

ANOVA F-Test: Completely Randomized Design

(n = total number of observations, k = Number of treatments, & T = total for treatment)

Step 1: Express the Claim

Step 2: The Hypotheses: $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$
 $H_A : \text{At least two means differ}$

Step 3: Work with the data

Correction Factor: $CF = \frac{(\sum y_i)^2}{n} = \frac{(\text{Sum of all observations})^2}{n}$

Sum of Squares Total: $SS(Total) = \sum y_i^2 - CF$
= (Square each observation then add them up) – CF

Sum of Squares for Treatments: $SST = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \dots + \frac{T_k^2}{n_k} - CF$
= (Square each treatment total and divided by the number of observations in each treatment then add those results up) – CF

Sum of Square for Error: $SSE = SS(Total) - SST$

Mean Square Treatment: $MST = \frac{SST}{k - 1}$

Mean Square Error: $MSE = \frac{SSE}{n - k}$

Step 4: the Test Statistic: $F = \frac{MST}{MSE}$

Step 5: the Critical Value: In order to properly form our conclusion, we need either a p-value or a critical value. To get the critical value, we will consider how our F-test stat is formed. It is the ratio of MST and MSE. The MST is on top, so its degrees of freedom will be the numerator degrees of freedom. The degree of freedom for MSE will serve as the denominator degree of freedom (we need these quantities for the F-table). Then, we need an alpha value.

Step 6: Initial Conclusion: If $F = \frac{MST}{MSE} > f_{k-1, n-k, \alpha}$, we reject the null. *Note the F-test for ANOVA is always right-tailed since MST is always the top of the fraction which forms our F-statistic.

Step 7: Word the Final Conclusion

Put all of these calculations into the ANOVA table:

Source	Df	SS	MS	F
Treatments	$K - 1$	SST	$MST = SST/k-1$	MST/MSE
Error	$N - k$	SSE	$MSE = SSE/n-k$	
Total	$N - 1$	SS(total)		

Multiple Comparisons of the Means

Number of comparisons = $\binom{k}{2} = \frac{k!}{(k-2)!2!} = \frac{k(k-1)}{2}$ where k = number of treatments

Confidence interval for a treatment mean:

$\bar{X}_T \pm t_{\alpha/2} S \sqrt{1/n_T}$ where $S = \sqrt{MSE}$, n = number of repetitions for the treatment, and t has degrees of freedom = $n - k$

The Randomized Block Design

Step 1: Express the Claim: For example, all of the treatment means are the same.

Step 2: The Hypotheses:
 $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$
 $H_A : \text{At least two means differ}$

Step 3: Work with the data

Correction Factor: $CF = \frac{(\sum y_i)^2}{n} = \frac{(\text{Sum of all observations})^2}{n}$

Sum of Squares Total: $SS(\text{Total}) = \sum y_i^2 - CF$
= (Square each observation then add them up) – CF

Sum of Squares for Treatments: $SST = \frac{T_1^2}{b} + \frac{T_2^2}{b} + \dots + \frac{T_k^2}{b} - CF$ = (Sum of squared Treatment totals divided by the number of observations in each treatment) – CF

Sum of Squares for Blocks: $SSB = \frac{B_1^2}{k} + \frac{B_2^2}{k} + \dots + \frac{B_k^2}{k} - CF$ = (Sum of squared block totals divided by the number of observations in each block) – CF

Sum of Square for Error: $SSE = SS(\text{Total}) - SST - SSB$

Mean Square Treatment: $MST = \frac{SST}{k-1}$

Mean Square for Blocks: $MSB = \frac{SSB}{b-1}$

Mean Square Error: $MSE = \frac{SSE}{n - k - b + 1}$

Step 4: the Test Statistics: $F = \frac{MST}{MSE}$ & $F = \frac{MSB}{MSE}$

Step 5: the Critical Value: In order to properly form our conclusion, we need either a p-value or a critical value. To get the critical value, we will consider our how our F-test stat is formed. If it is the ratio of MST and MSE, then we use the degrees of freedom for treatments as the numerator degrees of freedom and the degree of freedom for Error will serve as the denominator degree of freedom (we would use the df for blocks as our numerator degrees of freedom if our ratio was MSB/MSE).

Step 6: Initial Conclusion: If $F = \frac{MST}{MSE} > f_{k-1, n-b-k+1, \alpha}$, we reject the null. If $F = \frac{MSB}{MSE} > f_{b-1, n-b-k+1, \alpha}$, we conclude the blocks are significantly different.

Step 7: Word the Final Conclusion

Place the answers to our calculations in an ANOVA table:

Source	Df	SS	MS	F
Treatment	K - 1	SST	MST	MST/MSE
Block	B - 1	SSB	MSB	MSB/MSE
Error	N - k - b + 1	SSE	MSE	
Total	N - 1	SS(total)		

Fitting the Model: The Least Squares Approach $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

Step 1: fill out the following table

Preliminary computations table				
	x_i	y_i	x_i^2	$x_i y_i$

Totals	$\sum x_i$	$\sum y_i$	$\sum x_i^2$	$\sum x_i y_i$

Step 2: Calculate $SS_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$ and $SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

Step 3: Calculate the Slope: $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$

Step 4: Calculate the y-intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Finding an estimate for σ^2

Steps to finding S^2 an estimate of σ^2 :

1. Create and fill in the preliminary calculation table below:

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	\vdots	\vdots	\vdots	\vdots	\vdots
Totals	$\sum x_i$	$\sum y_i$	$\sum x_i^2$	$\sum y_i^2$	$\sum x_i y_i$

2. Calculate $SS_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

3. Calculate $SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

4. $SS_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$

5. Find $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$

6. Calculate $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$

7. Finally, $S^2 = \frac{SSE}{n-2}$

Testing the Model's Utility by Testing the Slope

One Tailed Test

$$H_o : \beta_1 \geq 0 \text{ or } (H_a : \beta_1 \leq 0)$$

$$H_a : \beta_1 < 0 \text{ or } (H_a : \beta_1 > 0)$$

$$\text{Test Statistic: } t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{s / \sqrt{SS_{xx}}}$$

$$\text{Rejection region: } t < -t_a \quad \text{Or } (t > t_a \text{ when } H_a : \beta_1 > 0)$$

Where t_a is based on (n-2) degrees of freedom

Two Tailed Test

$$H_o : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

$$\text{Test Statistic: } t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{s / \sqrt{SS_{xx}}}$$

Rejection region: $|t| > t_{a/2}$ Where $t_{a/2}$ is based on (n-2) degrees of freedom

Step 1: Form the claim symbolically

Step 2: Get your Hypotheses

Step 3: Find SS_{yy} , SS_{xx} , SS_{xy} , $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$, $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$, $s = \sqrt{\frac{SSE}{n-2}}$, and $s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{xx}}}$

Step 4: Find the Test Stat $t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{s / \sqrt{SS_{xx}}}$

Step 5: Find your critical t-value by looking up α on the t-table with $n - 2$ d.f.

Step 6: Form your initial conclusion

Step 7: State your final conclusion

A 100(1-a)% Confidence Interval for the Sample Linear Regression Slope β_1

$$\hat{\beta}_1 \pm t_{a/2} s_{\hat{\beta}_1}$$

Where the estimated standard error of $\hat{\beta}_1$ is calculated by $s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{xx}}}$ and $t_{a/2}$ is based on $(n-2)$ degrees of freedom.

The Coefficient of Correlation

The coefficient of correlation, $r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$ is a measure of the strength of the linear relationship between two variables x and y.

The coefficient of determination is

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}} = \frac{\text{Explained sample variability}}{\text{Total sample variability}}$$

Steps to finding r and r^2 :

1. Create and fill in the preliminary calculation table below:

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	\vdots	\vdots	\vdots	\vdots	\vdots
Totals	$\sum x_i$	$\sum y_i$	$\sum x_i^2$	$\sum y_i^2$	$\sum x_i y_i$

2. Calculate $SS_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

3. Calculate $SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

4. $SS_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$
5. Find $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$
6. Calculate $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$
7. Calculate $r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$ and calculate $r^2 = \frac{SS_{yy} - SSE}{SS_{yy}}$

Using the model for estimation and prediction

1. Our estimation intervals to estimate the average value of y for a specific value of x

$$\hat{y} \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

where $t_{\alpha/2}$ is based on n-2 degrees of freedom.

2. Our prediction interval to estimate a specific y value for a given x value

$$\hat{y} \pm t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

where $t_{\alpha/2}$ is based on n-2 degrees of freedom.

Steps to forming a prediction interval (the steps are nearly the same for the estimation interval):

Step 1: Use the least squares line to find \hat{y} by plugging x into the equation

Step 2: Find $t_{\alpha/2}$ using degrees of freedom is n - 2

Step 3: Find $S \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$

Step 4: Find the Margin of Error = ME = $t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} = t_{\alpha/2}$ (the result from step 3)

Step 5: Finish by getting: $[\hat{y} - ME, \hat{y} + ME]$