

Inferences for Contrasts and Treatment Means

Hao Zhang

- Comparison of Means
- Contrast
- Common Types of Contrast
 - Pairwise Comparisons
 - Treatment Versus Control
 - Difference of Averages
 - Trend Contrast
- Inferences for Individual Contrast and Treatment Means
 - Least square estimates
 - Confidence Interval
 - Hypothesis Testing
- Multiple Comparisons
 - Multiple Confidence Intervals
 - Multiple Comparison Methods
- SAS Example

Comparison of Means

ANOVA tells us differences exist - but not where.

↓
* need to test contrasts.

The F-test in ANOVA only tests for the equality of all means. What about the specific differences among the means?

Example: The battery experiment. Four treatments are coded as 1=alkaline, name brand; 2 = alkaline, store brand; 3=heavy duty; name brand; 4 = heavy duty, store brand. Four batteries of each treatment were tested to obtain the lifetime per cent. The F-test has a p-values < 0.0001 that indicates a significant difference among the treatment means. However, we might be interested in the following differences as well:

* pairwise $\mu_2 - \mu_1, \mu_2 - \mu_3, \mu_2 - \mu_4$ or possibly $\frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$.

Each is a contrast, and each has a specific interpretation.

Contrast

Recall:

μ_i : treatment mean; τ_i : treatment effect.

A linear combination $\sum_{i=1}^v c_i \mu_i$ is a contrast if $\sum_i c_i = 0$. The vector (c_1, c_2, \dots, c_v) is the contrast coefficients.

For a contrast,

$$\sum_{i=1}^v c_i \mu_i = \sum_{i=1}^v c_i \tau_i.$$

A contrast can be written in terms of treatment means or treatment effects. It is often expressed in terms of treatment effects particularly for two-way or multi-way models.

Common Types of Contrast

Pairwise Comparisons

every possible $\mu_i - \mu_j$

A pairwise comparison is a difference between any two treatment effects $\tau_i - \tau_j$. The contrast coefficient is $(0, 0, \dots, 1, 0, \dots, -1, 0, \dots)$ where the 1 and -1 are in the i th and j th position, respectively.

Q: If there are $v = 4$ treatments, how many pairwise comparisons that need to be investigated?

Treatment Versus Control

baseline

compare every treatment to baseline

If treatment 1 is control, then the treatment versus control contrasts are $\tau_2 - \tau_1, \dots, \tau_v - \tau_1$.

Difference of Averages

$$\frac{\mu_1 + \mu_4}{2} - \frac{\mu_3 + \mu_4}{2}$$

When the treatments divide naturally into two or more groups and the experimenter is interested in the difference of averages, the difference of averages contrasts are used. For example, in the pedestrian light experiment, it is of interest to use the contrast

$$(\tau_2 + \tau_3 + \tau_4)/3 - \tau_1$$

Trend Contrast when treatments level are ordered (e.g. dose levels)

Used when the levels of the treatment factor are quantitative and have a natural ordering as in the hear-lung experiment. Table A.2 in the textbook lists contrast coefficients for orthogonal polynomial trend contrasts when the levels are equally spaced and the replicates are equal.

$\nu = 3$				$\nu = 4$				
Trend	c_1	c_2	c_3	Trend	c_1	c_2	c_3	c_4
Linear	-1	0	1	Linear	-3	-1	1	3
Quadratic	1	-2	1	Quadratic	1	-1	-1	1
				Cubic	-1	3	-3	1

These coefficients are obtained by fitting a regression model to the uncoded levels of the treatment factor.

Formula for Linear Trend Contrast: Suppose the treatments are ν values x_1, x_2, \dots, x_ν in ascending order. The contrast coefficients are

$$c_i = r_i (x_i - \bar{x}), \text{ where } \bar{x} = (\sum r_i x_i) / n$$

where r_i is the number of observations taken on the i th uncoded level x_i of the treatment factor, and $n = \sum r_i$.

Example The five treatments are 50, 75, 100, 125, 150 revolutions per minutes, but the number of replicates for the treatments are $r_1 = r_3 = r_5 = 5$, $r_2 = 3$ and $r_4 = 2$. Now $n = \sum r_i = 20$, and

$$\bar{x} = (\sum r_i x_i) / n = (5(50) + 3(75) + 5(100) + 2(125) + 5(150)) / 20 = 98.75$$

x_i	$r_i(x_i - \bar{x})$
50	$5 \times (50 - 98.75) = -243.75$
75	$3 \times (75 - 98.75) = -71.25$
100	$5 \times (100 - 98.75) = 6.25$
125	$2 \times (125 - 98.75) = 52.50$
150	$5 \times (150 - 98.75) = 256.25$

The coefficients of contrast are $[-243.75, -71.25, 6.25, 52.50, 256.25]$.

Inferences for Individual Contrast and Treatment Means

Least square estimates

A contrast $\sum c_i \tau_i$ or $\sum c_i \mu_i$ is estimated by $\sum c_i \bar{Y}_i$, which has the normal distribution with mean $\sum c_i \tau_i$ and variance $\sigma^2 \sum c_i^2 / r_i$. The variance σ^2 is estimated by MSE that has $n - \nu$ degrees of freedom. Hence the inferences about the contrast involves the $t_{n-\nu}$ distribution.

Confidence Interval

A $100(1-\alpha)\%$ confidence interval for $\sum c_i \mu_i$ is

$$\left[\underbrace{\sum c_i \bar{Y}_i}_{SE} \pm \underbrace{t_{n-\nu, \alpha/2}}_{\text{standard error}} \underbrace{\sqrt{MSE \sum c_i^2 / r_i}}_{SE} \right] \rightarrow \text{C.I.}$$

Derived from $\sum c_i \bar{Y}_i \sim N(\sum c_i \mu_i, \sigma^2 \sum c_i^2 / r_i)$. It applies to any linear combination $\sum c_i \mu_i$ (e.g., each individual μ_i).

Example: In the battery experiment, find a 95% confidence interval for $\mu_2 - \mu_1$. The point estimate is $\bar{y}_2 - \bar{y}_1 = 289.75$. The

standard error of the estimator is $\sqrt{MSE * \left(\frac{1}{r_2} + \frac{1}{r_1} \right)} = \sqrt{2367.7083 * \left(\frac{1}{4} + \frac{1}{4} \right)} = 34.41$. The number of

degrees of freedom is $n - \nu = 12$, and $t_{12, 0.025} = 2.179$. The minimum significant difference is

$msd = 2.179 * 34.41 = 74.98$. The confidence interval is 289.75 ± 74.98 or $(214.77, 364.73)$.

Hypothesis Testing

Hypothesis Testing

$$H_0 : \sum c_i \tau_i = 0, \quad H_1 : \sum c_i \tau_i \neq 0.$$

reject H_0 if

$$\left| \frac{\sum c_i \bar{Y}_i}{\sqrt{MSE \sum c_i^2 / r_i}} \right| > t_{n-v, \alpha/2}.$$

Multiple Comparisons

Multiple Confidence Intervals

Suppose we want to have pairwise comparisons, say, $\mu_2 - \mu_1, \mu_3 - \mu_2, \mu_4 - \mu_1$. Therefore we would have several confidence intervals. If each one is 95% correct, the probability that these confidence intervals are simultaneously correct is much smaller than 95%. We would like to control the probability that they are simultaneously correct. Several methods have been developed. These are called simultaneous confidence intervals.

Multiple Comparison Methods

1. Bonferroni method for preplanned comparisons: Applied to any m preplanned estimable contrasts. Can be used for any design. Can not be used for data snooping (which means to choose the contrasts that data seem to suggest to be highly significant.)
2. Scheffe method for all comparisons. Applied to any m contrasts. Can be used for any design, and allows for data snooping.
3. Tukey method for all pairwise comparisons: Best for all pairwise comparisons. Can be used for completely randomized designs, randomized block designs, and balanced incomplete block designs.
4. Dunnett method for treatment-versus-control contrasts: Best for all treatment-versus-control contrasts. Can be used for completely randomized designs, randomized block designs, and balanced incomplete block designs.
5. Hsu method for multiple comparisons with the best treatment: Selects the best treatment and identifies those treatments that are significantly worse than the best. Can be used for completely randomized designs, randomized block designs, and balanced incomplete block designs.

Bonferroni method:

Suppose there are m pre-planned contrasts and we want a set of m simultaneous $100(1 - \alpha)\%$ confidence intervals. The Bonferroni method makes each individual interval estimate to be a $100 \left(1 - \frac{\alpha}{2m}\right)\%$ confidence interval. Hence the intervals are given by

$$\text{estimate} \pm t_{n-v, \alpha/(2m)} \times \text{standard error}.$$

A larger critical value is used.

Example: For the Battery Experiment, find the simultaneous 95% confidence intervals for $\mu_2 - \mu_1, \mu_2 - \mu_3, \mu_2 - \mu_4$ using the Bonferroni method. Here $m = 3$, and the critical value is $t_{12, 0.05/(2*3)} = 2.779$. This critical coefficient is found in SAS by the function `tinv`:

```
data;
prob=1-0.05/6;
df=12;
tcv=tinv(prob, df);
proc print;
run;
```

The standard error of the difference between two sample means is $\sqrt{MSE * \frac{2}{r}} = 48.67/\sqrt{2} = 34.41$ (why?). The margin errors are $2.779 * 34.41 = 95.63$.

The estimates of $\mu_2 - \mu_1, \mu_2 - \mu_3, \mu_2 - \mu_4$ are $\bar{y}_2 - \bar{y}_1 = 289.75, \bar{y}_2 - \bar{y}_3 = 427.50, \bar{y}_2 - \bar{y}_4 = 364.25$. Hence the simultaneous 95% confidence intervals are $289.75 \pm 95.63, 427.50 \pm 95.63$ and 364.25 ± 95.63 .

Note these confidence intervals have the same margin of error because of equal sample sizes.

Two drawbacks:

1. It is dangerous to apply the Bonferroni method after the data are collected, because one might choose some comparisons that appear to

be significant. This is called data snooping.

2. The confidence intervals become increasingly wider as m increases.

Scheffé Method

Applies to all contrasts. The simultaneous confidence intervals for $\sum_i c_i \mu_i$ are given by

$$\sum_i c_i \bar{y}_{i.} \pm \sqrt{(v-1)F_{v-1, n-v, \alpha}} \sqrt{MSE \sum_i c_i^2 / r_i}$$

It is based on the following result due to Henry Scheffe (1953): Let

$$K = \max_{\sum c_i = 0} \frac{\left(\sum c_i \bar{Y}_{i.} - \sum c_i \mu_i \right)^2}{MSE \sum c_i^2 / r_i}.$$

Then

$$\frac{K}{v-1} \sim F_{v-1, n-v}$$

where $(v-1)$ is the dimension of the contrasts or the number of independent contrasts, and $(n-v)$ is the degrees of freedom of MSE. Therefore we have

$$P(K \leq (v-1)F_{v-1, n-v, \alpha}) = 1 - \alpha, \text{ or } P\left(\sqrt{K} \leq \sqrt{(v-1)F_{v-1, n-v, \alpha}}\right) = 1 - \alpha.$$

Equivalently,

$$P\left(|\sum c_i \bar{y}_{i.} - \sum c_i \mu_i| \leq \sqrt{(v-1)F_{v-1, n-v, \alpha}} \sqrt{MSE * \sum \left(\frac{c_i^2}{r_i}\right)} \text{ for all } \sum c_i = 0\right) = 1 - \alpha$$

Scheffé's method usually produces wide intervals. Use it only when the number of m is very large.

Tukey's method for all pairwise comparisons

The best method for all pairwise comparisons. It is based on

$$Q = \frac{\max_{i,j} \left| (\bar{Y}_{i.} - \bar{Y}_{j.}) - (\mu_i - \mu_j) \right|}{\sqrt{MSE/r}}$$

where r is the common sample size. For each pair, the ratio has a t -distribution. This maximum has the Studentized Range Distribution, whose upper critical value is denoted by $q_{v, n-v, \alpha}$ and is provided in Table A.8.

The confidence interval is given by

$$\text{estimate} \pm \frac{1}{\sqrt{2}} q_{v, n-v, \alpha} * (\text{standard error}).$$

Example: In the Battery Experiment, $v = 4, n - v = 12, \frac{q_{v, n-v, 0.05}}{\sqrt{2}} = \frac{4.20}{\sqrt{2}} = 2.970$. The critical coefficient for the Tukey SCIs for all pairwise comparison is **2.97**.

Using the Bonferroni method, since there are $m = 6$ all pairwise comparisons, the critical coefficient is $t_{12, 0.05/(2m)} = 3.152$.

For the Scheffe's method, the critical coefficient is

$$\sqrt{(v-1)F_{v-1, n-v, 0.05}} = \sqrt{(3-1) * 3.49} = 3.24.$$

Tukey’s method is the best for ALL pairwise comparisons. If not all but only some pairwise comparisons are needed, Tukey’s method may not be the best one.

Dunnett Method for Treatment-versus-Control Comparisons

Let treatment 1 be the control, then the $(v - 1)$ treatment-versus-control contrasts are $\mu_i - \mu_1, i = 1, \dots, v$. Dunnett (1955) developed a method for MC that is based on the joint distribution of the estimators $\bar{Y}_{i\cdot} - \bar{Y}_{1\cdot}, i = 2, \dots, v$, which is a multivariate t -distribution and depends on the correlation between the differences $\bar{Y}_{i\cdot} - \bar{Y}_{1\cdot}$. When all sample sizes equal, the critical coefficient w_D is given in Table A.10 for two-sided SCIs and Table A.9 for one sided SCIs. The one-sided SCIs are used to select those treatments that have higher means than the control.

SAS Example

In the battery experiment, there are two treatment factors each having two levels. These are battery “duty” (level 1 = alkaline, level 2 = heavy duty) and “brand” (level 1 = name brand, level 2 = store brand). This gives four treatment combinations coded 11, 12, 21, 22. We will recode these treatment combinations as 1, 2, 3, 4. Thus, the levels of battery type are:

Level	Treatment Combination
1	alkaline, name brand (11)
2	alkaline, store brand (12)
3	heavy duty, name brand (21)
4	heavy duty, store brand (22)

```
* battery.sas, battery experiment, Table 4.2 (page 94);
options ls=85;
DATA BATTERY;
  INPUT TYPEBAT LIFEUC ORDER;
  LINES;
  1 611 1
  1 537 3
  4 476 4
  1 542 5
  1 593 6
  2 794 7
  3 445 8
  4 569 9
  2 827 10
  2 898 11
  3 490 12
  4 480 13
  3 384 14
  4 460 15
  3 413 16
;
run;
PROC GLM DATA=BATTERY;
  CLASSES TYPEBAT;
  MODEL LIFEUC = TYPEBAT;
  ESTIMATE 'DUTY'
    TYPEBAT 1 1 -1 -1 /DIVISOR = 2;
  ESTIMATE 'BRAND'
    TYPEBAT 1 -1 1 -1 /DIVISOR = 2;
  lsmeans TYPEBAT/cl pdiff adjust=TUKEY alpha=0.05;
  lsmeans TYPEBAT/cl pdiff=control("2") adjust=Dunnett;
  lsmeans TYPEBAT/pdiff=controlL("2") adjust=Dunnett;
run;
```

5. Multiple Comparisons

When doing **many contrasts**, adjust for Type I error.

Methods:

Method	Best for	Notes
Bonferroni	≤10 pre-planned contrasts	Widens CI
Scheffé	All possible contrasts	Very conservative
Tukey	All pairwise	Best for pairwise
Dunnnett	Treatment vs control	Focused
Hsu	Multiple vs best	Used to find best performer