Introduction to the Bootstrap and Robust Statistics

PSY711/712

Winter Term 2009

1 Comparing Multiple Independent Groups

Most of the methods described in the previous sections of the course can be generalized to the case where there are more than two independent groups. Details are presented in Chapter 7 of Wilcox [7].

A common measure of the difference among several means is the so-called F ratio:

$$F = \frac{\mathrm{MS_{BG}}}{\mathrm{MS_{WG}}} = \frac{n\sum_{j=1}^{J}(\bar{Y}_j - \bar{Y}_{..})^2}{\frac{1}{N-J}\sum_{j=1}^{J}\left(\sum_{i=1}^{n}(Y_{ij}-\bar{Y}_j)^2\right)} \tag{1}$$

where $\mathrm{MS_{BG}}$ and $\mathrm{MS_{WG}}$ are, respectively, Mean Square Between-Groups and Mean Square Within-Groups, J is the number of groups, n is the number of observations per group, N is the total number of observations, Y_{ij} is the ith score in the jth group, $\bar{Y}_j$ is the mean of group j, and $Y_{..}$ is the mean of the group means. Notice that the numerator of F is zero when $\bar{Y}_1 = \bar{Y}_2 = \cdots = \bar{Y}_J$ and increases as the differences among group means increases. The denominator is a measure of variability within groups that, like the denominator in the formula defining t, scales the between-group differences in terms of a measure of error.

When the scores in each group are distributed normally with the same variance, F follows the theoretical F distribution with $(J-1)$ and $N-J$ degrees of freedom. When the scores are not distributed normally, or when they are normally distributed with different variances, the F statistic does not follow the theoretical F distribution. Deviations from normality and/or constant variance can alter the Type I and Type II error rates significantly, especially when the groups have different n[6, 7].

1.1 Welch Test on Group Means

An independent-samples t test assumes that both samples are drawn from normal distributions with equal variance. In the case where the two variances differ, the t statistic does not follow the t distribution with $n_1 + n_2 - 2$ degrees of freedom. However, in previous sections we saw that the statistic *does* follow (at least approximately) a t distribution with reduced degrees of freedom calculated with the Welch-Satterthwaite formula. Welch [5] generalized this method to the case of several means. The following code constructs several sets of data, which are concatenated into single variable `Y`, a grouping variable (i.e., a factor), `group`, and then performs the Welch test with the `oneway.test` function. Note that sample size can differ across groups:

```
> set.seed(88)
> y1 <- rnorm(20, 0, 1)
> y2 <- rnorm(30, 3, 4)
> y3 <- rnorm(18, 0, 3)
> Y <- c(y1, y2, y3)
> group <- as.factor(c(rep("a", 20), rep("b", 30), rep("c", 18)))
> theData <- data.frame(Y, group)
> oneway.test(Y ~ group, data = theData)
```

```
        One-way analysis of means (not assuming equal variances)

data:  Y and group
F = 5.8873, num df = 2.000, denom df = 35.139, p-value = 0.006235
```

The null hypothesis being evaluated by `oneway.test` is that all group *means* are equal. Again, the analysis assumes that the scores are drawn from normal distributions that may differ in variance. If this assumption is invalid – for example, if the data are drawn from skewed distributions, or have heavy tails – then the p-value may be misleading.

1.2 Generalization of Yuen's Test

The following command shows how to test the null hypothesis that all of the group trimmed means are equal. The test calculates F_t, which is like the F statistic used in classical ANOVA except that F_t is based on differences among group trimmed means, rather than group means. See the box on the following page for details.

In the following code, the measures from different groups are stored in a *list*, Y. Also, notice that the different parts of the list are accessed using double brackets (e.g., Y[[1]]):

```
> set.seed(8)
> y1 <- rnorm(20, 0, 1)
> y2 <- rnorm(30, 3, 4)
> y3 <- rnorm(18, 0, 3)
> Y <- list()
> Y[[1]] <- y1
> Y[[2]] <- y2
> Y[[3]] <- y3
> t1way(Y, tr = 0.2)

$TEST
[1] 8.921138

$nu1
[1] 2

$nu2
[1] 19.50166

$siglevel
[1] 0.001773446
```

The value of F_t is returned as `TEST`, the degrees of freedom in the numerator and denominator are `nu1` and `nu2`, and the p-value for the null hypothesis test of no difference among trimmed means is `siglevel`. The p-value assumes that F_t follows an F distribution with `nu1` and `nu2` degrees of freedom.

The following equations, which are used by `t1way`, come from Wilcox [7] (page 267).
We want to test the null hypothesis

$$H_0 : \mu_{t1} = \cdots = \mu_{tJ}$$

First, we let

$$d_j = \frac{(n_j - 1)s_{wj}^2}{h_j \times (h_j - 1)}$$

where s_{wj}^2 is the sample Winsorized variance for group j, and h_j is the effective sample size of the jth group. Then we compute the following quantities:

$$w_j = \frac{1}{d_j} \quad (2)$$

$$U = \sum_{j=1}^{J} w_j \quad (3)$$

$$\tilde{Y} = \frac{1}{U} \sum^{J} j = 1 w_j \bar{Y}_{tj} \quad (4)$$

$$A = \frac{1}{(J-1)} \sum_{j=1}^{J} w_j (\bar{Y}_{tj} - \tilde{Y}_{t..})^2 \quad (5)$$

$$B = \frac{2(J-2)}{J^2 - 1} \sum_{j=1}^{J} \frac{(1 - \frac{w_j}{U})^2}{h_j - 1} \quad (6)$$

$$F_t = \frac{A}{1 + B} \quad (7)$$

When the null hypothesis is true, F_t is distributed approximately as F with ν_1 and ν_2 degrees of freedom:

$$\nu_1 = J - 1 \quad (8)$$

$$\nu_2 = \left[\frac{3}{J^2 - 1} \sum \frac{(1 - w_j/U)^2}{h_j - 1} \right]^{-1} \quad (9)$$

As in the two-sample case, it is possible to abandon the assumption that F_t follows the F distribution, and instead use bootstrap methods to evaluate the null hypothesis. The function `t1waybt` uses the a bootstrap method that is similar to the percentile-t method:

```
> t1waybt(Y, tr = 0.2, nboot = 1999)

[1] "Taking bootstrap samples. Please wait."
[1] "Working on group  1"
[1] "Working on group  2"
[1] "Working on group  3"
$test
[1] 8.921138

$p.value
[1] 0.007003502
```

It is important to understand how `t1waybt` works. First, the sample trimmed means are calculated for each group. Next, the trimmed mean of the jth group is subtracted from each score in group j, which produces J groups of *centered* scores that all have trimmed means of zero. It is important to note that the subtraction alters only the trimmed means: the variances, skewness, etc. of the sample distributions are unaffected. Next, we construct bootstrapped samples by sampling each group randomly with replacement and calculate F_t on the bootstrapped sample. The resampling and F_t calculations are repeated a large number of times, yielding an bootstrap estimate of the sampling distribution of F_t when — and this is the important part — *the null hypothesis of no difference among trimmed means is true.* Why is the null hypothesis true? Because we centered the data in all of the groups, and therefore the group means are all zero. `t1waybt` returns the value of F_t calculated on the original sample, as well as the bootstrapped estimate of the probability of obtaining a value of F_t that is equal to or greater than the observed value. The null hypothesis is rejected when $p < \alpha$.

An alternative to `t1waybt` is the function `b1way`:

```
> b1way(Y, nboot = 1999, est = tmean, tr = 0.2)

[1] "Taking bootstrap samples. Please wait."
[1] "Working on group  1"
[1] "Working on group  2"
[1] "Working on group  3"
$teststat
[1] 3.003067

$p.value
[1] 0.0005002501
```

The test statistic (`teststat`) is

$$H = (1/N) \sum_{j=1}^{J} n_j (\hat{\theta}_j - \bar{\theta})^2 \tag{10}$$

where J is the number of groups, N is the total number of observations, $\hat{\theta}_j$ is the statistic (in this case, the 20% trimmed mean) calculated for group j, and

$$\bar{\theta} = \frac{1}{J} \sum_{j=1}^{J} \hat{\theta}_j$$

The basic procedure used by `b1way` is the same as the one used by `t1waybt`: The data from each group are centered so that the test statistic (i.e., the trimmed mean) is zero in all groups, and the bootstrapped samples are drawn from the centered data to generate an estimate of the sampling distribution of H. The p-value is an estimate of the probability of obtaining a value of H equal to or more extreme than the observed value of H when the null hypothesis is true.

Wilcox recommends using the using `b1way` over `t1waybt`.

1.3 One-way Tests of M-estimators

The function `b1way` is used to test the null hypothesis of no differences among group M-estimators. In this case, however, Eq 10 refers to variation among M-estimators rather than trimmed means.

```
> b1way(Y, nboot = 1999, est = mom)

[1] "Taking bootstrap samples. Please wait."
[1] "Working on group  1"
[1] "Working on group  2"
```

```
[1] "Working on group  3"
$teststat
[1] 2.905077

$p.value
[1] 0.002501251
```

References

[1] Carl Gaspar, Allison B Sekuler, and Patrick J Bennett. Spatial frequency tuning of upright and inverted face identification. *Vision Res*, 48(28):2817–2826, 2008.

[2] Roger E Kirk. *Experimental design: procedures for the behavioral sciences*. Brooks/Cole, Pacific Grove, Calif., 3rd ed edition, 1995.

[3] Scott E Maxwell and Harold D Delaney. *Designing experiments and analyzing data: a model comparison perspective*. Lawrence Erlbaum Associates, Mahwah, N.J., 2nd ed edition, 2004.

[4] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org), 2007.

[5] B. L. Welch. On the comparison of several mean values: an alternative approach. *Biometrika*, 38:330–336, 1951.

[6] R. R. Wilcox. Understanding the practical advantages of modern anova methods. *J Clin Child Adolesc Psychol*, 31(3):399–412, 2002.

[7] Rand R. Wilcox. *Introduction to robust estimation and hypothesis testing*. Elsevier Academic Press, 2005.