects. In a fertilizer experiment this is easy to understand. Plants on a unit with low fertilizer intensity will have roots extending into the fertilized neighbor plot. There can also be neighbor effects between varieties or species. A tall variety will be more competitive than a short one, and in perennial

experiments a competitive species or variety can compete with neighboring treatments both for nutrients and space.

To separate individual plots in a field experiment we often leave a larger space between plots than between rows within plots. Such border effects are normally affecting all treatments equally, but these effects can also be avioded by harvesting only the center part of the plots.

To avoid neighbor effects in field experiments we install border rows or border plots between the treatments. Evidently, neighbor effects and border effects are influenced by the shape of the plots. Long and narrow plots are technically easiest to plant and harvest, but they also have largest borders to other plots.

It is not only field experiments that can have border effects and neighbor effects. An example is an experiment with different feed compositions for milking cows. To avoid the problem with differences between different cows, one can test a certain ration over one week. The unit in that case is not "cow", but "cow in one week". The effect of a new ration will be somewhat delayed, and we have to wait for some days to start measuring the effect of the new ration. This waiting period can be seen as a border or neighbor effect between treatments.

During the planning of an experiment it is important to carefully evaluate the size and shape of the experimental units. The aim should always be to use a representative unit, and to reduce as much as possible the variation between units within replication. Variation within the unit itself does not influence the error variance of the experiment.

2. Replications and / or parallels

The standard error of the mean will decrease with increasing number of observations per treatment. The decrease is proportional to the square root of the number of observations. Hence, the increase from 2 to 3 or 4 observations have the largest effect. Further increases by one observation have less and less effect on the standard error. Increase in the number of observations also has another effect. The number of DF for the error increases which means that our estimate of the error is more reliable. Also this effect is strongest for an increase on the low end of the scale. From an F-table or a t-table we find that there is only a slight effect of increases in DF above 20 - 30. In relation to the effect of DF it is most important to have many observations per treatment if we have few treatments. If we have

many treatments, there will always be enough DF for the error variance.

If we have an approximate estimate of the magnitude of the error variance, we can calculate the number of observations needed for a predetermined level of the least significant difference.

The observations must be random samples of the universe we are interested in. It does not help us to have many observations in one particular field if we are interested in giving advice for a number of farms in an area for the coming year. The samples must then represent the variation between locations and between years.

In a one-way table the observations of the same treatment are called parallels. In that case there are no connection between observation no. 1 for one treatment and observation no.1 for another treatment. All the variation within treatment will then be regarded as random variation and is included in the error variance. In a two-way table the first observations of each treatment are observed in the same block or replication, and differences between blocks can be eliminated from the error variance. In this case the observations for the same treatment are called replications.

We will come back to the discussion of different types of replications later in the course.

VI. EXPERIMENTAL DESIGNS

Experimental designs are characterized by the principles used for assigning treatments to the experimental units. We will discuss:

Complete random distribution
Complete block design
Latin square
Split plot / split block
Incomplete block designs (confounding)

Split plot / split block and confounding are designs used *only* for factorial questions, but complete block designs can also be used for factorial questions.

1. Complete random distribution

In a complete random distribution design the treatments are randomly distributed on all available experimental units. It can be different numbers of observation for different treatments. All the variation between experimental units will contribute to the experimental error, and this design is, therefore, mainly used where the units are homogeneous.

An example of a complete random distribution design is a pot experiment with fertilizer treatments where the soil for all treatments is carefully mixed and divided between pots (the units). Finally the treatments (the fertilizer levels) are added to the different pots before planting.

The statistical analysis of a complete random distribution design is a one-way analysis of variance.

It should be stressed that in most types of experiments in natural sciences there are many reasons for arranging the treatments in blocks, and that the complete random distribution design, therefore, should be used only in exeptional cases.

2. Complete block design

It is often difficult to obtain a large enough number of homogeneous experimental units for completely random distribution of the treatments. In complete block designs variation

between blocks is removed from the error variance. The requirement for homogeneous experimental units for the whole experiment is then reduced to a requirement for homogeneous experimental units within blocks. For this design we normally use the same number of replications for all the treatments, and all treatments must be represented once per block. It is only the variation between experimental units within blocks that contribute to the error variance.

The most common example of a block experiment is a field experiment. If nothing is known about the soil variation, we try to make the blocks as compact as possible. However, if there is a gradient across a field we should define the blocks such that this systematic variation is between blocks or within plots. The objective is to reduce as much as possible the variation between plots within blocks.

If there is stripe variation due to plowing or drainage, it is important to place the blocks along these stripes and the plots perpendicular to the stripes. In field experiments it is customary to place the blocks side by side, but this is not necessary. If there is a specifically variable area in our experimental field, we can place one block on each side of such a problem area.

In other types of experiments we must use other criteria for the construction of blocks. If we have only two treatments in an animal husbandry experiment, we could select identical twins as blocks to eliminate genetic variability and age differences. In an experiment with methods for manual labor, the individual workers could be the blocks in that all persons involved performed all the job methods. A prerequisite is that all the workers had a good knowledge of all the methods. To avoid "border effects" each worker must be equally fit at the beginning of each method, for instance that each method was tested over a period of one day or one week and that the skill in performing one method did not effect the skill of performing the next method. Compared to a complete randomized design the complete block design has two weaknesses:

1. There must be the same number of replications for all the treatments. Some times we wish to vary the number of replications in order to make use of all available material.
2. The number of DF for the error variance will be reduced compared to the completely random distribution design. This is particularly important in small experiments where both the number of treatments and the number of replications are low. This argument is valid only if we want to make predictions on the basis of a single experiment. If

the experiment is one of a series of experiments, we are less interested in the within experiment error variance.

3. Latin square

In a complete block design we can eliminate systematic variation in one direction. In a latin quare design we eliminate systematic variation in two directions. Latin squares are limited to experiments with the same number of treatments and replications. The experimental units are arranged in rows and columns, and the treatments must appear once per row and once per column. An example with 4 treatments and four replications is shown in table VI.1.

Table VI.1. Distribution of four treatments, A - D, in a latin square with four rows, I - IV, and four columns, 1 - 4.

	Column			
Row	1 (2)	2 (4)	3 (1)	4 (3)
I (III)	A	B	C	D
II (I)	D	A	B	C
III (IV)	C	D	A	B
IV (II)	B	C	D	A

In order to achieve random distribution of the treatments in rows and columns, we assign new random row and column numbers as shown in parenthesis, and we rearrange the treatments as shown in table VI.2.

Table VI.2. Random distribution of four treatments, A - D, in a latin square with four rows, I - IV, and four columns, 1 - 4.

	Column			
Row	(1)	(2)	(3)	(4)
(I)	B	D	C	A
(II)	D	B	A	C
(III)	C	A	D	B
(IV)	A	C	B	D

In an analysis of variance of a latin square the total variation is divided into four components; one between rows, one between columns, one between treatments and an experimental error.

$CT = X_{...}^2 / t^2$ where t is the number of treatments

$$SS_{total} = \sum_{i=1}^{t} \sum_{j=1}^{t} \sum_{k=1}^{t} X_{ijk}^2 - CT$$

$$SS_{row} = \sum_{i=1}^{t} X_{i..}^2 / t - CT$$

$$SS_{col.} = \sum_{j=1}^{t} X_{.j.}^2 / t - CT$$

$$SS_{tr.mt.} = \sum_{k=1}^{t} X_{..k}^2 / t - CT$$

$$SS_{error} = SS_{total} - SS_{row} - SS_{column} - SS_{treatment}$$

Given that we have two different types of systematic variation that we want to eliminate, a Latin square will give a reduced error variance compared to a block design. In field experiments there are often systematic variation in two directions. These can be due to tillage, fertility or topography.

In latin squares it is normally advantageous to use near square plots. This design is most efficient when the differences between columns are the same for all rows.

Latin squares can also be used for other than field experiments. If we want to compare four different working techniques, we can use four workers to test all four techniques. We also know that climatic conditions may influence the results, and, therefore, the tests are carried out over four days such that all methods are tested by all the workers and all the days. The variation due to workers and days will then not influence the standard error of the means of techniques.

The disadvantages with a latin square design compared to a block design are that we have to have the same number of replications and treatments, and that the number of DF for the error variance is reduced both by the DF for rows and columns. For these reasons the latin square designs are practical only for 4 - 6 treatments.

4. Split plot designs

Some experimental factors need larger plots or border areas than others. In a complete randomized block experiment with factorial questions, the plot size is decided by the factor requiring the largest plots including border areas. In a split plot design the largest plots are used for the treatments of the factor requiring this plot size, and these plots are split into smaller plots for the treatments of another factor. The treatments on the large plots are randomized within replications, and the treatments on small plots are randomized within the large plots.

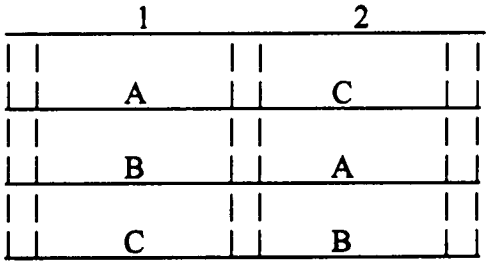
Fertilizer treatments need border areas whereas varieties normally does not. An experiment with two fertilizer treatments and three varieties could be laid out as a split plot experiment. Each replication is divided into two large plots with border areas for the fertilizer treatments, and each large plot is divided into three small plots for the varieties. With four replications and the proper randomization, the experimental plan would look as shown in table VI.4.

Table VI.4. Split plot experiment with 2 fertilizer levels (1 and 2), 3 varieties (A,B and C), and 4 replications (I - IV).

I						II					
1			2			2			1		
A	B	C	B	C	A	C	B	A	A	C	B

III						IV					
1			2			2			1		
C	A	B	A	C	B	A	C	B	C	B	A

There will be a larger distance between the fertilizer plots than between the varietal plots. We must therefore expect a larger error variance for the comparison between fertilizer treatments than for the comparison between varieties. This is particularly true if the division of the big plots into small plots is in the same direction as the division of the replications into big plots. If the small plots are split perpendicular on the split of replications into big plots, and the random variation occur in stripes along the small plots, the error on the small plots may be larger than the error on the big plots.



We call the error variances in a split plot error (a) and error (b) for the errors on large and small plots respectively.

The comparisons between treatments on large plots are less accurate than the comparisons of treatments on small plots for two reasons. Firstly the error variance is larger, and secondly the number of DF is lower. There are fewer replications of the large plots than of the small plots. The number of DF in the analysis of variance is as follows;

Analysis of variance

Source of variation	DF
Total	23
Rep.	3
Fertilizer levels	1
error (a)	3
Varieties	2
Fert. x Varieties	2
error (b)	12

The use of split plot designs is not limited to field experiments. Let us assume that

we want to test three recipes for cakes and two baking temperatures. The oven available can bake three cakes at a time. The oven would then be the big plot, where the temperatures could be regulated, and the different cakes would be the small plots, and for each temperature they would be baked with the three different recipes. The whole experiment had to be repeated over time or in different ovens for replications.

Split split plot

If we have three experimental factors that require three different plot sizes, we can use a split split plot design. An example could be to compare two different tillage systems, three different fertilizer treatments and three different cereal species in one experiment. The tillage systems require very large plots, the fertilizer treatments require medium size plots and the comparisons of species can be done on small plots.

With three replications the analysis of variance for such a split split plot experiment would be:

Analysis of variance

Source of variation	DF
Total	53
Replications	2
Tillage	1
Error (a)	2
Fertilizer	2
Tillage x Fertilizer	2
Error (b)	8
Species	2
Tillage x Species	2
Fertilizer x Species	4
Tillage x Fertilizer x Species	4
Error (c)	24

Error (a) is the interaction between tillage and replication, error (b) contains the two factor interaction between fertilizer and replications and the three factor interaction between tillage, fertilizer and replication. Error (c) contains 4 components, the interactions between replication and species, replication, tillage and species, replication, fertilizer and species, and replication, tillage, fertilizer and species.

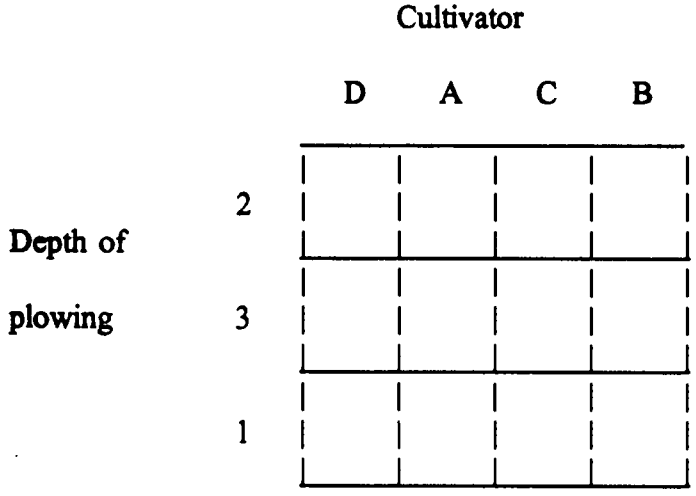
All main effects and interactions between experimental factors are tested against the error immediately below the effects in the analysis of variance table.

Error (a) has normally the largest variance and error (c) the smallest. In addition there will always be few DF for error (a) which means that we need large differences between levels of the factor on the largest plots in order to find significant effects.

Split split plot can also be an appropriate design for other that field experiments. Assume that we want to investigate the effects of relative humidity, cage placements and different feed compositions for chicken. The "plots" for climatic differences must evidently be a room, cages can be arranged in different heights, and feed compositions can be given to individual birds.

Split block

For two factors that both need large plots we can divide a replication into two sets of big plots perpendicular to each other. Let us assume that we will test three depths of plowing (1, 2 and 3), and four different cultivators (A, B, C and D). For such experimental questions a split block plan is a good experimental design. One replication can be:



Both factors are applied to large plots, and the question of which one has the highest error variance depends on the soil variation and the precision of the plowing and cultivation operations. Part of the soil variation will be eliminated from the error variance for the small plots, error (c), on which the interaction between the two factors is determined. Error (c) will also have more DF than the other two error terms. A split block design, therefore, give more accurate estimates of the interaction than of the main factors.

With four replications, each with new randomization of both factors, the analysis of

variance will be:

Analysis of variance

Source of variation	DF
Total	47
Replication	3
Plowing	2
Error (a)	6
Cultivation	3
Error (b)	9
Plowing x Cultivation	6
Error (c)	18

When should we use split plot / split block designs?

If we compare the analyses of variance for a complete block design and a split plot or split bloc design, we find that the error variance of the complete block design is the sum of the different error variances for the split plot / split block designs. If error (b) in a split plot experiment is reduced compared to a similar experiment laid out as a complete block design, the error (a) will have increased.

We should, therefore, only use split plot / split block designs when it has practical advantages, or if we are not very interested in the main effect of the factor on the large plots. Split block should only be used when it is necessary from a technical point of view, or if we are mainly interested in the interaction between two factors and not the main effects.

5. Incomplete blocks, - confounding

In order to eliminate some of the variation within replications from the error variation, we can divide the replications into smaller blocks. The treatments are assigned to the experimental units in such a way that the differences between the small blocks are the same as the effects of less important interactions. We confound these interactions with the block effects.

Confounding is mainly used for factorial questions with more than three factors and with two levels of each factor. With three factors, A, B and C, and with levels 0 and 1 for

each factor, We have the following eight treatment combinations:

$A_0B_0C_0$	$A_1B_0C_0$	$A_0B_1C_0$	$A_1B_1C_0$	$A_0B_0C_1$	$A_1B_0C_1$	$A_0B_1C_1$	$A_1B_1C_1$
1	a	b	ab	c	ac	bc	abc

The combination with the lowest values of all factors is called 1, and the other combinations are given names in lower case letters according to the factors having the highest level.

The main effect of A can be determined by four differences in each replication, namely (a-1), (ab-b), (ac-c) and (abc-bc). On the average

$$A = 1/4\ [(a - 1) + (ab - b) + (ac - c) + (abc - bc)]$$
$$= 1/4\ (a + ab + ac + abc - 1 - b - c - bc)$$

All the treatments with a are given "+", and all the treatments without a are given "-". Equally the main effects of B and C are given "+" for all the treatments where the factor has the highest level and "-" for treatments with the lowest level.

The interaction between A and B is the difference between the effect of A with and without B. In each replication the AB interaction is determined twice, with and without C. On the average

$$AB = 1/2\ [\ 1/2\ [(ab - b) - (a - 1)] + 1/2\ [(abc - bc) - (ac - c)]]$$
$$= 1/4\ (+1 - a - b + ab + c - ac - bc + abc)$$

Note that there are "+" signs for the treatments where none or both of A and B are present, and "-" signs for the treatments where only one of A or B are present.

The three factor interaction, ABC, is the difference between the two factor interaction AB with and without C.

$$ABC = 1/4\ (-1 + a + b - ab + c - ac - bc + abc)$$

For a three factor interaction there are "+" signs for treatments where 1 or 3 of the

factors are present, and "-" signs if 0 or 2 factors are present. The estimates for all the effects are presented in table VI.5.

Such a table is easy to construct. We start with the signs for the main effects as described above. In the column for two-factor interactions we multiply the signs for the two appropriate main factors, and the signs for the three factor interaction is obtained by multiplying the signs for a two factor interaction and the main effect of the third factor.

Table VI.5. Main effects and interactions in experiments with 3 factors with 2 levels each.

	Effects						
Treatment	A	B	AB	C	AC	BC	ABC
1	-	-	+	-	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
ab	+	+	+	-	-	-	-
c	-	-	+	+	-	-	+
ac	+	-	-	+	+	-	-
bc	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+

It is easy to extend the table to more than three factors. For a fourth factor we add the lines d, ad, bd, abd, cd, acd, bcd and abcd, and the columns D, AD, BD, ABD, CD, ACD, BCD and ABCD.

In an experiment with three factors and two levels of each factor, there are eight different treatments. Let us divide each replication into two blocks, and assign treatments 1, ab, ac and bc to one of the blocks, and a, b, c and abc to the other. From table VI.5 we find that all the main effects and all two factor interactions have two "+" and two "-" in each block. Differences between the block will, therefore not influence these effects. The effect of the three factor interaction, ABC, is identical to the difference between the blocks. We have confounded the effects of the three factor interaction and the blocks.

With a complete block design we would have had two DF for replication, seven DF for the different treatment effects, and fourteen DF for the error variance. The random

variation within block of four units is usually less than in blocks of eight units. The confounded design will, therefore, give a more accurate determination of the main effects and the two factor interaction. On the other hand we sacrifice the three factor interaction, and we loose two DF for the error variance. The three factor interactions are normally small and difficult to interpret and to utilize in practical extension work. Confounding this interaction is not a big loss. Higher interaction have even less interest and can be confounded in order to increase the accuracy of the determination of main effects and two factor interactions.

With three replications, six blocks, the analysis of variance for the three factor experiment is:

Analysis of variance

Source of variation	DF
Total	23
Blocks	5
A	1
B	1
AB	1
C	1
AC	1
BC	1
Error	12

The effect of blocks comprise the effects of replication with two DF and blocks within replication with three DF. Technically the latter component comprise the effect of ABC and the interaction between ABC and interaction.

Comparison between confounding and split plot

In the above example we confounded the three factor interaction with blocks. We could have confounded any of the other interactions, for instance AB, by assigning treatments 1, ab, c and abc to one block, and a, b, ac and bc to the other. In that case all the main effects and AC, BC and ABC would have been determined within blocks. In a confounded design we confound the least interesting interaction which is normally the highest one.

Let us confound one of the main effects with blocks. If we assign all the treatments

with "+" for factor A to one block and the ones with "-" to the other, we have a split plot experiment with factor A on big plots (blocks). The difference between confounding and split plot is that in a split plot design we also "split" the block effects into an effect of replication, a treatment effect and an error term on this treatment effect in order to test the effect of the factor confounded with the big plots (blocks).

VII. SUBDIVIDING SS AND DF FOR TREATMENT AND ERROR

In the subchapter on factorial designs we divided the effect of treatment into main effects and interactions. Also in a one factor experiment we can have á priori questions that can be answered by dividing the treatment SS and DF into different components. Sometimes these different components also have different errors.

1. Numerical example

In an experiment with five barley varieties (exercise 4) the varieties A and B were two row barley, and varieties C, D and E were six row barley. The effect of varieties can á priori be divided into three components, the difference between two row and six row, the differences within two row and the differences within six row varieties.

The yield in sum over three replications were:

Variety:	A	B	C	D	E	Σ
	10,44	10,47	7,99	9,46	9,18	47,54
Two row						20,91
Six row						26,63

The sum of squares for the different comparisons are:

$SS_{2r-6r} = 20{,}91^2 / 6 + 26{,}63^2 / 9 - CT = 0{,}99645$

$SS_{within\ 2r} = (10{,}44^2 + 10{,}47^2) / 3 - 20{,}91^2 / 6 = 0{,}00015$

$SS_{within\ 6r} = (7{,}99^2 + 9{,}46^2 + 9{,}18^2) / 3 - 26{,}63^2 / 9 = 0{,}40616$

with 1, 1 and 2 DF respectively. The total of these three sum squares is 1.40276 which is the SS_{var} in exercise 4.

2. Contrasts

Contrasts are comparisons with 1 DF. Let's call the yields of the 5 barley varieties discussed above a, b, c, d and e. a and b are the yields of the two row varieties and c, d and e are the yields of the six row varieties. The mean difference between the two groups is 1/2 (a + b) -

1/3 (c + d + e). If we multiply by 6 to avoid fractions, we can write 3a + 3b - 2c - 2d - 2e. The comparison must have equal weight for both components, the sum of the coefficients must be zero (+3 +3 -2 -2 -2 = 0).

The difference a - b is also a contrast. When the experiment has more than two treatments, we can construct a number of different contrasts, but we are specially interested in independent or *orthogonal* contrasts. We define contrasts by assigning coefficients to the treatment means such that the coefficients add up to zero. For two contrasts to be orthogonal the sum of the product of the two sets of coefficients must also be zero.

	Treatment				
	A	B	C	D	E
Coeff. for contrast 1	+3	+3	-2	-2	-2
Coeff. for contrast 2	+1	-1	0	0	0
Product	+3	-3	0	0	0

We can make only as many orthogonal contrasts as we have degrees of freedom for treatment. The contrast a - c, for instance, is not orthogonal to the two previous ones. If we define the third contrast as c - d, which is orthogonal to the contrasts listed in the table, the only possible fourth contrast is c + d - 2e.

We can make a number of different sets of orthogonal contrasts, but they should also be meaningful comparisons. In the example contrast 1 is the difference between 2 row and 6 row barley, contrast 2 is the difference within 2 row varieties and contrasts 3 and 4 together represent the differences within the 6 row varieties.

If a contrast (C) is based on treatment totals, the following formula gives the sum square:

$$SS_C = C^2 / r\Sigma c_i^2$$

where r is the number of replications and c_i are the coefficients. In exercise 4 we had these treatment totals: a = 10.44, b = 10.47, c = 7.99, d = 9.46, e = 9.18, and the 4 orthogonal contrasts and their sum squares are:

1. $C_1 = 3a+3b-2c-2d-2e = 9.47$	$r\Sigma c^2_{1i} = 3(9+9+4+4+4) = 90$	$SS_{C1} = 0.99645$
2. $C_2 = 1a - 1b = -0.03$	$r\Sigma c^2_{2i} = 3(1 + 1) = 6$	$SS_{C2} = 0.00015$
3. $C_3 = 1c - 1d = -1.47$	$r\Sigma c^2_{3i} = 3(1 + 1) = 6$	$SS_{C3} = 0.36019$
4. $C_4 = 1c + 1d - 2e = -0.91$	$r\Sigma c^2_{4i} = 3(1 + 1 + 4) = 18$	$SS_{C4} = 0.04601$

The sum squares for the four contrasts add up to 1.4028, which is the $SS_{treatment}$ in exercise 4. Since the contrasts have only 1 DF, MS = SS, and in this case we test all contrast MS' against the common MS_{error} of 0.0436.

The conclusions from this analysis is that the 2 row varieties gave higher yields than the 6 row varieties, the two 2 row varieties did not give significantly different yields, and the differences between the 6 row varieties were due to a significant difference between varieties C and D.

2. Orthogonal polynomials

In experiments with quantitative factors we can use orthogonal contrasts to isolate for instance linear, second and third degree relationships between the treatments and the effects. Also for these contrasts we can construct as many orthogonal contrasts as we have number of DF for treatments. Let us start with a polynomial where the effect (x) is expressed as a function of the treatments (t)

$$x = a + bt + ct^2 + dt^3 + ...etc.$$

The coefficients for the first contrast are set to express the linear effect of treatments, and the coefficients for the second contrast to express the second degree effects etc. Data from biological experiments seldom follow a 2. degree function, but we use the polynomials as an approximation because they are easy to calculate and give satisfactory estimates of a deviation from a linear relationship within a limited range.

In order to arrive at the simplest possible sets of coefficients we assume that there are equal distances between treatments. In a fertilizer experiment with different nitrogen levels, for instance, the levels are 80, 100, 120 and 140 kg / ha, the equal distance between treatments is 20 kg N / ha.

To estimate the first contrast, C_1 , which express x as a linear function of t, we

calculate the coefficients, c_i, from the simplest possible linear function of t.

$$c_i = a + t_i$$

The only condition we set for this function is that it represents a linear relationship between the coefficients and the treatments.

t_i	c'_{1i}	c''_{1i}	c_{1i}
0	a	-3/2	-3
1	a + 1	-1/2	-1
2	a + 2	1/2	1
3	a + 3	3/2	3
Σ	4a + 6	0	0

In this example t varies from 0 to 3 in equal intervals of 1. We assume that the combination of treatment effects given by the coefficients should be a contrast with one DF, which mean that the sum of the coefficients should be 0. Starting with c'_{1i} this means that

$$4a + 6 = 0, \text{ or that } a = -3/2$$

Substituting -3/2 for a gives c''_{1i} , and multiplying by 2 to avoid fractions, gives c_{1i}. These coefficients are general for all quantitative experiments with four treatments and equal differences between treatments, and the contrast

$$C_1 = \Sigma c_{1i} x_i$$

express the linear effect of t_i on x_i.

Similarly we can set coefficients for the quadratic effect of t_i on x_i. again we calculate the coefficients from the simplest possible 2. degree equation

$$c_{2i} = a + bt + t^2$$

In addition to the condition that the sum of the coefficients should be 0, we add the condition that the sum of the product of c_{1i} and c_{2i} should be 0.

t_i	c_{1i}	c'_{2i}	$c_{1i} c'_{2i}$	c_{2i}
0	-3	a	-3a	1
1	-1	a+b+1	-a-b-1	-1
2	1	a+2b+4	a+2b+4	-1
3	3	a+3b+9	3a+9b+27	1
Σ	0	4a+6b+14	10b+30	0

$10b + 30 = 0;$ $b = -3$

$4a + 6b + 14 = 0;$ $a = 1$

Substituting the values for a and b in c'_{2i} gives c_{2i}. Using these coefficients we can calculate the contrast

$$C_2 = \Sigma c_{2i} x_i$$

which express the quadratic deviation from the linear effect of t_i on x_i.

The coefficients for orthogonal polynomials are given in a set of tables. A short extract of such tables is given in table VII.1.

Table VII.1. Coefficients for 1. and 2. degree effects for experiments with 3 to 7 treatments.

	Number of treatments									
	3		4		5		6		7	
Degree	1.	2.	1.	2.	1.	2.	1.	2.	1.	2.
	- 1	+1	- 3	+1	- 2	+2	- 5	+5	- 3	+5
	0	- 2	- 1	- 1	- 1	- 1	- 3	- 1	- 2	0
	+1	+1	+1	- 1	0	- 2	- 1	- 4	- 1	- 3
			+3	+1	+1	- 1	+1	- 4	0	- 4
					+2	+2	+3	- 1	+1	- 3
							+5	+5	+2	0
									+3	+5
Σc^2_i	2	6	20	4	10	14	70	84	28	84

Estimating regression curves

In addition to using the coefficients to calculate variance components for treatment effects, we can use them to construct regression curves for the effects of t_i on x_i. First we estimate an average linear effect (Lin) and an average quadratic effect(Quad).

$$Lin = C_1 / r\Sigma c^2_{1i}$$
$$Quad = C_2 / r\Sigma c^2_{2i}$$

In this case r is a scaling factor. In the analysis of variance we must always operate on the same basic units, for instance yields per plot. We have shown the calculation of contrasts based on the treatment totals and must divide by r in order to obtain comparable sum squares for comparison.

To present the results in graphic forms we often transform our data to for instance kg / ha, and if the contrasts, C_1 and C_2, are calculated on that basis, the r = 1 in the above formula.

To calculate the coordinate points (x, t) for the linear and quadratic effects, the formulas are

$$x_{(Lin)i} = \bar{x} + c_{1i}(Lin)$$
$$x_{(Quad)i} = x_{(Lin)i} + c_{2i}(Quad)$$

where $x_{(Lin)\,i}$ are the points on a linear regression line, and $x_{(Quad)\,i}$ are the points on the second degree regression line.

VIII. SERIES OF EXPERIMENTS

Single experiments can only in exceptional cases give adequate answers to agricultural experimental questions. The rule is that advances in agricultural technology and advice to farmers have to be based on series of experiments over locations and years. Statistically there are interactions between experimental factors and climatic and edaphic factors, and in the case that we cannot control these factors we have to look on our experiments as samples of a universe including the expected variation in climate as well as site specific variation. In order to be valid samples of the universe of interest, our experiments must cover different years and different locations.

For animal husbandry research these problems are less pronounced than for plant production, especially where the animals are kept in controlled environments. However, there are always variation from herd to herd both in relation to genetic and environmental differences.

Advice to farmers must in most cases be based on series of experiments rather than single trials.

1. Annual series

The models for series of experiments are given on page 29. Results from experimental series within years are only valid for the climatic conditions of that particular year. If the area covered include many climatic zones, as is the case for series carried out by many of the international agricultural research centers, the results can have more general applications. The within trial errors are normally not of interest for the analysis of series of experiments, and to simplify the analysis the calculations can be based on treatment means from each trial. An annual series is then simplified to a two way analysis over treatments and locations.

If the cost of running the experiments were proportional to the number of plots, it would be most economical to install only one replication per location. However, certain costs per location are independent of number of plots, and an evaluation of how to construct a series in relation to replications per location and number of locations must take this into account. The error mean square for a treatment average is MS_2 / nr where n is the number of locations and r is the number of replications within location. From the model we find that this translates to

$\sigma^2 / nr + \sigma^2_{Lv} / n$

The three most important factor for the evaluation of the number of replications and number of locations are:

1. *The cost per location compared to the cost per plot.* High costs per location favors more replications per trial. However, if the available experimental material is the limiting factor, as is often the case for testing of new genetic stocks, it is most important to cover as many locations as possible.
2. *The relationship between σ and σ_{Lv}.* Often the interaction term is much larger than the within location error which favors few replications per location.
3. *Number of treatments.* Many treatments favor few replications per location. If we have many treatments, the per location costs are relatively low compared to the cost per plot, and there will be a high number of DF for the error variance.

As a rule we try to have as many locations as possible and reduce the number of replications per location. Estimates have shown that we gain very little in accuracy by having more than two replications per location. In many cases the standard error on treatment means is lowest using only one plot per treatment per location. In that case we have no estimate of the within location error or the location x treatment interaction. Our error estimate contain both these components.

2. Perennial series

Experimental series used to give advice for the coming year or years must include samples from different years. Even samples from a few years may not be representative for the climatic conditions next year. Two or three years in a row can be relatively extreme climatically, and there are often climatic trends that makes it difficult to predict the conditions in future years.

Climatic and technological trends have the same effects. Increased fertilizer levels influence the choice of varieties, and the availability of more lodging resistant varieties in cereals will change the optimum fertilizer level. These factors limit the time span for which experiments will be meaningful for future advice.

With reference to the model (page 30), the interaction between treatments and years is the most correct error variance for the treatment effects. It is not necessary to have exactly

the same number of trials each year, and the results will be weighted according to the number of trials per year.

If the years are very different, it may be best to calculate treatment averages for each year and base the final analysis on a two way table of treatments and years.

If the number of experiments differ importantly between years, it is an alternative to analyze a perennial series as a two way table of treatments and locations. This gives the same treatment means as the weighted analysis, and the error term will be a weighted mean of year x treatment (MS_2) and location within year x treatment (MS_3). If MS_2 is clearly larger than MS_3, this method underestimate the standard error of the means and also increase the number of DF for the underestimated error.

IX. REPEATED TREATMENTS OVER YEARS

In some perennial experiments we repeat the treatments or the measurements or both each year. We can apply a treatment once and measure the effects over several years as in experiments with perennial species. Alternatively we can repeat the application of a treatment over time as in perennial fertilizer experiments.

If the results are measured only once, the analysis of the experiment is equivalent to an annual experiment. If we measure the effects several times (years), times (or years) becomes an experimental factor.

There is an important difference between such experiments and the perennial series discussed in the previous subchapter. With repeated treatments or measurements years are not independent or random. The error is not dependent on the variation between years, but on the variation between replications. In a perennial experiment with different gras species there may be large differences between years, and large interactions between species and years. In this case years is a fixed variable. We are not interested in the variation between years, but in the total yield over years.

In perennial experiments we are normally most interested in the results during the early years of the experiment. A meadow or a strawberry field may be plowed after a certain number of years. It is, therefore, of interest to analyze the results of the first year, the two first years, the three first years etc.

X. STUDENT EXERCISES

Exercise 1

Read both part a) and part b) before you collect the data for the exercise.

a) Find the average length of words (no. of letters per word) in a book. Count the letters in 50 words selected at random. Arrange the data in a frequency distribution, draw a histogram, and calculate the mean and the standard deviation.

b) Calculate the mean length of 10 and 10 consecutive words. Calculate the mean and the standard deviation for these 5 means and compare the results of the calculations in a) and b)

Exercise 2

Which words have most letters, those starting with a consonant or those starting with a vowel?

a. Select a random sample of 50 words from a book (how is this done ?). Count the number of letters, separately for words starting with a consonant and for words starting with a vowel. Present the frequency tables, calculate the mean, standard deviation and confidence limits for each type and carry out a t-test on the difference.

b. Select a consonant word at random, and the first vowel word which follows. Find the difference in word length. Repeat this procedure 25 times, and calculate the mean the standard deviation, the standard error and the confidence limits for the differences.

How is the standard deviation on each single observation (in part a) related to the standard deviation on the differences (in part b)?

Exercise 3

In a pot-experiment with different types of lime and different quantities of lime, there were 9 different treatments with lime and a control treatment (0) without lime. The experimental crop was small grain, and the total yields of grain and straw in 0.1 g per pot were as shown in the table.

Treatment no									
0	1	2	3	4	5	6	7	8	9
598	623	614	683	591	631	570	591	588	608
563	621	620	619	633	658	607	590	628	642
465	627	635	632	643	638	620	541	595	627
1626	1871	1869	1934	1867	1927	1797	1722	1811	1877

a) Carry out a one-way analysis of variance for all treatments (0-9), and for the limed treatments (1-9).

b) Compare the error variances in the two analyses.

c) Give a short evaluation of the results.

Exercise 4

In a trial with 5 barley varieties (A, B, C, D and E; t=5) and 3 replications (I, II and III; r=3), the following yields in kg per plot of 10 m^2 were obtained:

Block					
I		II		III	
Variety	yield	Variety	yield	Variety	yield
A	3.46	C	2.86	B	3,28
C	2.71	B	3.51	E	3.27
D	3.14	E	2.85	C	2.42
B	3.68	A	3.56	D	2.86
E	3.06	D	3.46	A	3.42

a) Analyze the grain yields by a two way analysis of variance

b) Calculate yields per hectare for the five varieties

c) Calculate the $LSD_{5\%}$ and CV% and discuss the results obtained.

Exercise 5

An experiment with different quantities of fertilizer for potatoes was laid out according to a latin square design. A field map with the treatments in kg complete fertilizer per hectare and harvested yield in kg per 20 m^2 plot is presented below.

		Column				
Row		A	B	C	D	E
1	fertilizer	0	100	300	400	200
	yield	63	64	77	64	83
2	fertilizer	200	0	400	100	300
	yield	72	65	70	59	83
3	fertilizer	300	200	100	0	400
	yield	76	88	74	58	81
4	fertilizer	400	300	0	200	100
	yield	78	83	65	53	75
5	fertilizer	100	400	200	300	0
	yield	84	85	74	66	64

Analyze the yields by an analysis of variance, calculate the coefficient of variation and the yield results in kg /ha. Present the yields graphically and discuss the results of the experiment.

Exercise 6

A fertilizer experiment with different quantities of nitrogen and potassium to small grains was laid out as a randomized block design with 3 replications. The treatments were:

Nitrogen: N_0 = 15 kg N / ha
N_1 = 30 " — "
N_2 = 45 " — "

Potassium: K_0 = 0 kg K_2O / ha
K_1 = 100 " — "
K_2 = 200 " — "

Yields in kg grains per 20 m^2 plots are given below.

	Replication		
Treatment	I	II	III
NO K0	6,30	9,05	7,20
N0 K1	7,90	7,75	6,75
N0 K2	5,75	8,25	5,60
N1 K0	6,75	8,40	5,95
N1 K1	8,00	10,05	7,75
N1 K2	7,75	9,05	8,15
N2 K0	7,15	9,30	7,00
N2 K1	8,45	9,35	8,35
N2 K2	8,90	10,50	9,40

1. Analyze the yields by an analysis of variance, and divide the treatment variance into the main effect of N, the main effect of K and the interaction N x K. Test the different effects against the common MS_{error}.
2. Calculate the CV%.
3. Present the means for the significant effects in kg / ha, and discuss the results.

Exercise 7

In an experiment with different herbicides, the following number of weeds per m^2 was counted.

	Replication					
Treatment	I	II	III	IV	V	Σ
a. Control	350	200	470	310	260	1590
b. Hormone 1		12	15	10	14	14
c. Hormone 2	18	12	20	16	14	80
d. Nitro 1	75	50	70	60	55	310
e. Nitro 2	60	50	50	55	45	260

65

1. Analyze the data by an analysis of variance.
2. Divide the sum of squares for treatment according to the following contrasts:
   a. Control against treatments
   b. Hormone against nitro
   c. Within hormone
   d. Within nitro
3. Test if these contrasts are orthogonal
4. Divide the error into:
   a. An error for the comparison between control and treated plots.
   b. An error for the comparison between treatments within treated plots.
5. Test the different effects against their respective errors and discuss the result.

Exercise 8

In exercise 3 the treatments 1-9 are limed according to the following table. The table give the treatment numbers referred to in exercise 3.

Quantity of lime	Type of lime		
	Powder	Crushed	Granules
1	1	4	7
2	2	5	8
3	3	6	9

a. Analyze the data of exercise 3 using this new information, and divide the treatment effects into the two main factors lime type and lime quantity and the interaction between them.

b. For the main effects, calculate and test those contrasts that are of interest.

c. Discuss the results.

Exercise 9

A split plot experiment with 3 fertilizer levels, A, B and C, 5 wheat varieties, 1, 2, 3, 4 and 5, and 4 replications, I, II, III and IV, gave the following results in kg grains per 36 m^2 plot.

The table is arranged according to the field map.

I		II		III		IV	
B 5	11,0	C 3	13,0	A 4	11,3	C 3	12,2
B 4	11,6	C 1	15,2	A 2	14,6	C 2	14,2
B 1	14,3	C 5	11,3	A 3	12,5	C 5	10,0
B 2	15,9	C 4	12,0	A 5	10,3	C 1	14,9
B 3	12,0	C 2	16,0	A 1	15,7	C 4	11,3
C 5	11,9	A 2	12,6	B 3	11,9	A 5	7,9
C 1	14,1	A 3	9,8	B 2	15,1	A 3	10,0
C 2	16,2	A 5	8,8	B 4	11,8	A 1	12,5
C 4	12,7	A 1	13,4	B 5	10,8	A 2	11,8
C 3	13,9	A 4	10,3	B 1	15,5	A 4	11,0
A 2	14,1	B 5	8,9	C 3	13,2	B 1	14,3
A 4	10,7	B 1	13,5	C 5	11,3	B 3	11,1
A 5	8,7	B 4	10,8	C 1	15,8	B 2	11,6
A 1	12,4	B 3	12,2	C 2	16,1	B 4	10,0
A 3	10,4	B 2	13,6	C 4	12,8	B 5	8,0

a) Analyze the experiment by an appropriate analysis of variance.

b) Calculate the standard error and the $LSD_{5\%}$ for the comparison between varietal means.

c) Test for linear and quadratic effect of the fertilizer treatment.

d) Present in tabular form the means for the significant effects and discuss the results.

Exercise 10

In a series of experiments with 5 spring wheat varieties, at 10 locations and with 2 replications per location, the following yields in kg per dekaar were obtained.

		Location									
Var.	Rep.	1	2	3	4	5	6	7	8	9	10
1. Runar	I	278	431	344	362	384	529	327	448	359	256
"	II	193	392	254	335	457	468	308	497	372	283
2. Reno	I	259	531	341	355	463	527	330	423	368	263
"	II	155	445	270	404	478	516	343	448	338	273
3. T0382	I	322	526	417	394	482	583	409	530	444	311
"	II	208	447	308	387	420	520	319	595	396	269
4. T9027	I	325	523	397	387	434	600	371	529	432	254
"	II	192	524	446	427	446	552	362	567	417	364
5. Sv505	I	257	470	394	383	450	536	386	493	415	334
"	II	205	499	372	414	411	496	321	513	425	429

1. Analyze the results from this series of experiments with an appropriate analysis of variance and discuss the results.
2. How many single experiments (locations) would be required for yield differences of 20 kg to be significant? (P = 5%)
3. How large would the $LSD_{5\%}$ have been if we had had 20 experiments without replication instead of 10 experiments with 2 replications?

Exercise No 11

The table below gives results from a series of experiments with 4 species of grasses, Rye-grass, Timothy, Meadow fescue, and Cocksfoot (species are indicated by the subscript $_i = 1 - 4$), and 2 seed rates, $_j = 1 - 2$, over 2 meadow years, $_k = 1 - 2$, and 4 locations, $_l = 1 - 4$. The results are given in kg dry matter per dekaar. In each location there were seed rates on main plots and species on subplots.

Loc.	Species	1. meadow year Seed rate 1	1. meadow year Seed rate 2	2. meadow year Seed rate 1	2. meadow year Seed rate 2
	Rye-grass	913	971	996	893
1	Timothy	846	723	746	690
	Meadow fescue	558	742	994	990
	Cocksfoot	860	700	1327	1192
	Rye-grass	1162	1200	1451	1667
2	Timothy	832	1043	1696	1545
	Meadow fescue	830	922	1500	1504
	Cocksfoot	1236	1279	2021	1665
	Rye-grass	1618	1452	829	819
3	Timothy	1330	1319	991	1011
	Meadow fescue	1353	1243	959	830
	Cocksfoot	1612	1791	1251	1050
	Rye-grass	1066	1254	464	626
4	Timothy	1129	1174	639	733
	Meadow fescue	996	1131	845	814
	Cocksfoot	1365	1281	921	1204

a. Analyze the results from these experiments by an appropriate analysis of variance.

b. Present significant effects in a table and discuss the results.

Exercise 12

Select a field crop that you know and design an experiment for the following experimental questions:

A: 5 varieties

B: 3 seeding rates or planting densities

C: 4 levels of fertilizer

Specify the plot size and quantities/densities of factor B and C.

Propose an analysis of variance with degrees of freedom for the different effects, and suggest error terms for the different main effects and interactions.