Application of Least Significance Difference (LSD)

We use the analysis of variance (ANOVA) to determine if there is a significant effect of a factor. However, if there are more than 2 instances of the factor, then when a significant difference is found, ANOVA does not tell us which of those instances contributes significantly to the difference.

For example, suppose we use ANOVA to decide if there is a significant difference in reported values amongst 5 laboratories (k =5) in the analysis of a whole milk sample for its protein level with 4 replicates (n' = 4) each. In this case, the factor is "laboratory". In here, the total number of data under consideration is n = 4 x 5 = 20. The question is: if the ANOVA did conclude there was a significant difference, was there just one laboratory that was different from the other four, or were there more?

A simple approach to identify which laboratory or laboratories are different in their mean results is by the method of least significant difference (*LSD*). It is used to compute the smallest difference (i.e., the *LSD*) between two mean values as if these means have been the only means to be compared with a Student's *t*-test. We can declare the difference between the two means to be significantly different when the t-value calculated is larger than the *LSD*.

Suppose we carry out a hypothesis testing on two reported average results, $\bar{x}_1$ and $\bar{x}_2$ with n_i replicates, where $\bar{x}_1$is the smallest value amongst the means reported.

The relevant hypothesis is:

$H_o : \bar{x}_1 = \bar{x}_2$ or $H_o : \bar{x}_1 - \bar{x}_2 = 0$,

$H_1 : \bar{x}_1 - \bar{x}_2 \neq 0$,

The Student's t-test is carried out by equation:

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

The t_{obs} value should be compared to the $t_{(df:\, n\text{-}k)}$ distribution, and the difference is significant on the 5% significance level if and only if $|t_{obs}| > t_{(\alpha=0.025,df=n\text{-}k)}$.
This is the case if and only if

$$|\bar{x}_1 - \bar{x}_2| \geq t_{(0.025,n-k)}.\, s\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

The right-hand side of this equation is called the least significant difference – or the 95% LSD-value – for the difference between these two mean values.

If the numbers of repeats, n_i in all the laboratories are the same, i.e. $n_1 = n_2 = n'$, then the LSD-value is:

$$LSD_{0.95} = t_{0.025,n-k}.\, s\sqrt{\frac{2}{n'}}$$

where s is the standard deviation of the means. This equation is the same for all pairs of $\bar{x}'s$.

Now, we can compare differences of the $\bar{x}'s$ to the LSD-value, and see if there are significant differences.

Worked example

Five test laboratories took part in an inter-laboratory comparison exercise on the analysis of a whole milk sample for its protein content (%m/m) with four replicated results reported by each participant. The results collated are summarized below, with one-way (one-factor) analysis of variance (ANOVA) as carried out by the Excel Data Analysis Tools.

Trial #	Lab A	Lab B	Lab C	Lab D	Lab E
1	3.08	3.34	3.21	3.27	3.35
2	3.18	3.25	3.36	3.32	3.38
3	3.04	3.29	3.42	3.38	3.39
4	3.09	3.36	3.05	3.12	3.42

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Lab A	4	12.39	3.098	0.0035
Lab B	4	13.24	3.310	0.0025
Lab C	4	13.04	3.260	0.0274
Lab D	4	13.09	3.273	0.0124
Lab E	4	13.54	3.385	0.0008

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Labs	0.17825	4	0.0445625	4.787	0.0109	3.056
Within Labs	0.13965	15	0.00931			
Total	0.3179	19				

As the F-value was significantly different from the F critical value at $\alpha = 0.05$ with 4 and 15 degrees of freedom, it was concluded that there were significant differences amongst the laboratory results with 95% confidence. To find out which laboratory's mean result was different from the rest, the LSD method was used.

The reported mean results were first re-arranged in ascending order:

Lab A	Lab C	Lab D	Lab B	Lab E
3.098	3.260	3.273	3.310	3.385

The standard deviation of the means, s, calculated was 0.1054 and the t-critical value at (20-5) or 15 degrees of freedom at α=0.025 (2-tailed) is 2.131. The LSD-value calculated is, therefore, $LSD = 2.131 \times 0.1054 \times \sqrt{(2/4)} = 0.159$.

Now, the differences between paired results are found to be:

|Lab A – Lab C| = 0.162;

|Lab C – Lab D| = 0.013;

|Lab D – Lab B| = 0.037;

|Lab B – Lab E| = 0.075.

It is obvious therefore that the difference between Lab A and Lab C is significantly larger than *LSD*-value of 0.159, indicating that Lab A's reported mean value was at odds with the other participants.

One must take note of some limitations when the LSD method is to be applied, namely:

1. The *LSD* is valid only when the *F*-test under ANOVA confirms that significant differences between sample means do exist;
2. The *LSD* value is most robust when we compare the adjacent mean values, i.e., when the sample means under study are arranged in the order of magnitude.