



TWC/25/9 Rev.

ORIGINAL: English

DATE: October 3, 2007

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

**TECHNICAL WORKING PARTY ON AUTOMATION AND
COMPUTER PROGRAMS**

Twenty-Fifth Session
Sibiu, Romania, September 3 to 6, 2007

LEAST SIGNIFICANT DIFFERENCE (LSD) (REVISED)

Document prepared by an expert from Australia

Least Significant Difference (LSD)

LSDs are basically an extended Student's *t* test which uses a more comprehensive estimate of the “noise” or error in the data (the pooled error variance or the square of the more commonly calculated ‘mean square of the error’ or Error Mean Square) for making statistically valid mean comparisons. LSD (Least Significant Difference) is the value at a particular level of statistical probability (e.g. $P \leq 0.01$ - means with 99% accuracy) when exceeded by the difference between two varietal means for a particular characteristic, then the two varieties are said to be distinct for that characteristic at that or lesser levels of probability.

The difference between two means is declared significant at any desired level of significance if it exceeds the value derived from the general formula:

$$LSD = \frac{t(s\sqrt{2})}{\sqrt{n}}$$

Usefulness of LSDs

LSDs allow data to be eyeballed, without having training in statistics first. LSD is a simple calculation that allows the means of two or more pre-determined varieties to be compared. At a glance, the probability that the difference between the means is the result of chance can be evaluated and confidence is gained that the inference(s) drawn from the data are correct.

For plant breeders' rights (PBR), LSDs are very useful in determining the distinctness of a candidate variety from successive comparator varieties considering two varieties (candidate and one of the comparators) at a time. It is not legitimate to use the LSD when one or more candidates need to be simultaneously compared with multiple comparators.

Limitations to using LSDs

Like many statistics LSDs have restrictions on their use. However, if the following precautions and recommendations are considered LSDs are appropriate.

- LSDs must not be applied unless the F-test (or *t* test) indicates that significant differences between means are present.
- LSDs are only valid for testing mean comparisons that were pre-determined in the objective of the experiment.
- LSD is reasonably satisfactory for comparing each variety individually (when drawn from a set of varieties) with a standard control (for PBR the candidate variety would be considered the control).
- In its strictest sense, LSDs are at their most robust if, when more than two varieties are involved, the LSD is used to compare adjacent variety means when variety means are arranged in order of magnitude.

Calculating LSDs at $P \leq 0.01$ for Non Replicated Trials

For examining distinctness, it is essential to calculate LSDs at 1% ($P \leq 0.01$) level of significance. Anything beyond this level (for e.g. 5% - $P \leq 0.05$) is considered as non-significant for PBR purposes. There are many ways to calculate LSDs depending on the test situation. For a non-replicated trial, four different cases are discussed below for calculating LSDs. All these methods are basically extensions of Student's t-test

Finding the right value of t for LSD calculations:

The t tables found in the back of many statistical texts are often reprinted from Fisher and Yates *Statistical Tables for Biological, Agricultural and Medical Research*. Most are "two tailed" tables. This allows the difference between means to be tested regardless of direction (i.e. is mean A different from mean B regardless of whether mean A is greater than or less than mean B).

However PBR requires that the direction of the difference between the means be nominated (e.g. mean A greater than mean B). This restricts the t test to a "one tailed" test and changes the way that the correct t value is found from most tables.

Using a "two tailed table", the t values are found from the left hand column which is the degrees of freedom (d.f.). Deciding how to get the correct degrees freedom is discussed individually for each of the different test cases below. Now move across the columns to the appropriate probability value. In this case the required probability is $P \leq 0.01$. However as a "one tailed" test is required, the correct value (i.e. the heading of appropriate column from a two tailed table) is:

$$\begin{aligned} \text{Required column} &= \text{required P value} \times 2 \\ &= 0.01 \times 2 = 0.02 \end{aligned}$$

Testing of equality of variances: The following procedure can be used to test the equality of variances of the candidate and the comparator and to find which of the four alternatives discussed below should be used:

Calculate

$$F = \frac{\text{larger value } s^2}{\text{smaller value } s^2}$$

and compare this value with the tabulated F value (refer to an F table in a statistical table) at $P \leq 0.01$ reading degrees of freedom ($n_1 - 1$) horizontally, degrees of freedom ($n_2 - 1$) vertically.

where,

s^2 = square of the respective sample standard deviation (σ_{n-1} in your calculator) which is the sample variance

$(n_1 - 1)$ = degrees of freedom for the **numerator** s^2

$(n_2 - 1)$ = degrees of freedom for the **denominator** s^2

If calculated **F** is more than the tabulated **F** value, then the variances are unequal and then use either Case 3 or Case 4; if it is less, use either Case 1 or Case 2, depending on the equality/non-equality of sample sizes.

Case 1: Comparison of means of a candidate and a comparator for a characteristic with **unequal number of samples and assuming equal variances**

The test: when $n_1 \neq n_2$ (i.e., the sample numbers of candidate and the comparator are not same) calculate s_d (pooled standard deviation of the difference between the means) using the following formula:

$$s_d = \sqrt{s^2 (1/n_1 + 1/n_2)}$$

$$\text{where } s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

and LSD is = $t_{0.01} \times s_d$

in which

LSD = Least Significant Difference between the means at the required level of probability (e.g. **P≤0.01**)

n_1 = number of observations for the candidate variety

n_2 = number of observations for the comparator variety

s_1 = sample standard deviation for the candidate variety (get σ_{n-1} in the statistical mode if you are using a calculator)

s_2 = sample standard deviation for the comparator variety

s_d = pooled standard deviation of difference between the means

$t_{0.01}$ = tabulated 't' value (one tailed) at probability level $P \leq 0.01$ with degrees of freedom ($n_1 + n_2 - 2$)

(Look into tabulated Student's 't' values in any book on Statistics).

Case 2: Comparison of means of a candidate and a comparator for a characteristic with **equal number observations and assuming equal variances**

The test: when $n_1 = n_2 = n$ (i.e., the sample numbers of the candidate and the comparator are the same)

Calculate:

$$s_d = \sqrt{2s^2 / n} \quad \text{and} \quad \text{LSD} = t_{0.01} \times s_d$$

however,

$t_{0.01}$ = tabulated 't' value (one tailed) at probability level $P \leq 0.01$ with degrees of freedom **2 (n-1)**

Case 3: Comparison of means of a candidate and a comparator for a characteristic with **unequal number of samples and assuming unequal variances**

The test: when $n_1 \neq n_2$ (i.e., the sample numbers of candidate and the comparator are not the same)

Calculate:

$$s_d = \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

and also calculate

$$t' = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$

where $w_1 = s_1^2/t_1$ and $w_2 = s_2^2/t_2$

in which

$t_1 =$ **Student's t** value for **($n_1 - 1$)** degrees of freedom at $P \leq 0.01$ (refer to Student's t table)

$t_2 =$ **Student's t** value for **($n_2 - 1$)** degrees of freedom at $P \leq 0.01$ (refer to Student's t table)

and

$$LSD = t' \times s_d$$

Case 4: Comparison of means of a candidate and a comparator for a characteristic with **equal number of samples and assuming unequal variances**

The test: when $n_1 = n_2 = n$ (i.e., the sample numbers of candidate and the comparator are the same)

the t' is the tabular t for **($n-1$)** degrees of freedom without any further calculation. Then calculate

$$LSD = t' \times s_d$$

Determining significance

The difference between two mean values is compared to the LSD value. If the difference is greater than the LSD value, then the means are significantly different.

Calculating LSDs at $P \leq 0.01$ for Replicated Trials

For calculating LSDs at $P \leq 0.01$ for a replicated trial a fair amount of calculation is needed. Two worked examples are given below, the first with equal number of replicates and second with unequal number of replicates:

Example No: 1: Calculation of LSD using a replicated trial having varieties in **equal number of replicates:**

We want to compare the mean leaf width of a new variety (A) with two comparators (B & C). The trial consists of 4 replications (replicates) arranged in a randomized complete block (RCB) design. This example assumes that there are no data missing.

The data is presented as:

	leaf width (mm)			
	Replicates			
	1	2	3	4
Variety A	17	15	10	14
Variety B	34	26	23	22
Variety C	23	21	8	16

First, calculate the variety totals and means. Then, calculate the replicate totals.

	1	2	3	4	Variety totals	Variety means
Variety A	17	15	10	14	56	14
Variety B	34	26	23	22	105	26.25
Variety C	23	21	8	16	68	17
Replicate totals	74	62	41	52		
					229	Grand total

The total number of observations is 3 varieties \times 4 replicates = 12

As indicated, the LSD formula is derived from the Student t test. There are many different representations of the formula for the Student t test, however the one presented below seems to be the most widely used.

$$\text{where } t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{2}/\sqrt{n}}$$

$$\text{this can be re-written as } \bar{X}_1 - \bar{X}_2 = \frac{t(s\sqrt{2})}{\sqrt{n}}$$

$$\text{hence } LSD = \frac{t(s\sqrt{2})}{\sqrt{n}}$$

where

t = tabulated t value at probability level (e.g. $P \leq 0.01$) and with degrees of freedom equal to (number of varieties minus 1) \times (number of measurements for each variety minus 1)

\bar{x}_1 = mean of variety 1

\bar{x}_2 = mean of variety 2

s = standard deviation of all the plots. This is NOT the standard deviation of the varieties. It is the $\sqrt{\text{mean square of the error}}$.

n = number of measurements (observations) in each variety, usually equal to the number of replicates (in this example n = 4).

However, before the LSD can be calculated, a number of other calculations aimed at estimating the mean square of the error (MSE) must be performed. To be consistent with most statistical texts, varieties will be referred to as treatments and replicates as blocks.

Most modern spreadsheet packages have two way analysis of variance (ANOVA) functions that calculate the MSE. However for completeness, the long hand calculations are set out below.

$$MSE = \frac{SSE}{(blocks - 1)(treatments - 1)}$$

However, as

$SSE = \text{Total SS} - SSB - SST$ more calculations are required.

$$\text{Total SS} = (\text{sum of squares of all data values}) - \frac{(\text{grand total})^2}{\text{total number of observations}}$$

$$= (17)^2 + (15)^2 + \dots + (16)^2 - (229)^2 / 12$$

$$= \mathbf{574.917}$$

$$SSB = \frac{(\text{sum of squares of block totals})}{\text{number of observations in a single block total}} - \frac{(\text{grand total})^2}{\text{total number of observations}}$$

$$= \frac{(74)^2 + (62)^2 + (41)^2 + (52)^2}{3} - \frac{(229)^2}{12}$$

$$= \mathbf{198.250}$$

$$SST = \frac{(\text{sum of squares of treatment totals})}{\text{number of observations in a single treatment total}} - \frac{(\text{grand total})^2}{\text{total number of observations}}$$

$$= \frac{(56)^2 + (105)^2 + (68)^2}{4} - \frac{(229)^2}{12}$$

$$= \mathbf{326.167}$$

$$MST = \frac{SST}{(v-1)}$$

$$= 326.167/2 = \mathbf{163.08}$$

$$\begin{aligned} SSE &= \text{Total SS} - \text{SSB} - \text{SST} \\ &= 574.917 - 198.250 - 326.167 = \mathbf{50.5} \end{aligned}$$

$$\text{thus } MSE = \frac{50.5}{(4-1)(3-1)}$$

$$= \mathbf{8.417}$$

$$\mathbf{F \text{ ratio}} = \frac{MSV}{MSE} = 163.08/8.417 = \mathbf{19.37}$$

Compare this calculated F with tabulated F at (v-1), (b-1)(v-1) df, i.e. at 2, 6 df ($P \leq 0.01$) = **10.92**

The calculated F value 19.37 is **more than** the table F value 10.92. This reveals that there are significant differences among the three varieties i.e. **at least one variety is different from any of others**. The next step is to calculate LSD.

$$\text{Standard deviation per plot} = \sqrt{MSE} = 2.901$$

The tabular value of t for a “one tailed” test with d.f. = 6 is 3.143

$t = 3.143$ (from Student’s t tables @ 6 d.f. and P (or α) ≤ 0.01)

$s = \sqrt{MSE} = 2.901$ (from the calculation shown above)

$n = 4$

$$LSD = \frac{t(s\sqrt{2})}{\sqrt{n}}$$

and in this example, $LSD(P \leq 0.01) = \mathbf{6.447}$

Determining if the two varietal means are significantly different:

The difference between two mean values is compared to the LSD value. If the difference is greater than the LSD value, then the means are significantly different. The variety means for the above example are A (candidate)= 14, B (comparator 1)=26.25, C (comparator 2)=17.

The absolute difference between A and B is **12.25** which is greater than 6.447: therefore A and B are **significantly different** at $P \leq 0.01$. This **confirms** the results of F test.

The difference between A and C is **3** which is less than 6.447: therefore A and C are **not significantly different** at $P \leq 0.01$.

Example No. 2 : Calculation of LSD using a replicated trial having varieties in **unequal number of replicates:**

	leaf width (mm)			
	Replicates			
	1	2	3	4
Variety A	17	15		
Variety B	34	26	23	22
Variety C	23	21	8	

First, calculate the variety totals and means. Then calculate the grand total.

	1	2	3	4	Variety totals	Variety means
Variety A	17	15			32	16
Variety B	34	26	23	22	105	26.25
Variety C	23	21	8		52	17.33

189 Grand total

The total number of observations is $2 + 4 + 3 = 9$

$$\begin{aligned} \text{Total sums of squares (TSS)} &= (17)^2 + (15)^2 + \dots + (8)^2 - (189)^2/9 \\ &= 4393 - 3969 = \mathbf{424} \end{aligned}$$

$$\begin{aligned} \text{Variety sums of squares (VSS)} &= (32)^2/2 + (105)^2/4 + (52)^2/3 - (189)^2/9 \\ &= 1024/2 + 11025/4 + 2704/3 - 33721/9 \\ &= 512 + 2756.25 + 901.33 - 3969 \\ &= 4169.58 - 3969 = \mathbf{200.58} \end{aligned}$$

$$\text{Error Sums of squares (SSE)} = \text{TSS} - \text{VSS}$$

$$= 424 - 200.58 = \mathbf{223.42}$$

$$\text{MSV} = \frac{\text{VSS}}{(v-1)} \text{ or } \text{VSS}/\text{variety df}$$

$$= 200.58/2 = \mathbf{100.29}$$

$$\text{MSE} = \frac{\text{SSE}}{(t-1) - (v-1)} \text{ or } \text{SSE}/\text{error df}$$

$$= 223.42 / (8) - (2) = \mathbf{37.23}$$

$$\mathbf{F \text{ ratio}} = \frac{\text{MSV}}{\text{MSE}} = 100.29/37.23 = \mathbf{2.69}$$

Compare this calculated F with tabulated F at (v-1), (t-1)-(v-1) df

i.e. at 2, 6 df ($P \leq 0.01$) = **10.92**

The calculated F value 2.69 is **less** than the table F value 10.92. This reveals that there are no statistically significant differences among the varieties. Hence there is no need to calculate LSD.

However, to demonstrate how to calculate LSD in this case, the following procedure is done:

Standard deviation of varietal difference $S_{d(ij)}$

$$= \sqrt{\text{MSE} (1/r_i + 1/r_j)}$$

where

r_i = number of replications for the first variety (for e.g. the candidate)

r_j = number of replications for the second variety (for e.g. one of the comparators)

Then

$$\text{LSD}(ij) = S_{d(ij)} \times t_{(\text{error df}) (p=0.01)}$$

Determining if the two varietal means are significantly different:

The difference between two mean values is compared to the LSD value. If the difference is greater than the LSD value, then the means are significantly different. The variety means for the above example are A (candidate)=16, B(comparator 1)=26.25, C(comparator 2)=17.33

LSD for comparison of varietal means of A and B in the above table

$$= \sqrt{37.23 (1/2 + 1/4)} \times 3.143$$

$$= 5.28 \times 3.143 = \mathbf{16.61}$$

The absolute difference between A and B is **10.25** which is smaller than 16.61: therefore, A and B are not significantly different at $P \leq 0.01$.

LSD for comparison of varietal means of A and C in the above table

$$= \sqrt{37.23 (1/2 + 1/3)} \times 3.143$$

$$= 5.56 \times 3.143 = \mathbf{17.47}$$

Obviously the difference between the varietal means A and C (**=1.33**) is not statistically significant. This confirms the results of F test.

[End of document]