In Neil Salkind (Ed.), *Encyclopedia of Research Design.*
Thousand Oaks, CA: Sage. 2010

Tukey's Honestly Significant Difference (HSD) Test

Hervé Abdi · Lynne J. Williams

1 Overview

When an analysis of variance (ANOVA) gives a significant result, this indicates that at least one group differs from the other groups. Yet, the omnibus test does not inform on the pattern of differences between the means. In order to analyze the pattern of difference between means, the ANOVA is often followed by specific comparisons, and the most commonly used involves comparing two means (the so called "pairwise comparisons").

An easy and frequently used pairwise comparison technique was developed by Tukey under the name of the *honestly significant difference* (HSD) test. The main idea of the HSD is to compute the honestly significant difference (*i.e.,* the HSD) between two means using a statistical distribution defined by Student and called the q distribution. This distribution gives the *exact* sampling distribution of the largest difference between a set of means originating from the same population. All pairwise differences are evaluated using the same sampling distribution used for the largest difference. This makes the HSD approach quite conservative.

Hervé Abdi
The University of Texas at Dallas

Lynne J. Williams
The University of Toronto Scarborough

Address correspondence to:
Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/∼herve

2 Notations

The data to be analyzed comprise A groups, a given group is denoted a. The number of observations of the a-th group is denoted S_a. If all groups have the same size it is denoted S. The total number of observations is denoted N. The mean of Group a is denoted M_{a+}. Obtained form a preliminary ANOVA, the error source (*i.e.,* within group) is denoted $\mathcal{S}(\mathcal{A})$, the effect (*i.e.,* between group) is denoted $\mathcal{A}$. The mean square of error is denoted $MS_{S(A)}$ and the mean square of effect is denoted MS_A.

3 Least significant difference

The rationale behind the HSD technique comes from the the observation that, when the null hypothesis is true, the value of the q statistics evaluating the difference between Groups a and a' is equal to

$$q = \frac{M_{a+} - M_{a'+}}{\sqrt{\frac{1}{2} MS_{S(A)} \left(\frac{1}{S_a} + \frac{1}{S_{a'}} \right)}} \ , \tag{1}$$

and follows, a studentized range q distribution with a range of A and $N-A$ degrees of freedom. The ratio t would therefore be declared significant at a given α level if the value of q is larger than the critical value for the α level obtained from the q distribution and denoted $q_{A,\,\alpha}$ where $\nu = N - A$ is the number of degrees of freedom of the error, and A is the range (*i.e.,* the number of groups). This value can be obtained from a table of the Studentized range distribution. Rewriting Equation 1 shows that a difference between the means of Group a and a' will be significant if

$$|M_{a+} - M_{a'+}| > \text{HSD} = q_{A,\,\alpha} \sqrt{\frac{1}{2} MS_{S(A)} \left(\frac{1}{S_a} + \frac{1}{S_{a'}} \right)} \tag{2}$$

When there is an equal number of observation per group, Equation 2 can be simplified as:

$$\text{HSD} = q_{A,\,\alpha} \sqrt{\frac{MS_{S(A)}}{S}} \tag{3}$$

In order to evaluate the difference between the means of Groups a and a', we take the absolute value of the difference between the means and compare it to the value of HSD. If

$$|M_{a+} - M_{a'+}| \geq \text{HSD} \ , \tag{4}$$

Table 1 Results for a fictitious replication of Loftus & Palmer (1974) in miles per hour

	Contact	Hit	Bump	Collide	Smash
	21	23	35	44	39
	20	30	35	40	44
	26	34	52	33	51
	46	51	29	45	47
	35	20	54	45	50
	13	38	32	30	45
	41	34	30	46	39
	30	44	42	34	51
	42	41	50	49	39
	26	35	21	44	55
$M_{.+}$	30	35	38	41	46

then the comparison is declared significant at the chosen α-level (usually .05 or .01). Then this procedure is repeated for all $\frac{A(A-1)}{2}$ comparisons.

Note that HSD has less power than almost all other post-hoc comparison methods (*e.g.,* Fisher's LSD or Newmann-Keuls) except the Sheffé approach and the Bonferonni method because the α level for each difference between means is set at the same level as the largest difference.

4 Example

In a series of experiments on eyewitness testimony, Elizabeth Loftus wanted to show that the wording of a question influenced witnesses' reports. She showed participants a film of a car accident, then asked them a series of questions. Among the questions was one of five versions of a critical question asking about the speed the vehicles were traveling:

1. How fast were the cars going when they *hit* each other?
2. How fast were the cars going when they *smashed into* each other?
3. How fast were the cars going when they *collided with* each other?
4. How fast were the cars going when they *bumped* each other?
5. How fast were the cars going when they *contacted* each other?

The data from a fictitious replication of Loftus' experiment are shown in Table 1. We have $A = 4$ groups and $S = 10$ participants *per* group.

The ANOVA found an effect of the verb used on participants' responses. The ANOVA table is shown in Table 2.

Table 2 ANOVA results for the replication of Loftus and Palmer (1974).

Source	*df*	*SS*	*MS*	*F*	*Pr*(*F*)
Between: $\mathcal{A}$	4	1,460.00	365.00	4.56	.0036
Error: $\mathcal{S}(\mathcal{A})$	45	3,600.00	80.00		
Total	49	5,060.00			

Table 3 HSD. Difference between means and significance of pairwise comparisions from the (fictitious) replication of Loftus and Palmer (1974). Differences larger than 11.37 are significant at the $\alpha = .05$ level and are indicated with *, differences larger than 13.86 are significant at the $\alpha = .01$ level and are indicated with **.

	Experimental Group				
	$M_{1.+}$ Contact 30	$M_{2.+}$ Hit 1 35	$M_{3.+}$ Bump 38	$M_{4.+}$ Collide 41	$M_{5.+}$ Smash 46
$M_{1.+} = 30$ Contact	0.00	5.00 *ns*	8.00 *ns*	11.00 *ns*	16.00**
$M_{2.+} = 35$ Hit		0.00	3.00 *ns*	6.00 *ns*	11.00 *ns*
$M_{3.+} = 38$ Bump			0.00	3.00 *ns*	8.00 *ns*
$M_{4.+} = 41$ Collide				0.00	5.00 *ns*
$M_{5.+} = 46$ Smash					0.00

For an α level of .05, the value of $q_{.05,A}$ is equal to 4.02 and the HSD for these data is computed as:

$$\text{HSD} = q_{\alpha,A}\sqrt{\frac{MS_{S(A)}}{S}} = 4.02 \times \sqrt{8} = 11.37 \ . \tag{5}$$

The value of $q_{.01,A} = 4.90$, and a similar computation will show that, for these data, the HSD for an α level of .01, is equal to $\text{HSD} = 4.90 \times \sqrt{8} = 13.86$.

For example, the difference between $M_{\text{CONTACT}+}$ and $M_{\text{HIT}+}$ is declared non significant because

$$|M_{\text{CONTACT}+} - M_{\text{HIT}+}| = |30 - 35| = 5 < 11.37 \ . \tag{6}$$

The differences and significance of all pairwise comparisons are shown in Table 3.

Related entries

Analysis of variance, Bonferroni procedure, Fisher's least significant difference (LSD) test, Multiple comparison test, Newman-Keuls test, Pairwise comparisons, Post-hoc comparisons, Scheffe's test.

Further readings

Abdi, H., Edelman, B., Valentin, D., & Dowling, W.J. (2009). *Experimental Design and Analysis for Psychology*. Oxford: Oxford University Press.

Hayter, A.J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association*, **81**, 1001–1004.

Seaman, M.A., Levin, J.R., & Serlin, R.C. (1991). New developments in pairwise multiple comparisons some powerful and practicable procedures. *Psychological Bulletin*, **110**, 577–586.