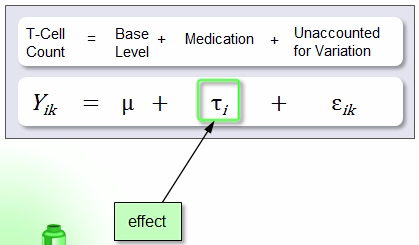
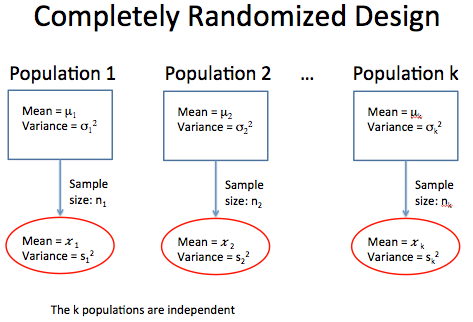
**ANOVA overview**

|  |  |
| --- | --- |
| ANOVA : The techniques used to analyze the means from a designed experiments. Variances of different sources are analyzed to make inferences. Different designs have different ANOVA tables. | |
| Purpose | To compare the means of several populations when independent samples have been chosen. |
| Assumption | * Samples are selected randomly and independently from the respective populations, * All *k* population variances are equal, and * All *k* population probability distributions are normal. |

* Balanced design – the same number of observations are made for each treatment.
* The fixed effect: This ANOVA model is appropriate when the k treatments being compared are the only treatments in which we are interested.





Error or

residual

Overall population mean

|  |  |  |
| --- | --- | --- |
| **ANOVA definitions** | | |
| Factor | Any feature thought to cause variation in response variable | |
| Level | Factor takes in particular trial | |
| Trial | Individual test run, resulting an observation on a given response variable | |
| Experiment | Complete set of trials | |
| Experimental Unit | The object that a trial is carried out | |
| Quantitative factor  Qualitative factor | Value can be arranged in order  Value can not …… | |
| Residual variation  or experimental error | Experimental variation caused by factors that have not been or can not be controlled by the experimenter | |
| Number of replication | The number that more than one observation be taken to estimate error | |
| Designed experiment  Observational experiment | | Analyst control the treatment and method of assigning unit  Analyst observe the treatment and response on units |
| Completely randomized design  Randomized block design | | Independent random samples of units are selected for treatment  Matched sets of units called block, each block with p(number of treatment) or k(unit number as similar as possible) unit, unit from each block is randomly assigned to each treatment, result in n=kp response |

**Hypothesis:**

|  |
| --- |
| Ho: µ1 = µ2 = µ3 = ………= µk  (The mean xxx for k treatment are the same)   * If Ho is true, each population mean µj is equal to the grand mean µ and there are no treatment effects. |
| Ha: at least two of the k treatment means differ |

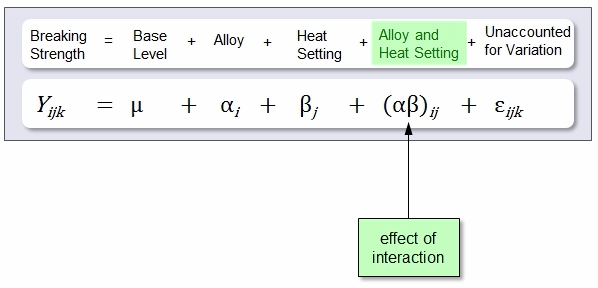
**Method:**

|  |  |
| --- | --- |
| The **variation between treatment means** is measured by the **sum of squares for treatments (SST),** which is calculated(between or across groups)    SST = | **The Sums of Squares for Error (SSE)** measures the **variability around the treatment means** that is attributed to sampling error, which is calculated using the following simplified formula  SSE = (within the groups) |
| SS(Total) = SST+SSE = SSbetween+SSwithin | |
| Note: k = number of groups or treatments and N = total number of observations     * Mean Square for Treatments (MST) =      * Mean Square for Error (MSE ) = * F = MST / MSE (between/within)   Values of the *F* statistics near 1 indicate that the two sources of variation, between treatment means and within treatments means are approximately equal. Thus, all F tests in ANOVA are right tailed tests. | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **The ANOVA Table**   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Source | DF | Sum Squares  SS | Mean Square  MS | F | | Treatments | k-1 | Σ(C2/n) - G2/N | MST | MST/MSE  F = Fv1/v2  v1=k-1  v2=N-k | | Error | N-k | By subtraction | MSE | | Total | N-1 | Σ(Σx2) – G2/N |  |   C2=(Total Value within each Treatment)2 = (Sum of each column)2  G2=(Total Value across treatments)2=(Sum of (sum of each column))2  Σx2=Total Value of squared each value from treatment=Sum of (value)2 from treatment  N=Total obersvation number n=each number under each treatment k=#n of treatment |
| **Assumptions Check:**   1. **Samples are selected independently and assigned randomly into the groups** 2. **All *k* population probability distributions are normal**   -Check graphically using a normal probability plot of the residuals and if the points follow the line we are comfortable that the error terms are approximately normal. If the point from QQplot of residuals do not follow the line, this may cause some concerns but as long as the deviation is slight we can proceed, since  -ANOVA is robust to moderate departures in the normally assumption.   1. **All variances of k treatment are equal**   -Check graphically by plotting the residuals versus the fitted values and compare the distributions. If there is no pattern on the scatter pot then we are good. If there is, then no good.  -Or using Levene’s Test for Homogeneity of Variance (high p value-var are same)  -Note: ANOVA is not robust with respect to departures to the equal variance assumption. The effect is slight when sample sizes are equal. |
| **Conduct a complete analysis** to determine if there is a difference in the mean  1)State Ho and Ha and define parameter  2)Check that all the assumptions for inference have been satisfied.  3)Use the computational formulas to complete an ANOVA table  4)Use the results in your ANOVA table to make a decision regarding the Ho  5)Follow up by Tuckey HSD test to see if any mean significantly differ from others  6) Interpret your finding in the context of the question. |

|  |  |
| --- | --- |
| P<0.05 or other significance level(a) | Reject the null |
| P>0.05 or other significance level(a) | Fail to reject the null |

**Two-Way ANOVA overview**



Effect B

Effect A

error

Response variable

* = the grand mean for X

Ai = the effect for the ith level of Factor A (i=1…..r)

βj = the effect for the jth level of Factor B (j=1……c)

Qij= the interaction effect for the ith level of A with the jth level of B

* = the random error

|  |  |
| --- | --- |
| Two Way ANOVA : The techniques used to analyze the means from a designed experiments Variances of different sources are analyzed to make inferences. Different designs have different ANOVA tables. | |
| Purpose | * The purpose of the test is to determine the effects of two factors on the response variable and, if there is interaction between the factors, to determine the nature of the interaction. |
| Assumption | * Samples are selected randomly and independently from the respective populations, * All *k* population variances are equal, and * All *k* population probability distributions are normal. |

* Balanced design – the same number of observations are made for each treatment.
* The factor interaction component is used to test whether the factors combine to affect the response while the factor main effect components are used to determine whether the factors separately affect the response.

**Hypothesis and Method:**

|  |  |
| --- | --- |
| Ho : All kj=0 ; Ha: Some kj ≠0 (Interaction)  **~The factor of A and B do not/do interact with each other to affect the response mean**  Ho : All kj=0 ; Ha: Some kj ≠0 (Factor A): **There is no/a difference between the mean xx for X**  Ho : All kj=0 ; Ha: Some kj ≠0 (Factor B): **There is no/a difference between the mean xx for Y** | |
| Test the first hypothesis initially. It is important to know whether or not interactions exist.  -If interactions do exist, our interpretation of the ANOVA results will be different from the case where no interactions exist.  -If there are not interactions and significant main effects, further analysis on the main effects may be conducted using the LSD post-hoc analysis or Fisher’s LSD test. | |
| Assumption Check | | 1. **Random and independent samples of experimental units are associated with each treatment.**   -Yes, we are assuming that subjects were independently and randomly selected assigned to the treatment groups.   1. **The residuals should be normally distributed with mean zero.**   -Check graphically using a normal probability plot of the residuals and if the points follow the line we are comfortable that the error terms are approximately normal. If the point from QQplot of residuals do not follow the line, this may cause some concerns but as long as the deviation is slight we can proceed, since  -ANOVA is robust to moderate departures in the normally assumption.  **3) The response variance is constant for all treatment combinations.**  -Conduct a scatterplot of residuals vs fitted values, if there is no pattern on the scatter pot then we are good. If there is, then no good.  -Or use 2-way Levene’s Test with a high p value to prove the var are the same  -Note: ANOVA is not robust with respect to departures to the equal variance assumption. The effect is slight when sample sizes are equal. |
| Interpretation  If there is an interaction between factor A and B | | Because the factors interact, we do not test the main effects for Brand and Club. We will compare the treatment means in an attempt to learn the nature of the interaction.   * Examine treatment means * conduct Tukey’s HSD test * Construct and examine interaction plot |

|  |  |
| --- | --- |
| P<0.05 or other significance level(a) | Reject the null |
| P>0.05 or other significance level(a) | Fail to reject the null |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **The Two way-ANOVA Table**   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Source of Variation** | **Degrees of Freedom=df** | **Sum of Squares=SS** | **Mean Square=MS** | **Variance Ratio=F** | | Factor A  Between Rows | r-1 | SSA=  ΣR2/nc - G2/N | MSA= |  | | Factor B  Between Columns | c-1 | SSB=  ΣC2/nr - G2/N | MSB= |  | | A\*B  Interaction | (r-1)(c-1) | SSAB=  Bysubtraction | MSAB= |  | | Residual or Error | rc(n-1) | SSE=  ΣX2 – ΣT2/n | MSE= |  | | Total | rcn-1 | ΣX2 – G2/N |  |  | |

\*LSD post-hoc analysis or Fisher’s LSD test: LSD = ta×, a=df of Error