

Experimental Type Research Questions

By: Ir. Togar A. Napitupulu, MS., MSc., Ph.D

Consider the following two cases:

1. Implementing a newly improved marketing technique might be claimed to reach and persuade customers to buy on average more than 100 customers per month. Sample of sales person were trained with the new technique and observed their monthly selling.

The hypothesis for such experimentation is

$$H_0 : \mu = 100$$

$$H_1: \mu > 100$$

where μ is the population average of the number of customers persuaded by the sales.

From the sample of say, n sales person, we calculate the average sales and the sample variance. Using these two information t -statistics and p -value for one sided test (see H_1 , having one direction "less than") can be calculated. If this is less than $\alpha/2$, then we enough evidence from the sample to reject H_0 .

2. We might want to compare two different method of teaching or training our employee and we would like to know the most effective method in terms of their productivity. The two methods can be thought of as two different treatment in lab such that one group of employee is being treated with one of the method and the other group being treated with the other method. The hypothesis to be tested might be:

$$H_0: \mu_1 = \mu_2 \text{ versus}$$

$$H_1: \mu_1 > \mu_2$$

Because we are suspecting that method one is more effective than method two. μ_1 and μ_2 are the population mean of the first and the second method respectively. This can be tested using t -statistics (t test), testing the mean of two populations. Again we can use p -value as decision rule whether to reject H_0 or not enough evidence from sample to reject H_0 .

Both cases are called *one-factor problem*. The factor is the treatment with two levels. In the first case, the treatment is the newly improved marketing method and the samples were taken. The other level is without sample, i.e., simply a constant where the value can be thought of as had been derived from previously taken samples. In the second case, the levels are method 1

and method 2 and two sets of samples were taken from both. The following is a case where there are more than two levels.

3. **One-Way ANOVA:** Sometimes we might want to test three or more treatments, or to test the mean of three or more population. In such case we can not use *t-test* anymore. We have to resort to a method called Analysis of Variance (ANOVA). The hypothesis for such problem is:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k \text{ versus}$$

$$H_1: \text{At least two of them are not equal to one another}$$

where k is the number of treatments or populations. The test statistic used is *F-stat, or F – test*. Again *p-value* can be used to decide whether to reject H_0 or not enough evidence to reject H_0 . When H_0 is rejected, then we still need to test which particular pairs is the one that is not equal to one another. To test which of the pairs is not equal, we use “Tukey-Kramer Procedure”. When there are n populations to test or n means, then there are $[n(n - 1)]/2$ pairs to test. For example, we might want to test more than two methods of marketing and want to know which one of them that has the greater impact. Another example is we might want to know which of three or more methods of impact of training on our employee and which one is the best among them.

Notes on strategy of experimental design: Experimental units (such as students, mice, company, products, etc., provide the heterogeneity that leads to experimental error . To avoid this, we **randomly** assigned the factor levels to the experimental units.

ASSUMPTION: The k populations are independent and normally distributed with mean $\mu_1, \mu_2, \dots, \mu_k$ and common variance σ^2 . This assumption is supported by the randomization.

THE MODEL:

Let y_{ij} denote the j th observation from the i th treatment. Each observation may be written in the form:

$$Y_{ij} = \mu_i + \epsilon_{ij} \dots\dots\dots ()$$

Where ϵ_{ij} measures the deviation of the j th observation of the i th sample from the corresponding treatment mean – the same as the error term in regression. Another way of writing this equation is by substituting $\mu_i = \mu + \alpha_i$, subject to the constraint $\sum_{i=1}^k \alpha_i = 0$. Hence we may write

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}.$$

And α_i is called the effect of the i th treatment; μ is the **grand mean** of all the μ_i 's. The above hypothesis then now can be replaced by :

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0.$$

H_1 : At least one of the α_i 's is not equal to zero.

Testing the hypothesis is based on comparison of two independent estimates of the common population variance σ^2 based on the following Sum-of-squares Identity:

$$\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

That is

$$SST = SSA + SSE$$

Where

SST = total sum of squares.

SSA = treatment sum of squares.

SSE = error sum of squares (within treatment sum of squares).

It can be shown that expected value of SSA

$$E(SSA) = (k - 1)\sigma^2 + n \sum_{i=1}^k \alpha_i^2.$$

If H_0 is true, and thus each α_i in the above formula is equal to zero, then

$$E(s_1^2) = E\left(\frac{SSA}{k-1}\right) = \sigma^2,$$

Or that s_1^2 is an unbiased estimate of σ^2 . However, if H_1 is true, we have

$$E(s_1^2) = E\left(\frac{SSA}{k-1}\right) = \sigma^2 + \frac{n}{k-1} \sum \alpha_i^2,$$

that is, s_1^2 estimates σ^2 plus an additional term, which measures variation due to the systematic effects. We also know that another independent estimator of σ^2 based on $k(n-1)$ degrees of freedom is the familiar s^2 ,

$$s^2 = \frac{SSE}{k(n-1)}.$$

Now, when H_0 is true then, the ratio $F = s_1^2/s^2$ is a random variable having the F-distribution with $k-1$ and $k(n-1)$ degrees of freedom. Since s_1^2 overestimates σ^2 when H_0 is false, we have a one-tailed test with the critical region entirely in the right tail of the F

distribution; That is, we have F -ratio for testing equality of the means (see hypothesis above).

EXAMPLE #1

Four suppliers of materials for parachutes are using different fiber materials. We need to test whether the strength of the fibers produced by each suppliers are equal or not. The following is the data on the tensile strength of the materials.

Supplier 1	Supplier 2	Supplier 3	Supplier 4
18.5	26.3	20.6	25.4
24.0	25.3	25.2	19.9
17.2	24.0	20.8	22.6
19.9	21.2	24.7	17.5
18.0	24.5	22.9	20.4

The hypothesis then is:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_4 = 0.$$

H_1 : At least one of the α_i 's is not equal to zero.

Here is the result using EXCEL:

Anova: Single Factor

SUMMARY				
Groups	Count	Sum	Average	Variance
Supplier 1	5	97.6	19.52	7.237
Supplier 2	5	121.3	24.26	3.683
Supplier 3	5	114.2	22.84	4.553
Supplier 4	5	105.8	21.16	8.903

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	63.2855	3	21.095167	3.461629	0.041366	3.238872
Within Groups	97.504	16	6.094			
Total	160.7895	19				

Since $F_{\text{stat}} = 3.46$ is greater than $F_{\text{table}(3,16)} = 3.24$, hence we reject H_0 , or at least one of the α'_i is not equal to zero. We can also use p -value to make conclusion about the hypothesis, that is the probability of getting an F statistic of 3.4616 or larger when the null hypothesis is true, is 0.0414. Because p -value is less than the specified α of say 5 % or 0.05, we can reject the null hypothesis, or at least one or more population means are significantly different. To determine which suppliers differ, we can use a multiple comparison procedure such as **Tukey-Kramer Procedure** as follows.

The Tukey-Kramer Procedure

This procedure is one among the procedures available. First we compute the differences, $\bar{X}_i - \bar{X}_j$ (where $i \neq j$) among all pairs of means. Then we compute **critical range** using the following formula:

$$\text{Critical range} = Q_U \sqrt{\frac{MSW}{2} \left[\frac{1}{n_i} + \frac{1}{n_j} \right]}$$

Where Q_U is the upper-tail critical value from a **studentized range distribution** (See Table in one of the Text) having c degrees of freedom in the numerator and $n - c$ degrees of freedom in the denominator. A specific pair is significantly different if the absolute difference in the sample means $|\bar{X}_i - \bar{X}_j|$ is greater than the critical range. In the example above, the following are the six comparison:

1. $|\bar{X}_1 - \bar{X}_2| = |19.52 - 24.26| = 4.74$
2. $|\bar{X}_1 - \bar{X}_3| = |19.52 - 22.84| = 3.32$
3. $|\bar{X}_1 - \bar{X}_4| = |19.52 - 21.16| = 1.64$
4. $|\bar{X}_2 - \bar{X}_3| = |24.26 - 22.84| = 1.42$
5. $|\bar{X}_2 - \bar{X}_4| = |24.26 - 21.16| = 3.10$
6. $|\bar{X}_3 - \bar{X}_4| = |22.84 - 21.16| = 1.68$

$$\text{Critical range} = 4.05 \sqrt{\left[\frac{6.049}{2} \right] \left[\frac{1}{5} + \frac{1}{5} \right]} = 4.4712.$$

If we look at the above six alternative comparisons, only alternative 1, i.e., the difference between Supplier 1 and 2, that are significantly different at 5 % significant level.

4. Question might also be raised as to whether two or more factor interacting one another in affecting a variable of interest to researcher. For example, there might be an interaction of types of teaching or training the length of training time. Such research question can be analyzed using two-way classification experimental design or two way factorial experimentation if you have two factors that we think are interacted one another.

EXAMPLE #2

Suppose the two types of course (methods) are Traditional and Online. The length of course are Condensed and Regular. The company assigned 10 clients randomly to each of the four cells and record the students' score on ACT. The following are the data:

Method	Condensed	Regular	Method	Condensed	Regular
traditional	26	34	online	27	24
traditional	18	28	online	21	21
traditional	27	24	online	29	16
traditional	24	21	online	32	19
traditional	25	35	online	30	22
traditional	19	23	online	20	19
traditional	21	31	online	24	20
traditional	20	29	online	28	24
traditional	21	28	online	30	23
traditional	18	26	online	29	25

The result using **EXCEL** is as follows:

Anova: Two-Factor With Replication

SUMMARY	Condensed	Regular	Total
<i>traditional</i>			
Count	10	10	20
Sum	219	279	498
Average	21.9	27.9	24.9
Variance	11.2111111	20.9889	24.7263158
<i>online</i>			
Count	10	10	20
Sum	270	213	483
Average	27	21.3	24.15
Variance	16.2222222	8.01111	20.0289474
<i>Total</i>			
Count	20	20	
Sum	489	492	
Average	24.45	24.6	
Variance	19.8394737	25.2	

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	5.625	1	5.625	0.39870053	0.53175028	4.11316522
Columns	0.225	1	0.225	0.01594802	0.90020864	4.11316522
Interaction	342.225	1	342.225	24.2569403	1.8882E-05	4.11316522
Within	507.9	36	14.1083333			
Total	855.975	39				

The first to check is the interaction column in the ANOVA. The *p-value* = 0.0000188 which is less than 0.05 (5 %) significant level. Hence there is interaction effect between length of course and type of course (the two factors). Since the interaction is significant, then we cannot conclude on the two main effects, i.e., type of course and length of course. For example, we cannot conclude that there is no effect even though both have *p-values* greater than 0.05. To ensure the highest ACT score, the company should use the traditional approach when offering regular course (27.9 vs 21.9); but use the online approach when offering the condensed courses (27 vs.21.3).

- In marketing, we might want to study contribution of an **Attribute** (or **Factor**) of a product on the utility (or part-worth) for the customers. For example, a product might have the following factors: (1) Package Design, labeled A, B, C, (given possible value of 1, 2, or 3 by the researcher); (2) Brand name, labeled BR1, BR2, BR3 (again, given possible value of 1, 3, or 3, by the researcher); (3) Price, labeled P1, P2, and P3 (having possible value of Rp. 100 rb, Rp. 200 rb, or Rp. 400 rb.); (4) Good house-keeping seal, labeled, yes or no (having possible value of 1 or 2); and (5) money-back-guarantee, labeled ye or no (also having value of 1 or 2). In fact we can also calculate the contribution or utility of a combination of attribute with a particular combination of their levels. The statistical tool that can address such research question is the **“Conjoint (consider it jointly)” Analysis**. You can use SPSS to do this for you.

The model is :

$$Y_{ij} = \sum_{k=1}^K \sum_{m=1}^{M_k} \beta_{ikm} X_{jkm} + \varepsilon_{ij} \dots\dots\dots (3)$$

Where

J = product,

Y_{ij} = rating provided by respondent i on product j,

α_{ikm} = utility or part-worth of the mth level (m=1,2,... M_k) of the kth attribute,

M_k = number of level of the kth attribute,

K = number of attribute,

X_{jkm} = dummy variables that take on the value 1 if the m th level of the k th attribute is present in product j and the value 0 otherwise,

ε_{ij} = error term assumed to be normally distributed with zero mean and variance σ^2 .

6. ONE COMMON MISS-USED OF F-test (ANOVA) IN REGRESSION: Testing Simultaneous Effect of Independent Variables

In regression analysis of the following model (without loss of generality, let's have three independent variables):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \dots \dots \dots (1)$$

one might want to test whether the three independent variables have an impact simultaneously on the dependent variable Y . It is common to have this tested (mistakenly) by using the **F-test** from the ANOVA of the regression, or equivalently by looking at corresponding *p-value*. It is mistaken because the hypothesis corresponding to such *F-test* is:

$$\begin{aligned} H_0: & \beta_1 = \beta_2 = \beta_3 = 0 \text{ against} \\ H_1: & \text{At least one of the betas is not equal to zero} \end{aligned}$$

So for example if we reject H_0 , it only means that one or more of the β_i is not equal to zero and to know which one that is not equal to zero, can be tested partially using **t-test**.

How then we can test hypothesis to find out the effect of all independent variables simultaneously? We approach this by introducing a new variable into equation (1), that is, multiplication of all independent variables as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 X_3 + \varepsilon \dots \dots \dots (2)$$

Now, to test such hypothesis then is equivalent to testing

$$\begin{aligned} H_0: & \beta_4 = 0 \text{ against} \\ H_1: & \beta_4 \neq 0 \end{aligned}$$

If from the data there enough evidence to reject H_0 then it can be concluded that the three independent variables interact (*hence the term to test the "interaction" of the independent variables*) or simultaneously has an impact on the dependent variable Y . Researchers might want to test the impact of interaction of pairs of the independent variables. In such case, we would have to introduce three more independent variables into equation (2) as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1 X_2 X_3 + \varepsilon \dots (3)$$

And testing the impact of the interaction would be testing the corresponding betas, i.e., β_i , $i = 4, 5, 6, 7$.