

Chapter 4 Inferences for Contrasts and Treatment Means

4.1. Introduction

The objective of an experiment is often more specific than merely determining whether all treatments have the same effects on the response. For example, a medical experiment might be concerned with the efficacy of each of several new drugs compared with a standard drug. In this chapter, we will study inferences about treatment comparisons and treatment means.

Example: The battery experiment. Four treatments are coded as

1=alkaline, name brand; 2=alkaline, store brand;

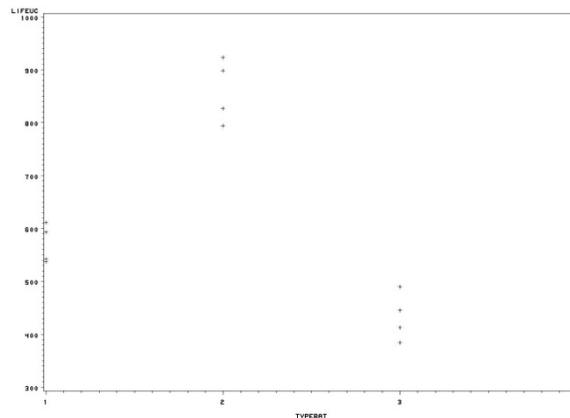
3=heavy duty; name brand; 4=heavy duty, store brand.

Four batteries of each treatment were tested to obtain the lifetime per cent.

The one-way ANOVA shows that there is sufficient evidence that the treatment means are different ($p\text{-value} < 0.0001$). What exactly are the differences? The plot shows that treatment 2 has the longest lifetime per cent. So one might be interested in the following:

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

Each of these is called a contrast. Meaning of each of the above contrasts?



4.2. Contrasts

A contrast is a linear combination of the form $\sum c_i \tau_i$ with $\sum c_i = 0$. The c_i are called contrast coefficients. For example

$$\tau_2 - \tau_1, \tau_3 - \tau_1, \tau_3 - (\tau_1 + \tau_2)/2 \text{ are all contrasts.}$$

Note that since $\sum c_i = 0$, $\sum c_i \tau_i = \sum c_i \mu_i$. Therefore a contrast of treatment effects is also a contrast of treatment means. When $\sum c_i \neq 0$, $\sum c_i \tau_i$ does not make sense and is said to be not estimable. However, $\sum c_i \mu_i$ is always estimable although it often is not of practical interest to estimate $\sum c_i \mu_i$ when $\sum c_i \neq 0$.

4.2.1. Pairwise Comparisons

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 $v=3$ treatments, how many pairwise comparisons that need to be investigated?

4.2.2. Treatment Versus Control

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

4.2.3. Difference of Averages

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

4.2.4. Trend Contrast

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

4.3. Inference for Individual Contrast and Treatment Means

In this section, we will focus of confidence intervals for individual contrast and treatment mean. Note that the contrast $\sum c_i \tau_i = \sum c_i \mu_i$ is estimable and is estimated by $\sum c_i \hat{\mu}_i = \sum c_i \bar{y}_i$. This estimator has a $N(\sum c_i \mu_i, \sigma^2 \sum c_i^2 / r_i)$ distribution under model assumptions.

Therefore, the variance of the estimator $\sum c_i \bar{y}_i$ is estimated by

$$MSE \sum_i \frac{c_i^2}{r_i}$$

which is independent of $\sum_i c_i \bar{y}_i$. (why?). Can be shown that

$$\frac{\sum_i c_i \bar{y}_i - \sum_i c_i \mu_i}{\sqrt{MSE \sum_i (c_i^2 / r_i)}}$$

has a t-distribution with $(n - v)$ degrees of freedom.

Confidence interval for $\sum c_i \mu_i$: *Estimate* $\pm t_{df, \alpha/2} \times$ (*standard error of the estimator*), where $t_{df, \alpha/2}$ is the critical value of t-distribution and the degree of freedom for one-way anova is n-v.

In particular, the confidence interval for the treatment mean μ_i is

$$\bar{y}_i \pm t_{df, \frac{\alpha}{2}} \sqrt{\frac{MSE}{r_i}}$$

Example:

1. 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

$$\sqrt{MSE * \left(\frac{1}{r_2} + \frac{1}{r_1}\right)} = \sqrt{2367.7083 * \left(\frac{1}{4} + \frac{1}{4}\right)} = 34.41. \text{ The number of degrees of freedom is}$$

$n - v = 12$, 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).

2. In the same experiment, find a 95% confidence interval for μ_2 .

We know $\bar{y}_2 = 860.50$, $\sqrt{MSE} = 48.67$, $r_2 = 4$, $t_{12, 0.025} = 2.179$. So the confidence interval is given by

$$860.50 \pm 2.179 \times \frac{48.67}{2} = 860.50 \pm 53.03.$$

Note that the estimate of a contrast and its standard error can be obtained in SAS by the estimate statement.

Confidence limits for a sample mean can be obtained by the lsmeans statement with the cl option in SAS.

Hypothesis testing of a contrast: one-sided or two-sided test.

4.4. Multiple Comparisons

4.4.1. Multiple Confidence Intervals

Suppose we want to have pairwise comparisons. Therefore we would have several confidence intervals, each for a pairwise comparison. 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.

4.4.2. 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. **Dunnnett 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

1. 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 `tinvt`:

```
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 $\sqrt{MSE * \frac{2}{r}} = 48.67/\sqrt{2} = 34.41$. 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 .

These intervals imply that μ_2 is significantly higher than other treatment means.

2. Scheffe Method is based on the following result due to Henry Scheffe (1953): Let

$$K = \max_{\sum c_i = 0} \frac{(\sum c_i \bar{y}_i - \sum c_i \mu_i)^2}{MSE \sum c_i^2 / r_i}.$$

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

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

Equivalently,

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

For any contrasts, the Scheffe SCI are constructed by

$\text{estimate} \pm \sqrt{(v-1)F_{v-1, n-v, \alpha}} \times (\text{standard error}).$

The Scheffe method can be applied to any contrasts.

3. Tukey's method for all pairwise comparisons is based on the distribution of

$$Q = \frac{\max_{i,j} |(\bar{Y}_i - \bar{Y}_j) - (\mu_i - \mu_j)|}{\sqrt{MSE/r}}$$

where r is the common sample size. For each pair, the ratio has a t -distribution. The 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

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

For 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

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

Therefore, Tukey's method yields the shortest SCIs. In general, 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.

When the sample sizes are unequal, Hayter (1984) showed that Tukey's method yields an overall confidence level to be at least $100(1 - \alpha)\%$. So it is a conservative method.

4. 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 $\bar{Y}_i - \bar{Y}_1, i = 2, \dots, v$, which is a multivariate t -distribution and depends on the correlation between the differences $\bar{Y}_i - \bar{Y}_1$. 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.

4.5 SAS programs

Here is the Example for Tukey's method for all pairwise comparison and Dunnett's method for treatment-versus-control comparisons.

```
* 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
;
PROC PLOT;
  PLOT LIFEUC*TYPEBAT
      / VPOS=19 HPOS=50;
;
PROC GLM;
  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/cl pdiff=controlL("2") adjust=Dunnett;
run;
```