Statistics for the Sciences

Multiple Comparisons

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Multiple Comparisons
  - Bonferroni test
  - Tukey test
  - Contrasts
  - Scheffé Comparison
  - Dunnett test: Multiple comparisons with a control
- Kruskal-Wallis Test
- Lab

Multiple Comparisons

After conducting one-way ANOVA, if H_0 is rejected, there are several informal methods for determining which means are different:

- Construct boxplots of the different samples to see if one or more of them is very different from the others.
- Construct confidence interval estimates of the means for the different samples, then compare those confidence intervals to see if one or more of them does not overlap with the others (pairwise comparison) or conduct pairwise hypotheses. The method is called LSD (Least Significant Difference).
  - LSD problem: In hypotheses test problems involving a single null hypothesis H_0 the statistical tests are often chosen to control the Type I error rate of incorrectly rejecting H_0 at a pre-specified significance level α. In general, when testing m null hypotheses using independent test statistics, the probability of committing at least one Type I error is $1-(1-\alpha)^m$.

Multiple Comparisons

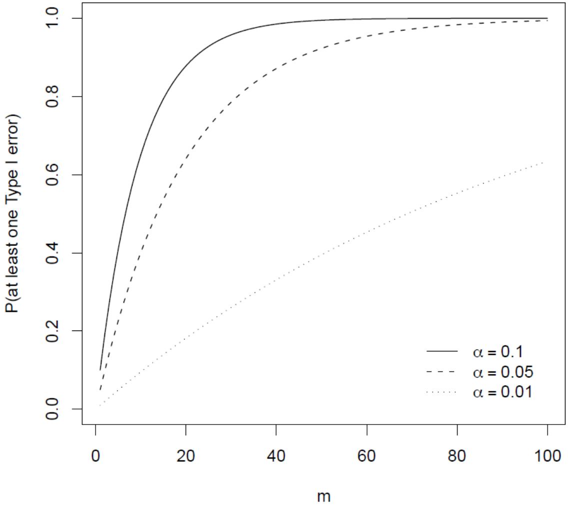


Figure 1: Probability of committing at least one Type I error for diferent significance levels α and number of hypotheses *m*.

Contrasts

- Many multiple comparison methods use the idea of a **contrast**.
- For example, testing

$$H_0 : \mu_3 = \mu_4 \text{ vs. } H_1 : \mu_3 \neq \mu_4$$

  is equivalent to testing

$$H_0 : \mu_3 - \mu_4 = 0 \text{ vs. } H_1 : \mu_3 - \mu_4 \neq 0$$

- testing

$$H_0 : \mu_1 + \mu_2 = \mu_3 + \mu_4 \text{ vs. } H_1 : \mu_1 + \mu_2 \neq \mu_3 + \mu_4$$

  is equivalent to testing

$$H_0 : \mu_1 + \mu_2 - \mu_3 - \mu_4 = 0 \text{ vs. } H_1 : \mu_1 + \mu_2 - \mu_3 - \mu_4 \neq 0$$

Contrasts

- In general, a **contrast** is a linear combination of the parameters of the form

$$\Gamma = \sum_{i=1}^{k} c_i \mu_i,$$

where the contrast constants $c_1, \ldots, c_k$ sum to zero, $\sum_{i=1}^{k} c_i = 0$.

- Both the above hypotheses can be expressed in terms of contrasts:

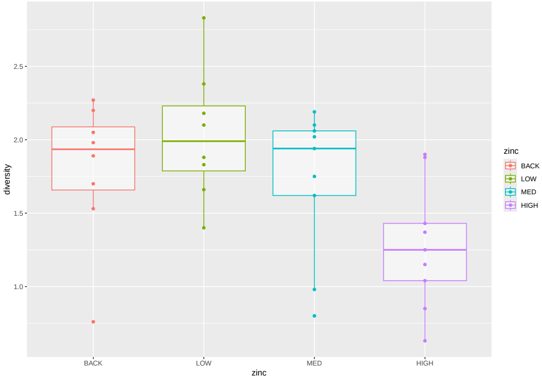
$$H_0 : \sum_{i=1}^{k} c_i \mu_i = 0 \text{ vs. } H_1 : \sum_{i=1}^{k} c_i \mu_i \neq 0$$

- Testing hypotheses involving contrasts can be done using the general t-test since we are testing a linear combination of independent population means.

Contrasts

- For the Medley example, we have

```
##   zinc     mean
## 1 BACK 1.797500
## 2  LOW 2.032500
## 3  MED 1.717778
## 4 HIGH 1.277778
```



Contrasts

- Suppose we test $H_0 : \mu_1 + \mu_3 = \mu_2 + \mu_4$ or $H_0 : \mu_1 - \mu_2 + \mu_3 - \mu_4 = 0$
  - We check the p-value of the test to decide if H_0 can be rejected

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
##
## Fit: aov(formula = Model1)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0   0.2050     0.3203    0.64    0.527
## (Adjusted p values reported -- single-step method)
```

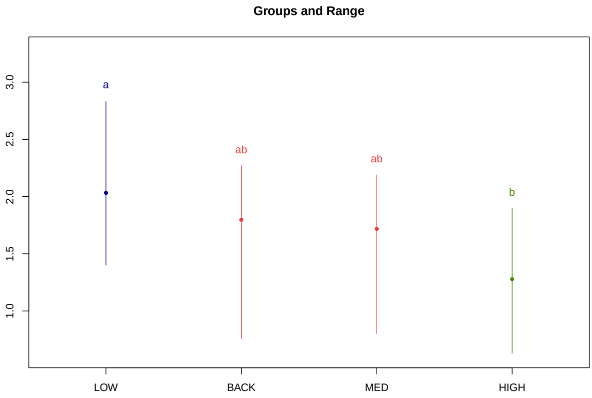
Scheffé Comparison

- Scheffé test was devised in order to be able to test all the possible contrasts a posteriori while maintaining the overall Type I error for the family at a reasonable level, as well as trying to have a relatively powerful test.
- Scheffé Comparison is based on F tests

```
##
##   Posthoc multiple comparisons of means: Scheffe Test
##     95% family-wise confidence level
##
## $zinc
##                  diff     lwr.ci      upr.ci   pval
## LOW-BACK   0.23500000 -0.4549775  0.92497753 0.7973
## MED-BACK  -0.07972222 -0.7502599  0.59081541 0.9886
## HIGH-BACK -0.51972222 -1.1902599  0.15081541 0.1769
## MED-LOW   -0.31472222 -0.9852599  0.35581541 0.5929
## HIGH-LOW  -0.75472222 -1.4252599 -0.08418459 0.0223 *
## HIGH-MED  -0.44000000 -1.0905171  0.21051705 0.2809
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Scheffé Comparison

- Visualization of the grouping



Many-to-one comparisons: Dunnett test

- Dunnett test is the standard method for the the classical many-to-one problem of comparing several groups with a common control group. Suppose there are $m+1$ treatment and let μ_0 be the mean of the control group. Then we are testing $H_{0i}: \mu_0 = \mu_i, i = 1, \ldots, m$ against one of the three alternatives $(>, <, \neq)$.
  - Rejecting any of the null hypotheses thus ensures that at least one of the alternative is supported at a given confidence level $1-\alpha$, if suitable multiple comparison procedures are employed.
- The one-sided Dunnett test takes the minimum (or the maximum, depending on the sideness of the test problem) of the m, say, pairwise t tests

$$t_i = \frac{\bar{y}_i - \bar{y}_0}{s\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}}, i = 1, \ldots, m.$$

Many-to-one comparisons: Dunnett test

- Each test statistic t_i is univariate t distributed. The vector of test statistics $t = (t_1, \ldots, t_m)$ follows an m-variate t distribution with degrees of freedom $\sum_{i=0}^{m} n_i - (m+1)$ (and a correlation matrix).
- In the Medley example, suppose we compare diversity from all other zinc concentrations to that from the BACK level at significance level 0.05.

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: aov(formula = diversity ~ zinc, data = medley)
##
## Linear Hypotheses:
##                  Estimate Std. Error t value Pr(>|t|)
## LOW - BACK == 0   0.23500    0.23303   1.008   0.6195
## MED - BACK == 0  -0.07972    0.22647  -0.352   0.9701
## HIGH - BACK == 0 -0.51972    0.22647  -2.295   0.0728 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Many-to-one comparisons: Dunnett test

- 95% confidence intervals

```
##                   Estimate        lwr        upr
## LOW - BACK   0.23500000 -0.3400227 0.81002269
## MED - BACK  -0.07972222 -0.6385438 0.47909939
## HIGH - BACK -0.51972222 -1.0785438 0.03909939
## attr(,"conf.level")
## [1] 0.95
## attr(,"calpha")
## [1] 2.467581
```

Nonparametric Statistics

- Parametric tests have requirements about the nature or shape of the populations involved.
- Nonparametric tests do not require that samples come from populations with specified distribution assumptions, for example normal distributions or any other particular distributions, and homogeneous variances. Consequently, nonparametric tests are called distribution-free tests.
- Nonparametric methods tend to waste information because exact numerical data are often reduced to a qualitative form.
  - Ranks of data are often used.
- Nonparametric tests are not as efficient (powerful) as parametric tests (if assumptions for parametric tests are satisfied), so with a nonparametric test we generally need stronger evidence (such as a larger sample or greater differences) in order to reject a null hypothesis.

Rank

- Data are sorted when they are arranged according to some criterion, such as smallest to the largest or best to worst.
- A rank is a number assigned to an individual sample item according to its order in the sorted list. The first item is assigned a rank of 1, the second is assigned a rank of 2, and so on.
- Nonparametric methods use the rank information in a data set.

Rank

- **Handling Ties in Ranks**: Find the mean of the ranks involved and assign this mean rank to each of the tied items.

Example The numbers 4, 5, 5, 5, 10, 11, 12, and 12 are given ranks of 1, 3, 3, 3, 5, 6, 7.5, and 7.5, respectively. The table below illustrates the procedure for handling ties.

Sorted Data	Preliminary Ranking		Rank
4	1		**1**
5	2	Mean is 3.	**3**
5	3		**3**
5	4		**3**
10	5		**5**
11	6		**6**
12	7	Mean is 7.5.	**7.5**
12	8		**7.5**

Kruskal-Wallis Test

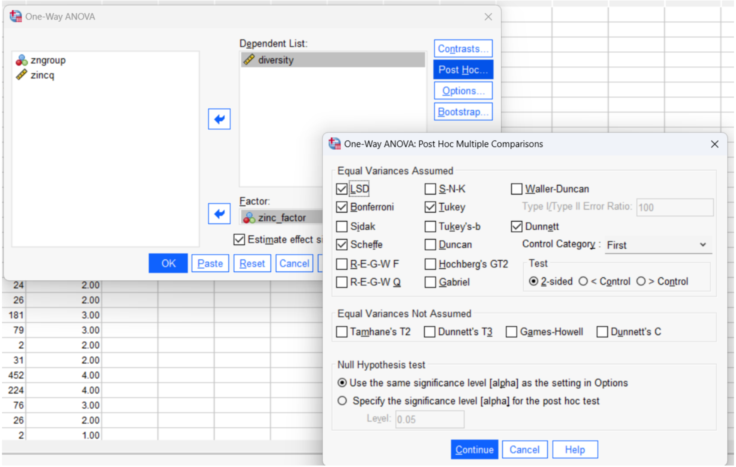
- The Kruskal-Wallis H Test is a **nonparametric** procedure that can be used to compare more than two populations in a completely randomized design.
- All $n = n_1 + n_2 + \ldots + n_k$ measurements are jointly ranked.
- We use the sums of the ranks of the k samples ($k > 2$) to compare the distributions.
- For the Medley example

```
##  BACK   LOW   MED  HIGH
## 1.935 1.990 1.940 1.250
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  diversity by zinc
## Kruskal-Wallis chi-squared = 8.7367, df = 3, p-value = 0.033
```

Lab

- Consider the data `medley.csv` with response `diversity` and fixed-effect factor `zinc`.
- `Analyze` $\rightarrow$ `Compare Means and Proportions` $\rightarrow$ `One-Way ANOVA ....`



Lab

- We read the results of various `Post Hoc Tests`
- For example, the results of Tukey tests

Multiple Comparisons

Dependent Variable: diversity

	(I) zinc_factor	(J) zinc_factor	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	1.00	2.00	-.23500	.23303	.746	-.8686	.3986
		3.00	.07972	.22647	.985	-.5361	.6955
		4.00	.51972	.22647	.122	-.0961	1.1355
	2.00	1.00	.23500	.23303	.746	-.3986	.8686
		3.00	.31472	.22647	.515	-.3011	.9305
		4.00	.75472*	.22647	.012	.1389	1.3705
	3.00	1.00	-.07972	.22647	.985	-.6955	.5361
		2.00	-.31472	.22647	.515	-.9305	.3011
		4.00	.44000	.21970	.210	-.1574	1.0374
	4.00	1.00	-.51972	.22647	.122	-1.1355	.0961
		2.00	-.75472*	.22647	.012	-1.3705	-.1389
		3.00	-.44000	.21970	.210	-1.0374	.1574

Lab

Homogeneous Subsets

diversity

	zinc_factor	N	Subset for alpha = 0.05 1	2
Tukey HSD[a,b]	4.00	9	1.2778	
	3.00	9	1.7178	1.7178
	1.00	8	1.7975	1.7975
	2.00	8		2.0325
	Sig.		.122	.515
Scheffe[a,b]	4.00	9	1.2778	
	3.00	9	1.7178	1.7178
	1.00	8	1.7975	1.7975
	2.00	8		2.0325
	Sig.		.177	.593

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 8.471.

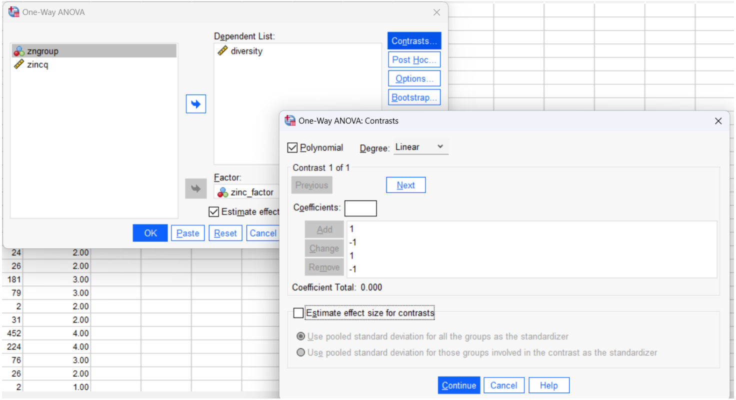
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Lab

- The Homogeneous Subsets table in SPSS output for post hoc tests helps to identify groups of means that are not significantly different from each other.
  - The table lists the means for each group (identified by the zinc_factor values 4.00, 3.00, 1.00, and 2.00).
  - Groups that appear in the same column subset (e.g., 1 or 2) are considered homogeneous, meaning their means are not significantly different from each other.
  - The Sig. row indicates the p-value for the comparisons within each subset.
- For both Tukey HSD and Scheffe tests, the means of groups 4.00, 3.00, and 1.00 are not significantly different from each other (Subset 1).
- Additionally, groups 3.00, 1.00, and 2.00 are also not significantly different from each other (Subset 2).

Lab

- Let's check one contrast $H_0 : \mu_1 - \mu_2 + \mu_3 - \mu_4 = 0$ (equivalent to $H_0 : \mu_1 + \mu_3 = \mu_2 + \mu_4$)



Lab

Contrast Coefficients

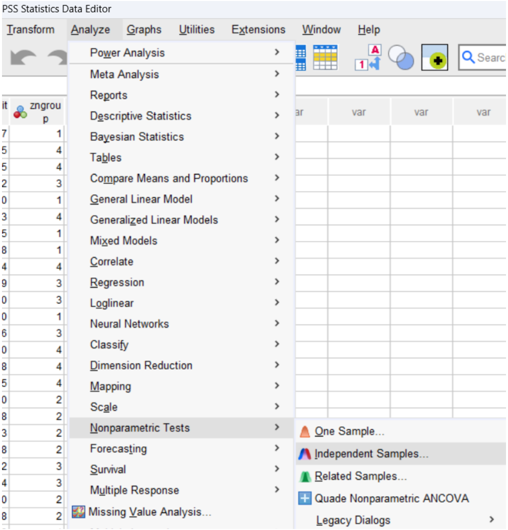
Contrast	zinc_factor 1.00	2.00	3.00	4.00
1	1	-1	1	-1

Contrast Tests

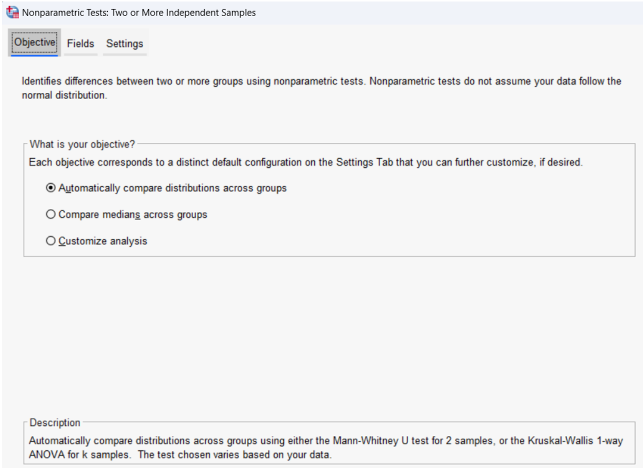
		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)	95% Confidence Interval Lower	Upper
diversity	Assumes equal variances	1	.2050	.32027	.640	30	.527	-.4491	.8591
	Does not assume equal variances	1	.2050	.32023	.640	29.103	.527	-.4498	.8598

Lab

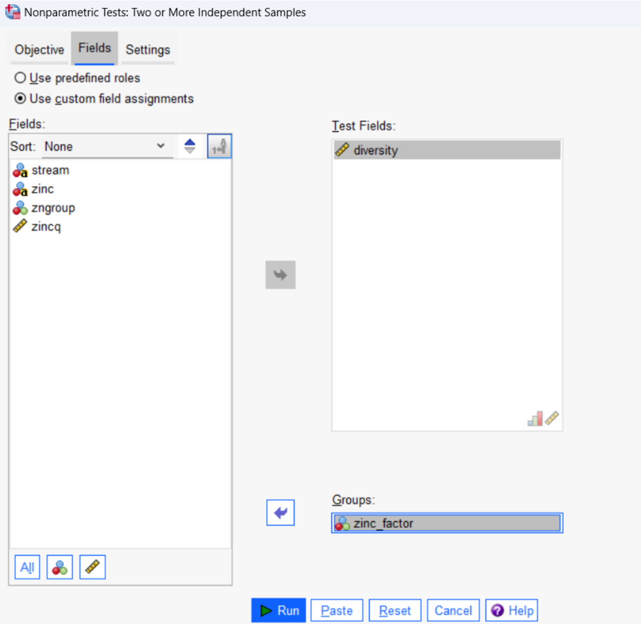
- Kruskal-Wallis Test: `Analyze` → `Nonparametric Tests` → `Independent Samples ....`



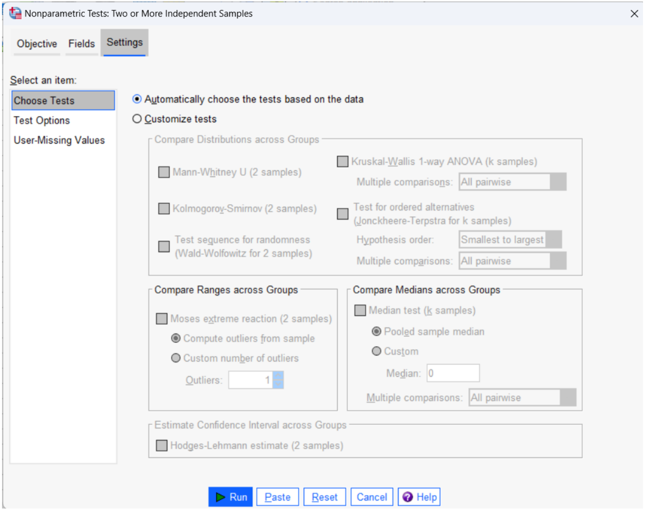
Lab


Nonparametric Tests: Two or More Independent Samples
Objective
Fields
Settings
Identifies differences between two or more groups using nonparametric tests. Nonparametric tests do not assume your data follow the normal distribution.
What is your objective?
Each objective corresponds to a distinct default configuration on the Settings Tab that you can further customize, if desired.
Automatically compare distributions across groups
Compare medians across groups
Customize analysis
Description
Automatically compare distributions across groups using either the Mann-Whitney U test for 2 samples, or the Kruskal-Wallis 1-way ANOVA for k samples. The test chosen varies based on your data.

Lab


Nonparametric Tests: Two or More Independent Samples
Objective
Fields
Settings
Use predefined roles
Use custom field assignments
Fields:
Sort: None
stream
zinc
zngroup
zincq
All
Test Fields:
diversity
Groups:
zinc_factor
Run
Paste
Reset
Cancel
Help

Lab


Nonparametric Tests: Two or More Independent Samples
Objective
Fields
Settings
Select an item:
Choose Tests
Test Options
User-Missing Values
Automatically choose the tests based on the data
Customize tests
Compare Distributions across Groups
Mann-Whitney U (2 samples)
Kolmogorov-Smirnov (2 samples)
Test sequence for randomness (Wald-Wolfowitz for 2 samples)
Kruskal-Wallis 1-way ANOVA (k samples)
Multiple comparisons: All pairwise
Test for ordered alternatives (Jonckheere-Terpstra for k samples)
Hypothesis order: Smallest to largest
Multiple comparisons: All pairwise
Compare Ranges across Groups
Moses extreme reaction (2 samples)
Compute outliers from sample
Custom number of outliers
Outliers: 1
Compare Medians across Groups
Median test (k samples)
Pooled sample median
Custom
Median: 0
Multiple comparisons: All pairwise
Estimate Confidence Interval across Groups
Hodges-Lehmann estimate (2 samples)
Run
Paste
Reset
Cancel
Help

Lab

Independent-Samples Kruskal-Wallis Test

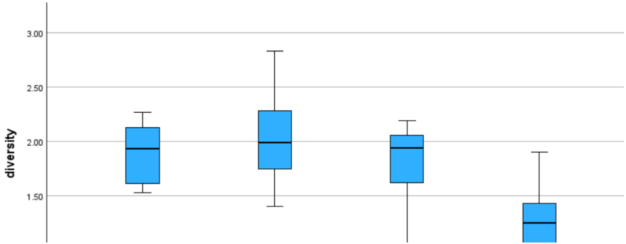
diversity across zinc_factor

Independent-Samples Kruskal-Wallis Test Summary

Total N	34
Test Statistic	8.737[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.033

a. The test statistic is adjusted for ties.

Independent-Samples Kruskal-Wallis Test



License