

MS Module 11: Single-Factor ANOVA – Tukey's honestly statistical difference (overview)

(The attached PDF file has better formatting.)

Reading: §11.2, Multiple Comparisons in ANOVA

An actuary analyzing accident frequencies or mortality rates by location has two concerns:

- whether location is a good predictor
- which locations differ significantly.

Comparing each pair of locations by a t test or by an ANOVA exacerbates Type I errors. If one has 100 pairs of locations, then even if the null hypothesis is true (all locations have the same mean), about 5 of the pairs will differ significantly at a 5% confidence level by random fluctuations.

If the pairs were independent, one might adjust the confidence levels so that the probability of a Type I error for at least one pair is 5% (or whatever value one uses). But the pairs are not independent, since an outlier in one group creates many pairs of apparently significant differences.

Tukey's honestly statistical difference incorporates this dependence in its test. A final exam problem may ask to calculate the w value or a critical value for Q , the studentized range distribution (from a table).

Know also the sub-section "Confidence Intervals for Other Parametric Functions." Territorial and age group clustering are used in actuarial pricing and risk classification. Mortality rates differ by age group, and motor insurance accident frequencies differ by territory; actuaries group ages and zip codes (or addresses) into age groups and territories. An actuary with data by census tract may

- compare each pair of census tracts (Tukey's honestly statistical difference)
- compare the urban tracts vs the rural tracts (other parametric functions)

The actuarial tasks are more complex than the examples in the textbook. Comparing existing territories is a statistical problem. Forming territories from experience by policyholders' addresses is a more difficult task.