# THE USES OF STATISTICS IN FIELD EXPERIMENTS WITH RUBBER TREES.*

There is an old gibe that there are three grades of falsehood:— "Lies, Damned Lies and Statistics." The intention here is to show that statistics, properly used, so far from being the acme of falsehood, is indeed common-sense in a highly developed degree. Huxley defined Science as "organised common-sense," and this is a definition which is specially appropriate here, for the application of statistics to field experiments has made it possible to carry these out in a scientific manner not before possible, and to interpret their results with a high degree of scientific precision. Statistics have in fact brought a large amount of common-sense to bear on the conduct and interpretation of field experiments.

It must of course be understood that the term statistics is used here not in the loose sense of the mere collection and classification of numerical data, but in the specialised sense of the use of higher mathematics for the analysis and interpretation of data, and for the design of experiments. It is not intended to attempt to explain the mathematical principles involved or the manner of their use. That is the job of the higher mathematician himself — so much so that although other scientists have to understand the general principles of the application of statistics to the design and interpretation of experiments, it is nowadays best left to the specialized mathematical statistician to undertake the actual work of applying his methods. It is for that reason that the Rubber Research Institute has just appointed a Statistician to its staff.

## WHAT IS AN EXPERIMENT?

An experiment is defined in Webster as "A trial . . . . . made to confirm or disprove something doubtful" — and in the case of rubber-growing, the "something doubtful" is whether a certain method of treatment or a certain type of material is better than others, and therefore worth while to adopt.

---

* The substance of a lecture given at the Annual Conference of the Incorporated Society of Planters at Kuala Lumpur, 1937, by Mr. H. J. Page, Director of the Rubber Research Institute of Malaya.

Reprinted, with permission, from the *Journal of the Rubber Research Institute of Malaya*, 1938, Vol. 8, pp. 197-216.

Anyone can carry out an experiment, and can obtain results therefrom. But that is less than half the battle. The experiment may have been so badly designed or so badly conducted that it cannot possibly give results capable of valid interpretation; or it may have been well designed and executed, and yet the meaning of the results may be obscure or misunderstood.

Let us take as our first example an experiment intended to compare only two things, — or "treatments" as they are usually termed, — for instance to compare the yielding power of one clone with that of another, or with that of seedlings, or to ascertain the relative merits of two tapping systems, or the effect of a given fertiliser on the yield of rubber trees. What would appear at first sight to be the obvious way of doing this would be to take two adjacent plots of land as alike as possible, and to apply the two treatments one to each, and then to compare the results obtained.

Now the crux of the whole matter is that more often than not this is an entirely misleading and unreliable way of doing a field experiment, and the chief object of this paper is to explain why this is so.

## "SINGLE-PLOT" EXPERIMENTS.

The practical man is strengthened in his unwillingness to accept this verdict by the fact that the earliest and best known agricultural field experments were carried out in this way, as "single-plot" experiments: that is, with only one plot for each treatment. Many of the most important advances in scientific agricultural practice are based on the results of such "single-plot" experiments. The world-famous manuring experiment on Broadbalk Field at Rothamsted for instance, from the results of which our knowledge of the main principles of manuring was first derived, is a "single-plot" experiment, started in 1843 and is still going on.

The reason for this is that if the effects which the experiment is intended to study are big enough, "single-plot" experiments will establish them with sufficient certainty. The Broadbalk experiments just referred to showed increases in yield of the order of 50 or 100 per cent or more — in fact at the present time the plot receiving no manure gives barely 10 bushels, whilst the plot getting a liberal "complete" manuring gives over 30 bushels: an increase of about 200 per cent over the control! If we knew beforehand that the effects to be studied in our field experiments were always as big as that, we could probably do well enough with "single-plot" experiments, if indeed we should then bother to do any experiments at all! But unfortunately the effects we wish to measure nowadays are more often only a matter of differences of 10 or 20 per cent instead of 100 or 200 per cent. They may be no less important for all that. A 20 per cent increase in yield of rubber, yielding before treatment 400 lb. per acre, *i.e.* an increase of 80 lb. per acre, may represent a handsome, profit, perhaps 100 per cent on the cost of getting it. But an increase of 20 per cent usually cannot be detected with certainty from a "single-plot" experiment.

Two adjacent tasks, one of which is manured and the other not, may give yields of 400 and 480 lb. rubber respectively, and yet the difference may have nothing whatever to do with the manuring. Even if previous records for the two tasks indicated that they then both gave 400 lb., the difference of 80 lb. after manuring may have been due to other factors entirely unconnected with the manuring. Moreover, unless special steps have been taken to ensure that the previous records do actually represent the true yields of the two tasks, there is no guarantee that they do; for tappers are notoriously prone to level up their deliveries of latex from adjoining tasks. All the same, as explained later, the existence of previous records does help to improve the reliability and to simplify the layout of experiments on mature rubber. But for "single-plot" experiments for which suitable past yield-records are not available, such as comparisons of different planting material; or cultivation and manuring experiments on young rubber; one is entirely at the mercy of the variations in the fertility of the soil over the area chosen for the experiment, as well as many other varying factors outside one's control.

### PLOT VARIABILITY: "UNIFORMITY" TRIALS.

Before going any further to explain this, a few examples follow:

Nos. I-V show the yields from the untreated control plots from various experiments.

Nos. VI-VIII show the yields from adjoining plots, or "uniformity trials."

I

| | | | |
|---|---|---|---|
| | | | 973 |
| | 926 | | |
| | | 807 | |
| 944 | | | |

Difference between adjacent plots in centre = 14 per cent.

II

| | | |
|---|---|---|
| | | 511 |
| 680 | | |
| | 531 | |

Difference between adjacent plots = 25 per cent.

III

| | | |
|---|---|---|
| | | 202 |
| 211 | | |
| | 206 | |

Exceptional uniformity.

IV

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 119 | | 120 | | | | | 162 |

Right hand plot 35 per cent *in* excess of other two.

V

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 179 | 179 | | | 155 | 153 | | | 106 | 121 |

Good agreement of adjacent plots in two cases, but 14 per cent difference in the third case, and a pronounced "drift" from high yields on left to low yields on right.

VI

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 102 | 93 | 90 | 92 | 97 | 98 | 91 | 89 | 99 | 100 |

A "uniformity trial" showing fair agreement, but a range of 13 per cent between highest and lowest yields.

VII

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 3.5 | 3.5 | 4.3 | 3.5 | 3.3 | 3.3 | 3.9 | 4.3 |

Yields from adjacent tasks (in gms per 100 trees per tapping) Differences between adjoining tasks, and the same as percentages, are shown below.

| | | | | | | |
|---|---|---|---|---|---|---|
| Nil | 0.8 | 0.8 | 0.2 | Nil | 0.6 | 0.4 |
| — | 20 | 20 | 5 | — | 15 | 10% |

## VIII

A uniformity trial of 78 plots. The yields showed a total range of variation of nearly 60 per cent, and examples of yields from adjacent plots are as follows:—

| | |
|---|---|
| 73<br>133 | Difference = 58% of mean |
| 103<br>84 | ,, ,, 18% ,, |
| 84<br>112 | ,, ,, 28% ,, |
| 89<br>117 | ,, ,, 27% ,, |
| 106<br>79 | ,, ,, 29% ,, |
| 115<br>83 | ,, ,, 32% ,, |

These are all actual instances of the yields obtained from adjoining or closely situated tasks of rubber, to which no different experimental treatments have been applied, *i.e.*, which have been treated, as nearly as practicable, exactly alike. Admittedly some of these are chosen deliberately as horrid examples, but they are none the less actual instances of the variation that may occur. Others of the examples displayed show that sometimes there may be a high degree of uniformity. The point is that one never knows with even reasonable certainty what the extent of this natural variation is. All the areas in question were chosen to represent the most uniform land and rubber available on the estate. This is an effectual answer to the planter who insists on using only single plots; and who says: "The other chap may have to use an elaborate modern experiment to get results, but I know my soil and my rubber so well that I can rely on the results of single plots." That is as often as not just sheer nonsense. Agricultural scientists have found by bitter experience that even on the most level and uniform-looking land, with apparently uniform planting material, be it an annual crop or perennial, differences in yield of the order of those shown on the diagrams are only too commonly found. Orginally agricultural scientists were just as unwilling and reluctant as the farmer or planter, to adopt the modern type of elaborate layout of field experiments. The remark of a certain distinguished professor of agriculture: "Damn your duplicate plot; give me single plots and I know where I am," is typical of the attitude which has had to be given up after hard experience. He thought he knew where he was, but more often than not he would have been quite wrong.

The reason for this uncertainty is that with all natural material, living or inanimate, there is a greater or less degree of variability, which is not the case with pure substances. The chemist working with pure substances is largely free from this trouble. He does not have to do an experiment, for instance, to find out whether sulphuric acid and caustic soda will react together and neutralise one another. He knows that they always will; that 98 grams of $H_2OS_4$ will always be exactly neutralised by 80 grams of NaOH to give 142 grams of $Na_2SO_4$; and that a fixed and constant amount of heat will be liberated. All molecules of sulphuric acid or of caustic soda are chemically alike. But no two individuals of any living kind are exactly the same, nor are any two samples of natural material, and the behaviour of an individual or a sample cannot be predicted with certainty. It is this innate variability of natural material, and of the environmental conditions, which places experimentation with such materials, and particularly with living materials, on an entirely different plane from experimentation with pure substances. The statistician's job is to study variability and its effects, and it is for that reason that the assistance of statistics is so necessary in the design and interpretation of field experiments.

## FACTORS RESPONSIBLE FOR VARIATION.

Let us examine the factors responsible for variation in field experiments with rubber trees. In the first place there is the variation in "fertility" between different plots of land. In a properly designed experiment this is the biggest source of variation. Even the most uniform-looking area of land may possess quite big variations from one place to another. Even if the surface soil is quite flat and apparently uniform, there may be invisible differences in the nature and depth of the subsoil. The sporadic occurrence of a hard pan, at varying depths, is a special case of this, and instances often occur in which the existence of such a hard pan is first discovered only as a result of poor sickly growth of trees in patches. Moreover, an apparently uniform top soil, of uniform depth, may yet hide differences in fertility through differences in the previous croping history of the area; a part may have been heavily depleted of available plant food by previous cropping with, say, tapioca. On many estates it is practically impossible to find any level areas of land, and experimental areas have to be on hillsides, with considerable differences in the soil, and in conditions such as drainage, between the higher and the lower land.

Added to these inherent variations in the soil, are variations in the rubber trees. By choosing plots of the size of a task, with over 200 trees, the effect of the large tree-to-tree variability, which occurs especially with seedling rubber, is largely eliminated; and differences in age or origin of trees should always be avoided by choosing an area which was all planted at the same time with uniform material (except of course in an experiment for the comparison of different sorts of planting material). There still remain the effects of uneven attack by pests and diseases, of differences in tapping history or in average height of tapping, whilst there is also the possibility of lack of uniformity in the conduct of the experiment and in the collection of the results.

## SIZE OF UNCONTROLLABLE VARIATION.

The uncontrollable variation due to differences in soil and subsoil, — or "soil heterogeneity" as it is usually called, — will sometimes be slight and sometimes be large, whilst the other factors instanced above will similarly differ in their incidence from place to place and time to time. Thus the sum-total of the effects of all these factors, as reflected in the growth or yield data obtained from the plots, will also vary between wide limits, as shown by the examples above, where in some cases the yields show a very satisfactory approach to uniformity, as between different plots, whilst in other cases the differences may be over 25 per cent. The important point is that one can never be sure beforehand what the position is.

Suppose that the two plots A and B of example VII above had been chosen to compare two clones, and that a difference of 20 per cent had shown up between them in growth, or in yield when they came to tapping, then one would in fact be reasonably safe in drawing the conclusion that the two clones did actually differ, on that soil, by about 20 per cent in their growth or yield. On the other hand, suppose that plots C and D had been chosen, without the knowledge which the "uniformity" test has here revealed. The two clones might be identical in their inherent yielding capacity, and yet the natural "soil heterogeneity" between the two plots would result in a difference of 20 per cent, which would be quite erroneously interpreted as a difference in the clones themselves, if the results of only that one single-plot experiment were accepted as valid.

This is sufficient to demonstrate how unsafe it is to rely on the results of "single-plot" experiments — save under certain special circumstances mentioned later.

## REGULAR AND IRREGULAR VARIATION.

Uniformity trials such as those which are represented above, are trials in which the area in question is divided up into plots which are harvested separately, without any difference in treatment. These enable *fertility contour diagrams* to be drawn. These are similar to the height contours so familiar on ordinary maps. There are peaks of high fertility, hollows and valleys of low fertility, and intermediate slopes of average fertility. Now if changes of fertility of this sort occur *within* one experimental area, it is clear that the variation from plot to plot is irregular, though contiguous plots are likely to be less different than those more widely separated.

On the other hand the scale of the fertility contour map may be smaller, so that the whole of an experimental area falls on a patch of regularly varying fertility. The pitfalls into which a regular fertility gradient or "drift" of this kind can lead, are well illustrated by a case recorded some years ago, which is illustrated below.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 10 | 20 | 30 | 40 | 50 | | Cwt. of lime per acre |
| A | B | C | D | E | F | X | |
| 20 | 22 | 24 | 26 | 28 | 30 | | Yield. |

There were seven plots in a row, the first receiving no treatment, and five others receiving, in order, 10, 20, 30, 40 and 50 cwt. of lime per acre. The yields obtained may be represented by the the figures underneath the plots, *e.g.*, a steady increase in yield from 20 to 22, 24, 26, 28 and 30 units — apparently a clear case of increasing benefit to the yield from increasing amounts of lime. Fortunately, (or unfortunately!) a second control plot was included at the other end, marked X on the diagram. The yield from that control plot was 32! The lime had really no effect on yield at all; all that the experiment showed was a uniform fertility drift increasing from left to right. But for the after-thought of adding a duplicate control plot at the other end, this would not have been discovered, and it would have been concluded that lime had an increasingly beneficial effect on that soil, in dressings at least up to 50 cwt. per acre!

It might have happened, however, that the experiment was laid out on land with a peak of high fertility in the middle, in which case the two control plots one at each end would have given similar yields, and it would have been concluded that the land was of uniform fertility, unless another control plot had been included in the middle to reveal a yield much higher than those from the plots at the two ends.

One of the earliest methods of guarding against misleading results of this kind was to make every other plot a control plot, as below:—

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | B | A | C | A | D | A | E | A | F | A |

A = Control plots. B, C, D, etc. = Treated plots.

Then by comparing the yield of each treated plot with the average of the yields of the two control plots one on each side, a more accurate measure of the treatment was obtained. But what a wasteful and cumbrous method of experimenting! Half the total number of plots is given up untreated to controls, and there is still only one single plot for each treatment. All the same, this was a definite move in the right direction, for it introduced the principle of repetition, or "replication" as it is now called, (as a useful word to cover duplication, triplication, and so on, up to any desired degree of repetition).

## REPLICATION.

In view of the chances of variation between plots, it is obvious that if there are *several* of each treatment, distributed over the whole area, the *average* result obtained from all the replicate plots of any one treatment is much more likely to represent the real yield obtainable with that treatment under the average conditions of that area, and comparison between such averages for different treatments is far more likely to be a true indication of the relative effects of those treatments.

Replication is the first essential of a reliable field experiment, and in all modern experiments each treatment (including the control, if any) is replicated the same number of times. Let us take four-fold replication, (that is, quadruplication), as an example. In the following diagram is an experiment with four treatments, A, B, C and D, with four plots for each treatment.

| | | | |
|---|---|---|---|
| A | B | C | D |
| D | A | B | C |
| C | D | A | B |
| B | C | D | A |

This is typical of the kind of experiment first adopted after the need for replication came to be recognised. It includes adequate replication, but its arrangement is *systematic*. A, B, C and D always follow one another in the same order.

## EXPERIMENTAL ERROR.

Now we have seen that owing to "soil heterogeneity" and other uncontrollable factors, the results obtained from an experiment are lacking in accuracy to a greater or less degree. The extent to which the results are therefore not completely accurate is the experimental error, termed mathematically the "standard error" or the "probable error" according to the mathematical convention used to calculate it. It is expressed as ± X, where X is a number; the higher X is, the more inaccurate the experiment is and the less reliable are the results. A probable error of say ± 10 per cent applying to an average yield of say 100 units from a given treatment, means that owing to the influence of uncontrolled variation in the conditions of the experiment, that average is inaccurate, and that there is a good chance that the true value might be only 90 on the one hand, or 110 on the other, or anywhere between; it is impossible to say, between those limits just what it really should be.

The object of a properly-designed field experiment, so far as the experimental error is concerned, is not only to make that error as low as possible, but also, *equally important*, to make it possible to calculate that error accurately. The use to which the results of an experiment are put is to decide how the effects of different treatments compare. This involves studying the *differences* between the average results for the two treatments under comparison. Supposing one treatment gives an average of 100 and another an average of 120. If the "probable error" of each is ± 10 per cent, that means that the treatment with the average of 100 may be taken as being possibly anything from 90-110, and the other as anything from 108-132. The difference between the averages for the treatments may therefore be anything from −2 to +42! The apparent superiority of 20 per cent of the second treatment over the first has changed into the complete uncertainty, whether perhaps it is no better, or whether perhaps it is very much better. This is taking rather an extreme and crude example — the mathematician has more elegant and exact methods of treatment — but it serves to illustrate my point. Supposing the probable error has been wrongly estimated and that it is really only about 5 per cent instead of 10 per cent, what a difference that would make to the value of the results.

The fault of the systematic arrangement of plots with replication, is that whilst it reduces the experimental error, it does not allow the true value of that error to be calculated, and as just explained, it is not much use improving the accuracy of results if you do not know what that accuracy is.

## RANDOMISATION.

The other requirement of a properly laid out experiment, in addition to replication, is "*randomisation.*" That simply means that the arrangement of the plots is not systematic, but at random, —

according to chance. If the previous diagram is compared with the following one

| | | | |
|---|---|---|---|
| A | D | B | C |
| D | C | A | B |
| C | B | D | A |
| B | A | C | D |

it is seen that whilst in the first one, A, B, C and D always come in the same order in each row or column, in this one that is not so, — they are arranged at random in each. The order in which the treatments are assigned, in a randomised arrangement, is one of the many in which they might come if dealt from a shuffled pack of cards, or thrown with dice, or drawn from a hat.

It is a requirement of the mathematical statistics whereby the standard error is worked out, that if the estimate of that error is to be *valid* the arrangement of treatments in each experiment, or section thereof, must be at random, not systematic. We have seen already that it is not any use calculating the experimental error unless it is really an indication of the true error, and here it must be taken on trust that randomisation is essential for that purpose. The mathematical arguments underlying that fact are perfectly sound but far too involved to explain now.

## RANDOMISED BLOCKS AND LATIN SQUARES

The commonest and best-known arrangements in which the great majority of properly-laid-out field experiments are included, are *Randomised Blocks* and *Latin Squares*. Many will be familiar with these from the schemes of experiments they are doing in co-operation with the Rubber Research Institute, the Department of Agriculture or with other concerns. It is hoped that after the attempt to explain why they are necessary, and the further attempt later on to explain how they work, planters will be happier in their minds about what must often have seemed the rather unreasonable requirements and conditions they are asked to conform to in such experiments.

In the case of the *Randomised Block* experiments illustrated below.

*Two Examples of Randomised Block Experiments*

(Plot yields shown only for Control (A) Plots)

| | | | | | | | Block Totals |
|---|---|---|---|---|---|---|---|
| Block | 1 | A 17.5 | C | D | B | E | 102 |
| „ | 2 | E | C | B | D | A 27.2 | 122 |
| „ | 3 | E | B | D | A 18.7 | C | 127 |
| „ | 4 | A 21.8 | B | D | C | E | 128 |
| „ | 5 | D | C | E | A 23.6 | B | 134 |

Standard Error of Difference between Treatment Averages=8.8%
Significant Difference between Treatment Averages=17.6%

An example of a Randomised Block experiment showing rather a high experimental error, as was to be expected from the big variation in yields from Control (A) plots.

| | | | | | | | Block totals |
|---|---|---|---|---|---|---|---|
| Block | 1 | D | C | E | A 3.2 | B | 16.6 |
| „ | 2 | E | C | D | B | A 3.2 | 16.9 |
| „ | 3 | C | A 3.2 | D | B | E | 17,3 |
| „ | 4 | A 3.1 | D | B | C | E | 16.7 |
| „ | 5 | E | B | D | C | A 3.1 | 16.9 |

Standard Error of Difference between Treatment Averages=2.6%
Significant Difference between Treatment Averages=5.2%

An example of Randomised Block experiment with a low experimental error, as was to be expected from the uniformity in yields from the Control (A) plots, and in the Block totals.

We have five treatments, A, B, C, D, E, and fivefold replication, so that there are five separate blocks. In each block there are five plots, one for each treatment, and the allocation of these treatments is entirely at random, that is, in the order in which they might be drawn from a hat. The random arrangement is different in each block, representing a fresh drawing from the hat for each.

*An Example of a Latin Square Experiment*

R. R. I. Experiment Station Block 6, 1936

Letters = Treatments

Numbers = Plot yields

| | | | | | | | Row Totals |
|---|---|---|---|---|---|---|---|
| | B 391 | A 330 | C 374 | D 362 | F 339 | E 372 | 2168 |
| | C 414 | D 399 | F 384 | E 387 | B 392 | A 375 | 2351 |
| | F 395 | C 445 | E 336 | A 391 | D 362 | B 329 | 2258 |
| | D 393 | B 394 | A 383 | C 440 | E 361 | F 386 | 2357 |
| | A 378 | E 373 | B 391 | F 399 | C 390 | D 373 | 2304 |
| | E 337 | F 372 | D 350 | B 397 | A 352 | C 454 | 2262 |
| Column totals | 2308 | 2313 | 2218 | 2376 | 2196 | 2289 | 2283 |

| | | | | |
|---|---|---|---|---|
| Treatment Averages | A | ... | 368 | 97% |
| ,, ,, | B | ... | 382 | 100% |
| ,, ,, | C | ... | 420 | 110% |
| ,, ,, | D | ... | 373 | 98% |
| ,, ,, | E | ... | 361 | 95% |
| ,, ,, | F | ... | 379 | 100% |

Standard error of Difference between
Treatment Averages ... 3.44%
Significant Difference ... 7.0 %

The *Latin Square* is a slightly more restricted arrangement, in which the number of replications is the same as the number of treatments. We have such a $6 \times 6$ Latin Square illustrated above.

It is not necessarily a geometrical square, for the individual plots may be longer than they are broad, but it is called a square because it has the same number of plots, 6, in each direction. It may be regarded as six randomised blocks, either vertically or horizontally, and it is so in fact, in so far that the arrangement of the treatments in each row or each column is at random, but with this important restriction, that any one treatment must not occur more than once in any one row or in any one column.

An essential feature of experiments of this type is that the *total* yield for a block (in a Randomised Block experiment) or from a row or column (in a Latin Square experiment) is contributed to by each treatment alike. In the above Latin Square, one-sixth of the area of each row or column is untreated control (A), one-sixth has received treatment B, one-sixth treatment C, and so on; so that all treatments have an equal share in their effect on the block, row or column total. The comparison of these totals thus does not involve main differential treatment effects, and is an indication of the variation due to other uncontrolled factors. Variations in "soil heterogeneity" can be observed from variations in block, row or column total, and by a mathematical treatment of those totals, a large part of the error caused by "soil heterogeneity" and the like can be calculated and can be eliminated from the error attaching to the comparison of the treatment totals, so increasing the accuracy of the experiment.

## "SIGNIFICANT DIFFERENCE."

An attempt will now be made to explain what is meant by another term which is used very frequently in stating and interpreting the results of an experiment, the term "Significant Difference." To do this, it must first be remembered that the irregularities due to "soil heterogeneity" and other causes are distributed more or less at random over the experimental area, and that it is a matter of *chance* whether any particular plot will be on an area which is above or below, and much or little above or below, the average. The mathematical treatment of the results must therefore be one which deals with the assessment of *chances* or *odds*, or as the mathematician calls it, *probability*.

When a "standard error" is said to be $\pm$ 10 per cent that does not mean that the result could not possibly be even less accurate than that, but that the odds are in favour of the inaccuracy not exceeding $\pm$ 10 per cent. Since there are so many variable factors not under complete control in any experiment: soil, weather, disease, Tamil coolies, etc., and since the incidence of these factors can never be exactly known, there *must* always be *some* uncertainty regarding the accuracy of the results obtained, however well the experiment has been designed and carried out. The object of the application of statistics to the design and to the interpretation of field experiments is to reduce that uncertainty as far as possible, and to define it. This is done by stating the experimental error and the "significant difference" in terms of probability or odds. It is open to anybody to decide what degree of probability he is willing to accept as good enough for him in practice, and circumstances may often dictate whether long odds or short odds are worth while.

The odds that are usually accepted by the scientist in the interpretation of field experiments are 20 to 1, or 95 per cent probability. ($P=0.05$). As explained already, there is always a chance that the difference between the average results for two treatments, even from the best-designed and executed experiment, is caused by chance, *i.e.*, by the operation of the uncontrolled variables included in "soil heterogeneity," etc., and not by the difference in the treatments applied. If the mathematical analysis of the data shows that the odds are 20 or more to 1 that this is not so, *i.e.*, that there is at least 95 per cent probability that the difference is caused by the difference in the treatments, then that difference is said to be "significant," and the "significant difference" is the minimum difference allowing these minimum odds.

Any difference between treatment averages which is less than this "significant difference" is "insignificant," *i.e.*, it is not safe enough to ascribe it to the difference in treatments, because the odds against its being caused merely by chance are less than 20:1. Anybody who is prepared to take more of a chance, is perfectly at liberty to do so. He may wish to accept 10:1 odds instead of 20:1, in which case a smaller difference is accepted as caused by the treatments under test.

As in other walks of life, the amount of chance one is prepared to take depends on what is at stake. To take action on a low probability may be worth the risk when the result of the gamble is big enough if it succeeds, but anybody doing so must admit that it is a gamble, involving risks bigger than the scientist is justified in giving his blessing to — even though he also may sometimes be forced to take them. When therefore a difference is stated by the scientist to be "significant," remember that it does not mean that it *must* be a real and not a chance difference, but only that the risk of its being a chance difference is so small as to be fairly safely accepted. Similarly when the result is said to be "not significant," it does not mean that it cannot possibly be real, but only that the risk of its being only a chance difference is too big for safety.

## WHEN SINGLE-PLOT EXPERIMENTS MAY HAVE SOME VALUE.

It was mentioned earlier in this paper that if the difference is big enough, even "single-plot" experiments can be relied upon. Thus if the yield of one plot is double that of the control, the odds are pretty high that this is chiefly the effect of the treatment, but if the difference is only 10 or 20 per cent, the odds are perhaps 5 or 10 to 1 *against* its being due to anything except chance.

There are two other factors which may greatly reduce the uncertainty of the results from single plots. The first of these is the *time factor*. If single plots have been carried on for a number of years, with careful recording of yields all the time, it may be possible by studying the relative behaviour of the plots during that time, to find differences in trend; the yield of one plot may have gone steadily up and that of another steadily down. Statistical analysis of such results is possible, and may often yield results of fair reliability. There is in fact replication here, though it is replication in time, not in space. But we do not want to wait for years to get definite results, if we can possibly avoid it, so whilst this is some consolation to those who have records of single-plot experiments started some years ago, it is no recommendation, and is not intended to be, that single-plot experiments should be laid down now.

The other factor is the *personal factor*. The planter who knows his rubber well can often see effects which have not or cannot be accurately recorded. It may not need more than a single plot to enable, *e.g.*, the effect of manures on foliage and canopy to be clearly seen. If an area in which the trees had small sickly-coloured leaves, thin canopy and obvious "stag heads" and die-back, develops in a thick dark green canopy of large well-formed leaves, with lots of a new growth in a few months after manuring, so that its improved condition leaps to the eye in comparison with the area not manured, then it would be sheer nonsense to deny, because only a single plot had been used, that the improvement was practically certainly caused by the manuring. Less spectacular differences in bark condition may also be detected with reasonable certainty by the man who knows his trees well. Should such improvements later be paralleled by the recording of a higher yield from the manured plot, then it is more likely that this is an effect of the manuring, even if it cannot be taken as sufficiently probable to satisfy the statistician.

We must remember that the statistician is the servant, not the master, of the experimenter, in much the same way that the architect is the servant, not the master, of the owner of a building. We expect the architect to design for us a building which is safe, and suitable for the use to which it is intended to be put, and so long as the building is properly constructed to that design, he takes the responsibility for its safety and suitability; but if we ask for a design for a dwelling house and he submits one more suitable for a block of offices, we do not have to accept it, whilst if he submits alternative designs, it is for us and not for him to decide which shall be accepted. We are perfectly at liberty to erect a building without any design or advice from an architect (that is, assuming we are outside the jurisdiction of the Municipal Authorities!) provided we do so on our own responsibility and do not ask him to accept any.

In just the same way, the experimenter is obviously quite at liberty to carry out "single-plot" experiments, and to draw his own conclusions from them, perhaps merely because he has a "hunch" that the result is so and so, or because he knows from his own observation that there have been effects which could not be recorded in figures, so long as he does not expect the statistician to take any responsibility for them, and so long as he does not expect others blindly to accept his conclusions. The rigorous mathematical tests which the statistician applies, on the other hand, serve as a guarantee which others can readily accept, and so they enable experimental results to be of use to the whole body of planters. But such tests can be validly applied *only* if the experiment is properly laid out in the manner already exemplified.

## VARIATION BETWEEN LOCALITIES.
## FIELD EXPERIMENTS AND SOIL SURVEYS.

It goes without saying that the results of an experiment in one locality cannot be unreservedly applied to another where conditions of soil, climate, trees, etc., are different. Due allowance must always be made, if it can, for such differences and it is here that the close relation between soil surveys and field experiments is of such importance. The two should always go hand in hand, for when the soils of a country have been properly mapped as the result of a soil survey, it is possible to decide how far results obtained from experiments on a given type of soil are applicable in other places, according as the soils in those other places resemble or differ from that of the locality where the experiment was carried out.

Until such knowledge of soils is available as a result of a soil survey, advice on rubber cultivation in localities where the soil has not been examined and experimented on must always be of less sureness. The scheme of experimental planting under the ægis of the Rubber Research Institute, along with the soil survey work which is being done, is for that reason of the utmost importance, and may be expected to yield results of the greatest value to the rubber-growing industry of Malaya in due course.

## LOGICAL SELECTION OF PROBLEMS.

There is another aspect of field experiments with rubber; that of the choice of treatments in the scheme of the experiment. This refers more particularly to manuring experiments. There is sometimes a tendency to jump ahead to the study of matters of secondary importance before those of overruling primary importance have been elucidated. Thus experiments have sometimes been laid down to test, for instance, which form of phosphatic manure is best — say rock-phosphate or super-phosphate — or what is the most effective amount to use, — before it has been found out whether on the soil in question there is need for any phosphate at all. The first essential question is, does the use of nitrogen, or of phosphate, or of potash, produce any result. Only if any one or more of these is beneficial, is it worth while comparing the relative merits of different forms of nitrogenous, or phosphatic or potassic manures as the case may be, or the relative effects of small and larger dressings and so on.

At the same time, in view of the relatively long time which must elapse in experiments with rubber trees before results of reasonable certainty are obtained, it is only natural that one should wish to avoid having to tackle the problems only in this logical order one after the other. An important recent advance in the design of experiments is one which enables many of these different questions to be answered together by an experimental layout which is not much more complicated than that which is needed to answer only one set of problems at a time. Such an experiment has the further advantage that it may supply important information on what is known as the "interaction", of factors.

No attempt will be made here to explain in detail what is meant by this "interaction," but the kind of thing it covers is well enough illustrated by the following example: A soil might be so very deficient in phosphate that plants could not make more than very poor growth on that account, and so long as no phosphatic manure was given, nitrogen manures had little or no effect. The use of phosphatic manures would have a marked effect, but the much improved growth or yield thus resulting might *then* be prevented from being still better because the soil was not too well supplied naturally with nitrogen. The use of a nitrogenous manure along with phosphate would then produce a pronounced effect, additional to that produced by phosphate alone. That would be called a nitrogen-phosphate interaction, for the effectiveness of the nitrogen was subject to the presence or absence of phosphate. Nitrogen and phosphate together then give a bigger effect than the sum of the effects either gives in the absence of the other.

## PRACTICAL ASPECTS OF FIELD EXPERIMENTS WITH RUBBER TREES

It is not possible to deal here with certain special aspects on the practical side of the conduct of experiments on rubber, such as the size and shape of plots, the effect of individual variations between tappers, the effect of the average tapping height, avoidance of interference or "poaching" between adjacent plots, and so on. There is however one important aspect to mention. Hitherto it has been customary to use plots consisting of whole tasks or at most half-tasks owing to the difficulty of organizing the tapping of smaller units without great uncertainty as to the reliability of the results. However, even when the results are obtained in the form of the yield of dry rubber from normal-sized tasks, it is often impossible on estates to supervise the collection and carrying-in of the latex by the coolies well enough to be reasonably sure that there has been no levelling-up of the amount of latex from different tasks. Such levelling-up will, of course, tend to vitiate the results of the whole experiment.

Another service which statisticians have done to field experimentation is the study and development of sampling methods for determining the yield of plots, instead of weighing the whole produce from them. Such methods should be particularly applicable to, and valuable for, experiments with rubber trees, when the annual yield of a given task is made up of perhaps 250 individual tree-tappings on each of say 150 days, or 37,500 separate samples. Work is now being done at the Institute on this subject, and also by Dr. Haines of Dunlop Plantations, Ltd.

It may be expected that by collecting the latex separately from only perhaps one tree in every ten, or perhaps only one or two days a month, or less, considerably more accurate results from experiments will be obtained. Moreover, such a method would have the further advantage that individual plots in an experiment need no longer be as big as tasks or even half-tasks. Subject to their not being so small as to introduce extra troubles from individual tree-variation, they can be any size desired. For budded rubber in particular, where the uniformity from tree to tree is much higher than with seedlings, very considerable reduction in the size of plots may be possible.

## CONCLUSION.

In spite of some digression from the original theme of the reason why the modern type of layout of field experiment is necessary, and how the statistician's services can be of enormous value to field experimentation with rubber, it is hoped that it has been convincingly demonstrated that the old idea of "single-plot" experiments, which still dies hard with many planters, is really of very little use. As to how the modern type of experiment — Randomised Blocks, Latin Square and the more complex type I mentioned a short time ago — do what they have been stated to do, this is a matter which can be properly explained only in terms of rather abstruse mathematics, and in this place it must suffice to refer to the notes and explanation appended to the samples above.

In conclusion, it is apt to cite the words of Robert Boyle, written in 1673, under the title of "Concerning the Unsuccessfulness of Experiments" as quoted by R. A. Fisher at the commencement of his book on "The Design of Field Experiments":—

> "I am very sorry, Pyrophilus, that to the many (elsewhere enumerated) difficulties which you may meet with, and must therefore surmount, in the serious and effectual prosecution of experimental philosophy I must add one discouragement more, which will perhaps as much surprise as dishearten you; and it is, that besides that you will find (as we elsewhere mention) many of the experiments published by authors, or related to you by the persons you converse with, false and unsuccessful (besides this, I say), you will meet with several observations and experiments which, though communicated for true by candid authors or undistrusted eye-witnesses, or perhaps recommended by your own experience may, upon further trial, disappoint your expectation, either not at all succeeding constantly or at least varying much from what you expected."