Lecture 32

Nancy Pfenning Stats 1000

Chapter 16: Analysis of Variance

Example

Suppose your instructor administers 3 different forms of a final exam. When scores are posted, you see the observed mean scores for those 3 different forms—82, 66, and 60—are not the same. Is this due to chance variation, or do the 3 exams not share the same level of difficulty?

Version 1: 65, 73, 78, 79, 86, 93, 100

Version 2: 39, 58, 63, 67, 69, 74, 92

Version 3: 39, 52, 62, 64, 66, 77

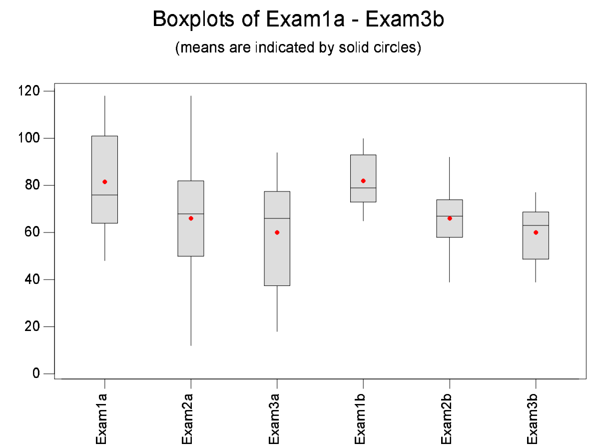
[Note: if there were only 2 different exams, a two-sample t test could be used.]

It would be unrealistic to expect 3 identical sample mean scores, even if the exams were all equally difficult: in other words, there's bound to be some variation *among* means in the 3 groups.

Also, of course there will be variation of scores *within* each group.

If the ratio of *variation among groups* to *variation within groups* is large enough, we will have evidence that the population mean scores for the 3 groups actually differ. Picture two possible configurations for the data, both of which represent 3 data sets with means 82, 66, and 60. Thus, variation *among* means (from the overall mean of 70) would be the same for both configurations.

1. There could be large *within*-group variation, such as we see in Exams 1a, 2a, and 3a, in which case the ratio among to within is small; in this case, the data could be coming from populations that share the same mean.
2. There could be small *within*-group variation, such as we see in Exams 1b, 2b, and 3b, in which case the ratio among to within is large; in this case, the population means probably differ.



Note that sample size, although not yet mentioned, plays a role, too. If there were 100 students in each group, we would set more store by the differences than if there were only 2 students in each group.

We use *One-Way Analysis of Variance* to compare several population means based on independent SRS's from each population. "One-way" refers to the fact that only one explanatory variable, or factor, is considered. Populations are assumed to be normal [robustness discussed on page 564] with equal standard deviations $\sigma_1 = \sigma_2 = \cdots$ [Rule of Thumb: check that largest sample standard deviation s_i is no more than twice the smallest], but possibly different means. Our test statistic will be

$$\frac{\text{variation among groups}}{\text{variation within groups}}.$$

We begin with the following sample statistics:

Factor(Group)	Sample Sizes n_i	Sample Means $\bar{x}_i$	Sample s.d.'s s_i
1	7	82	12
2	7	66	16
3	6	60	13
$I = 3$	$N = 20$	overall $\bar{x} = 70$	

Note that the largest sample standard deviation, 16, is not more than twice the smallest, 12, justifying our assumption of equal population standard deviations.

Now we must establish how to measure variation *among* group means (from the overall mean) and variation *within* groups (from each group mean): we find the "mean sum of squared deviations" among and within groups as follows.

Sum of Squared deviations among Groups (SSG):

$$7(82-70)^2 + 7(66-70)^2 + 6(60-70)^2 = 1720 = SSG$$

In general, $SSG = \Sigma_{i=1}^{I} n_i(\bar{x}_i - \bar{x})^2$ weights each squared deviation of a group mean $\bar{x}_i$ from the overall mean $\bar{x}$ with the number of observations n_i for that group, then takes the overall sum. In general, I is the number of groups studied; we have $I = 3$.

Mean Sum of squared deviations among Groups (MSG): For our example, if we know 2 out of 3 deviations among groups $\bar{x}_i - \bar{x}$, we can solve for the last one, since group means $\bar{x}_i$ average out to the overall mean $\bar{x}$. Thus, our SSG has $3-1$ degrees of freedom: $DFG = 2$. In general, $DFG = I - 1$ and $MSG = \frac{SSG}{DFG}$. We have $MSG = \frac{1720}{2} = 860$.

Sum of Squared Error within groups (SSE): For summing up squared deviations within groups, we must calculate each observation x_{ij} minus its group mean $\bar{x}_i$, square this difference, and sum over all observations in that group. Finally, sum these up over all groups:

$$SSE = \Sigma_{i=1}^{I} \Sigma_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2.$$

If we already have sample standard deviations s_i calculated for each group, this saves some work. Since $s_i^2 = \frac{1}{n_i - 1} \Sigma_j (x_{ij} - \bar{x}_i)^2$, $(n_i - 1)s_i^2 = \Sigma_j (x_{ij} - \bar{x}_i)^2$, or

$$SSE = \Sigma_{i=1}^{I} (n_i - 1)s_i^2 = (n_1 - 1)s_1^2 + \cdots + (n_I - 1)s_I^2$$

We have

$$SSE = (7-1)12^2 + (7-1)16^2 + (6-1)13^2 = 3245$$

Mean Sum of squared Error within groups (MSE): Here we are comparing $(7+7+6) = 20$ observations with 3 sample means, so we have $20 - 3 = 17$ df; in general, if we have I groups and a total of N observations, our SSE has degrees of freedom $DFE = N - I$. Thus, we have

$$MSE = \frac{SSE}{DFE} = \frac{3245}{17} = 191$$

Ratio of Among to Within Group Variation (F): Our original goal was to examine the ratio of variation *among* groups to variation *within* groups. Take

$$F = \frac{MSG}{MSE} = \frac{860}{191} = 4.5$$

This ratio is our **F statistic**, used to test

H_0 : population means for all groups are equal $[\mu_1 = \mu_2 = \cdots = \mu_I]$

H_a : not all the μ_i are equal

Important: it is incorrect to write H_a as $\mu_1 \neq \mu_2 \neq \cdots \neq \mu_I$. This is *not* the logical opposite of H_0! The opposite of H_0 is to say that at least two (not necessarily all) of the means differ.

Example

> A physician's advice column in the Pittsburgh Post-Gazette (October, 2000) features the question, "Dear Doctor, Does everyone with Parkinson's disease shake?" and Dr. Kasdan's answer, "All patients with Parkinson's disease do not shake..." Is that what Dr. Kasdan really wanted to say?

If $\frac{MSG}{MSE}$ is close enough to 1, this indicates variation among groups is not significantly greater than variation within groups, and we cannot refute the claim that group means are equal. In other words, if the F statistic is close to 1, we cannot reject H_0. This is the sort of situation illustrated by the first configuration discussed at the beginning of class.

If $\frac{MSG}{MSE}$ is large, this would indicate a relatively large amount of variation among groups, and we have reason to doubt that group means are all equal. In other words, if the F statistic is much larger than 1, we reject H_0. This would be the case in the second configuration.

How large is "large" depends on the particular distribution of F involved.

Recall: There is only one standard normal distribution, regardless of sample size. There is one t distribution for each sample size n; it has $n - 1$ df.

Now, the distribution of F depends on *df* for the numerator $I - 1$ and *df* for the denominator $N - I$. It is right-skewed and always positive, with a peak near 1. Table A.4 shows values F^* and accompanying right-tail probabilities p for various degrees of freedom in the numerator and in the denominator.

ANOVA F Test

ANOVA stands for ANalysis Of VAriance. To test $H_0 : \mu_1 = \cdots = \mu_I$ in a one-way ANOVA, calculate the test statistic $F = \frac{MSG}{MSE}$. When H_0 is true, F has the $F(I-1, N-I)$ distribution. When H_a is true, F tends to be large. We reject H_0 in favor of H_a if the F statistic is sufficiently large, as determined by the P-value. The P-value is the probability that a R.V. having the $F(I-1, N-I)$ distribution would take a value greater than or equal to the calculated value of the F statistic. [We look at the upper tail only; F is never negative]

Small P-value$\Rightarrow$small probability$\Rightarrow$unlikely$\Rightarrow$ reject H_0 $\Rightarrow$conclude not all populations have the same mean.

Large P-value$\Rightarrow$not too unlikely$\Rightarrow$do not reject H_0 $\Rightarrow$ equal means possible.

A special case is when the number of groups $I = 2$, sample sizes n and population standard deviations σ are equal. Then $F = t^2$!