

MS Module 12: Expected values and β 's for ANOVA + unequal sample sizes – practice problems

(The attached PDF file has better formatting.)

Exercise 12.1: Expected value of treatment mean square

- An experiment has five groups with eight observations in each group.
- The five groups have the same population variance $\sigma^2 = 2$.
- An analysis of variance is done on the five groups to test the null hypothesis $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

The true means of the five groups are $\mu_1 = 1, \mu_2 = 2, \mu_3 = 3, \mu_4 = 4, \mu_5 = 5$, but these values are not known:

- A. What is the expected value of MSE, the mean square error?
- B. What are the deviations of the group means from the overall mean?
- C. What is the expected value of MSTr, the treatment mean square?
- D. What is the non-centrality parameter for the analysis of variance?

Part A: The expected value of MSE is $\sigma^2 = 2$.

The MSE is the mean sample variance in the groups. By assumption, all the groups have the same σ^2 .

Question: If we estimate the variance from the data, is the answer to this question the same?

Answer: The sample variance is an unbiased estimator of σ^2 . If we assume the variance is the same for all groups but we test whether the group means are the same, we use a weighted average of the sample group variances as the estimator of σ^2 .

Part B: The group means are written as $\mu_i = \mu + \alpha_i$, and the null hypothesis is $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$.

The overall mean $\mu = (1 + 2 + 3 + 4 + 5) / 5 = 3$, so the values of α for the five groups are

$$\alpha_1 = -2, \alpha_2 = -1, \alpha_3 = 0, \alpha_4 = 1, \alpha_5 = 2$$

Part C: The expected value of MSTr, the treatment mean square, is

$$\begin{aligned} E(\text{MSTr}) &= \sigma^2 + [J / (I - 1)] \times \sum \alpha_i^2 \\ \sum \alpha_i^2 &= (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 = 10, \Rightarrow \\ E(\text{MSTr}) &= 2 + [8 / (5 - 1)] \times 10 = 2 + 20 = 22 \end{aligned}$$

Question: Is the F value equal to $22 / 2 = 11$?

Answer: The practice problem gives the true group means μ_i and the true random error σ^2 . The F value is a statistic based on the observed sample.

Question: Is the expected F value equal to $22 / 2 = 11$?

Answer: We know the expected value of MSTr and the expected value of MSE. The expected value of the ratio is not exactly equal to the ratio of the expected values.

Part D: The non-centrality parameter for the one-way ANOVA is the quantity $J \times \sum \alpha_i^2 / \sigma^2 = 8 \times 10 / 2 = 40$.

The non-centrality parameter is used in the textbook to estimate the probability of a Type II error (β), which is inversely correlated with the value of the non-centrality parameter.

- The more the true group means differ from each other, the higher the value of $\sum \alpha^{2i}$ and the lower the probability of a Type II error.
 - If the true group means are the same, any rejection of the null hypothesis is an error.
 - If the true group means are very different (compared to the random error term), the probability of failing to reject the null hypothesis is low.
- A higher σ^2 increases the probability that the ANOVA will incorrectly attribute observed group differences to the random error term.
- More observations in each group reduces the probability of errors in the statistical analysis.

Exercise 12.2: Expected value of treatment mean square

- An experiment has five groups with eight observations in each group.
- The distributions in the five groups have the same population variance $\sigma^2 = 2$.
- An analysis of variance is done on the five groups to test the null hypothesis $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

The true means of the five groups are $\mu_1 = 0.1$, $\mu_2 = 0.2$, $\mu_3 = 0.3$, $\mu_4 = 0.4$, $\mu_5 = 0.5$, but these values are not known by the experimenter.

- A. What is the expected value of MSE, the mean square error?
- B. What are the deviations of the group means from the overall mean?
- C. What is the expected value of MSTr, the treatment mean square?
- D. What is the non-centrality parameter for the analysis of variance?

This problem is the same as the preceding one except that the input values are 10% as large.

Part A: The expected value of MSE is $\sigma^2 = 2$.

Part B: The group means are written as $\mu_i = \mu + \alpha_i$, and the null hypothesis is $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$.

The overall mean $\mu = (0.1 + 0.2 + 0.3 + 0.4 + 0.5) / 5 = 0.3$, so the values of α for the five groups are

$$\alpha_1 = -0.2, \alpha_2 = -0.1, \alpha_3 = 0, \alpha_4 = 0.1, \alpha_5 = 0.2$$

Part C: The expected value of MSTr, the treatment mean square, is

$$\begin{aligned} E(\text{MSTr}) &= \sigma^2 + [J / (I - 1)] \times \sum \alpha^{2i} \\ \sum \alpha^{2i} &= (-0.2)^2 + (-0.1)^2 + 0^2 + 0.1^2 + 0.2^2 = 0.10, \Rightarrow \\ E(\text{MSTr}) &= 2 + [8 / (5 - 1)] \times 0.10 = 2 + 0.20 = 2.20 \end{aligned}$$

The expected F value is close to $2.20 / 2 = 1.10$, and the p value is high; we do not reject the null hypothesis at any reasonable significance level.

Question: The two practice problems suggest that the ANOVA depends on the magnitude of the group means. But if we change the units of measurement – say from meters to millimeters or from miles to kilometers – we change $\sum \alpha^{2i}$ without affecting the real differences among the groups.

Answer: If we change the units of measurement, $\sum \alpha^{2i}$ and σ^2 change in the same proportions, so the F value does not change.

Part D: The non-centrality parameter for the one-way ANOVA is the quantity $J \times \sum \alpha^{2i} / \sigma^2 = 8 \times 0.10 / 2 = 0.40$.

Exercise 12.3: ANOVA (one-way) unequal group sizes

An experiment has three groups; the number of observations per group and the group mean are shown below.

	size	mean
group 1	7	25
group 2	8	21
group 3	6	29

- The sum of the squares of the observations is 14,000.
 - The observations have normal distributions in each group, and the variance in each group is the same.
 - The null hypothesis is that the means of the groups are the same: $H_0: \mu_1 = \mu_2 = \mu_3$
- A. What is the square of the sum of all the observations, or $x_{..}^2$?
- B. What is the correction factor for SST and SSTr?
- C. What is SST, the total sum of squares?
- D. What is SSTr, the treatment sum of squares?
- E. What is SSE, the error sum of squares?
- F. What are the total degrees of freedom?
- G. What are the degrees of freedom for the groups?
- H. What are the degrees of freedom for the error sum of squares (SSE)?
- I. What is MSTr, the mean squared deviation for the groups?
- J. What is MSE, the mean squared error?
- K. What is the F value for testing the null hypothesis?
- L. What is the p value for this test of the null hypothesis?

Part A: Analysis of variance uses sums of squares and squares of sums. Some final exam problems give the observations; other problems give sums and sums of squares. We compute two values.

The sums of the observations by group are the group size $\times$ the group mean. The sum of all the observations is $7 \times 25 + 8 \times 21 + 6 \times 29 = 517$. The square of the sum of the observations is $517^2 = 267,289$

Question: If the sample group means are similar, do we infer that the F value is close to one and we do not reject the null hypothesis? And if the sample group means differ greatly, do we infer that the F value is not close to one and we reject the null hypothesis?

Answer: By changes in the units of measurement (or any linear transformation of the observations), we can make the group mean differences smaller or larger without changing the ANOVA. Focus on the relative size of (i) the group mean differences and (ii) the standard deviation of the error term.

Part B: The correction factor = the square of the sum of the observations / the number of observations:

$$267,289 / 21 = 12,728.0476$$

Part C: SST (total sum of squares) = the sum of the squares of the observations – the correction factor:

$$14,000 - 12,728.0476 = 1,271.9524$$

The sum of the squares of the observations is given in the exercise; it is not derived from other figures here.

Part D: We compute SSTr in steps:

- Step #1: Compute the square of the sum of each group.

- Step #2: Divide these squares of sums by the sizes of the groups and add the quotients.
- Step #3: Subtract the correction factor.

The sum of (the squares of the group totals / the group size) = the sum of (the squares of the group means $\times$ the group size) = $7 \times 25^2 + 8 \times 21^2 + 6 \times 29^2 = 12,949$

	size	mean	sum	sum ²	sum ² /size
group 1	7	25	175	30625	4375
group 2	8	21	168	28224	3528
group 3	6	29	174	30276	5046
total	21		517		12949

The SSTr = $12,949 - 12,728.0476 = 220.9524$

Part E: We compute the error sum of squares SSE two ways:

The computation formula (expanding the definition of the SSE and simplifying): $SSE = \sum_i \sum_j x_{ij}^2 - \sum_i x_{i.}^2 / i_j$

= the sum of squares of all the observations

– the sum of (squares of the group totals $\div$ the number of observations in the group)

= $14,000 - 12,949 = 1,051$

Alternatively, we derive SSE by subtraction.

$SST = SSTr + SSE \Rightarrow SSE = SST - SSTr = 1271.9524 - 220.9524 = 1,051$

Part F: The degrees of freedom for the total sum of squares = number of observations – 1 = $21 - 1 = 20$.

Part G: The degrees of freedom for groups (SSTr) = number of groups – 1 = $3 - 1 = 2$.

Part H: The degrees of freedom for the error sum of squares (SSE) = the sum of (the number of observations in each group – 1) = $(7 - 1) + (8 - 1) + (6 - 1) = 18$.

Part I: MSTr, the mean squared deviation for the groups, is $220.9524 / 2 = 110.4762$.

Part J: MSE, the mean squared error, is $1,051 / 18 = 58.3889$.

Part K: The F value for testing the null hypothesis is $110.4762 / 58.3889 = 1.8921$

Part L: The cumulative distribution function for the F distribution with 2 and 18 degrees of freedom gives 0.820449 for a value of 1.282076, so the p value is $1 - 0.820449 = 0.179551$.