22S:101 Biostatistics

Comparing more than two population means

Lecture 18
March 10, 2004

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

Comparing more than two population means

Example: Does the presence of pets or friends affect responses to stress?

- Allen, Blascovich, Tomaka, and Kelsey, 1988, *Journal of Personality and Social Psychology*
- subjects: 45 women who described themselves as dog lovers
- randomly assigned to three groups: to do a stressful task
  1. alone
  2. with a good friend present
  3. with their dog present
- Subjects' mean heart rate during the task was one measure of the effect of stress.

Goal: to compare population means under three different "treatments"

- a *three*-independent-sample problem
- Call the population mean heart rates μ_1 for when pets are present, μ_2 for when friends are present, and μ_3 for when women perform task alone: then
  - $H_0 : \mu_1 = \mu_2 = \mu_3$
  - $H_a : \mu_1 \neq \mu_2$ or $\mu_1 \neq \mu_3$ or $\mu_2 \neq \mu_3$
    - not one-sided or 2-sided

SAS descriptive statistics:

```
         Analysis Variable : BEATS

--------------- GROUP=C ----------------------------------

 N          Mean       Std Dev       Minimum       Maximum
----------------------------------------------------------
15    82.5240667     9.2415747    62.6460000    99.0460000
----------------------------------------------------------

--------------- GROUP=F ----------------------------------

 N          Mean       Std Dev       Minimum       Maximum
----------------------------------------------------------
15    91.3251333     8.3411341    76.9080000   102.1540000
----------------------------------------------------------

--------------- GROUP=P ----------------------------------

N          Mean       Std Dev       Minimum       Maximum

----------------------------------------------------------
15    73.4830667     9.9698202    58.6920000    97.5380000
----------------------------------------------------------
```

To infer about the three population means, we *might* use the two-independent-sample t test 3 times:

- Test $H_0 : \mu_1 = \mu_2$ to see if mean heart rate when pet is present differs from mean when friend is present.
- Test $H_0 : \mu_1 = \mu_3$ to see if mean heart rate when pet is present differs from mean when alone.
- Test $H_0 : \mu_2 = \mu_3$ to see if mean heart rate when friend is present differs from mean when alone.

Problem with this approach:

- 3 p-values for 3 different tests don't tell us how likely it is that *three* sample means are spread apart as far as these are.
- might be that $\bar{x}_1 = 73.48$ and $\bar{x}_2 = 91.32$ are significantly different if we look at just 2 groups but *not* significantly different if we know they are the smallest and largest means in 3 groups
  - As more and more groups are considered, we expect gap between smallest and largest sample mean to get larger.
  - (Imagine comparing heights of shortest and tallest person in larger and larger groups of people.)
- the probability of Type I error for the whole set of t-tests will be much bigger than the α level set for each one

Multiple comparisons procedures in statistics

- issue: how to do many comparisons at once with some overall measure of confidence in all our conclusions
- two steps
  - overall test of whether there is good evidence of *any* differences among parameters we wish to compare
  - follow-up analysis to decide which of parameters differ and to estimate size of differences

Step one: One-Way Analysis of Variance (ANOVA)

- step one (overall test) for *some* difference among 3 or more population means
- uses an *F test* to compute a p-value

Dogs, friends, and stress example:

```
                   Analysis of Variance Procedure

                      Class    Levels    Values

                      GROUP         3    C F P

               Number of observations in data set = 45

                   Analysis of Variance Procedure

Dependent Variable: BEATS
                                 Sum of            Mean
Source             DF           Squares          Square  F Value    Pr > F

Model               2      2387.6889920    1193.8444960    14.08    0.0001
Error              42      3561.2994916      84.7928450
Corrected Total    44      5948.9884836

           R-Square              C.V.         Root MSE          BEATS Mean
           0.401360          11.16915        9.2083030           82.444089


Source             DF          Anova SS     Mean Square  F Value    Pr > F

GROUP               2      2387.6889920    1193.8444960    14.08    0.0001
```


Main idea of ANOVA

What matters is how far apart sample means are *relative to variability of individual observations.*

- F statistic

$$F = \frac{\textit{variation among the sample means}}{\textit{variation among individuals in the same sample}}$$

- compare to a cutoff value in an **F distribution**

Notation:

- I = number of different populations whose means we are studying
- n_i = number of observations in sample from ith population
- N = total number of observations in all samples combined

F distributions

- many different F distributions, identified by two parameters
  - numerator degrees of freedom = I - 1
  - denominator degrees of freedom = N - I

Example

Do four varieties of tomato plant differ in mean yield? Agronomists grew 10 plants of each variety and recorded the yield of each plant in pounds of tomatoes.

What are

- the populations of interest
- the variable of interest
- I
- each n_i
- the degrees of freedom for the ANOVA F statistic

Assumptions for One-Way ANOVA

- We have I independent simple random samples, one from each of I populations.
- Each population i has a normal distribution with unknown mean μ_i.
  - As with t-tests, if sample sizes are large enough in each sample, Central Limit Theorem says inference based on sample means is OK even if population distributions are not exactly normal.

- All of the populations have the same standard deviation σ (unknown)
  - unlike t-tests, there is no general procedure when population standard deviations are not assumed to be equal
  - rough rule of thumb: if largest sample standard deviation is no more than twice the smallest sample standard deviation, then population standard deviations probably are close enough to equal that ANOVA procedure is OK

Step two: individual t-tests with correction for multiple comparisons

This is the *follow-up* test.

- should be carried out *only* if the F test from one-way ANOVA is significant at the chosen significance leve.

Goal: to set the *overall* probability of commiting a type I error at α when doing pairwise comparisons of k different means

- we will perform $\binom{k}{2}$ two-independent-sample t-tests
- we will conduct each one at the significance level
$$\alpha^* = \frac{\alpha}{\binom{k}{2}}$$
- This is called the *Bonferroni correction*

  - very conservative

Dogs, friends, and stress example

- There are $k = 3$ samples, so there are $\binom{k}{2} =$ 3 different pairs to compare.
- To get an overall significance level $\alpha = .05$ on all 3 tests considered together, we conduct each one at

$$\alpha^* = \frac{.05}{3} = .0167$$

  - That is, we would consider the difference between two population means to be significantly different from zero at the .05 level only if the p-value for the the t-test for that pair was less than .0167.

  - Equivalently, we could multiply the p-value from each t-test by 3.
    * If the result was less than .05, we would consider the difference between two population means to be significantly different from zero at the .05 level

SAS does the adjusting and prints a grouped list of the classes. Means with the same letter are not significantly different at the specified alpha level.

```
            Analysis of Variance Procedure

      Bonferroni (Dunn) T tests for variable: BEATS

NOTE: This test controls the type I experimentwise error rat
      generally has a higher type II error rate than REGWQ.

           Alpha= 0.05  df= 42  MSE= 84.79285
                Critical Value of T= 2.49
         Minimum Significant Difference= 8.3847

Means with the same letter are not significantly different.

      Bon Grouping             Mean      N  GROUP

                 A           91.325     15  F

                 B           82.524     15  C

                 C           73.483     15  P
```


One-way ANOVA in SAS

```
options linesize = 79 ;

data pet ;
infile '/temp/pet.dat' ;
input group $ beats ;
run ;

proc sort data = pet ;
by group ;
run ;

proc means data = pet ;
by group ;
var beats ;
run ;

proc anova data = pet ;
class group ;
model beats = group ;
run ;

proc anova data = pet ;
class group ;
model beats = group  ;
means group / bon alpha = .05 ;
run ;
```