STS 414 ANALYSIS OF VARIANCE (ANOVA)

REVIEW OF SIMPLEREGRESSION

Modeling refers to the development of mathematical expressions thatdescribe in some sense the behavior of a random variable of interest. Thisvariable may be the price of wheat in the world market, the number ofdeaths from lung cancer, the rate of growth of a particular type of tumor,or the tensile strength of metal wire. In all cases, this variable is called the**dependent variable** and denoted with Y_i. A subscript on Y identifies theparticular unit from which the observation was taken, the time at whichthe price was recorded, the county in which the deaths were recorded, theexperimental unit on which the tumor growth was recorded, and so forth.Most commonly the modeling is aimed at describing how the **mean** of thedependent variable $E(Y)$ changes with changing conditions; the varianceof the dependent variable is assumed to be unaffected by the changingconditions.

Other variables which are thought to provide information on the behaviour of the dependent variable are incorporated into the model as predictor orexplanatory variables. These variables are called the **independent variables**and are denoted by X with subscripts as needed to identify differentindependent variables. Additional subscripts denote the observational unitfrom which the data were taken. The Xs are assumed to be known constants. In addition to the Xs, all models involve unknown constants, called**parameters**, which control the behavior of the model. These parametersare denoted by Greek letters and are to be estimated from the data.

The mathematical complexity of the model and the degree to whichit is a realistic model depend on how much is known about the processbeing studied and on the purpose of the modeling exercise. In preliminarystudies of a process or in cases where prediction is the primary objective,the models usually fall into the class of models that are **linear in theparameters**. That is, the parameters enter the model as simple coefficientson the independent variables or functions of the independent variables.Such models are referred to loosely as **linear models**. The more realisticmodels, on the other hand, are often **nonlinear in the parameters**. Mostgrowth models, for example, are nonlinear models. Nonlinear models fallinto two categories: **intrinsically linear models**, which can be linearizedby an appropriate transformation on the dependent variable, and thosethat cannot be so transformed.

The Linear Model and Assumptions

The simplest linear model involves only one independent variable and states that the true mean of the dependent variable changes at a constant rateas the value of the independent variable increases or decreases. Thus, thefunctional relationship between the true mean of Y_i, denoted by $E(Y_i)$, and$X_i$is the equation of a straight line:

$$E(Y_i) = \beta_0 + \beta_1 X_i.$$

β_0 is the intercept, the value of $E(Y_i)$ when $X = 0$, and β_1 is the slope of the line, the rate of change in $E(Y_i)$ per unit change in X.

The observations on the dependent variable $Y_i$are assumed to be random observations from populations of random variables with the mean of eachpopulation given by $E(Y_i)$. The deviation of an observation $Y_i$from itspopulation mean $E(Yi)$ is taken into account by adding a random error $e_i$to give the statistical model

$$Y_i = \beta_0 + \beta_1 X_i + e_i.$$

The subscript i indicates the particular observational unit, $i = 1, 2, \ldots, n$.The $X_i$are the n observations on the independent variable and are assumedto be measured without error. That is, the observed values of X are assumedto be a set of known constants. The $Y_i$and $X_i$are paired observations; bothare measured on every observational unit.

The random errors $e_i$have zero mean and are assumed to have commonvariance σ^2 and to be pairwise independent. Since the only random elementin the model is e_i, these assumptions imply that the $Y_i$also have commonvariance σ^2 and are pairwise independent. For purposes of making testsof significance, the random errors are assumed to be normally distributed,which implies that the $Y_i$are also normally distributed. The random errorassumptions are frequently stated as

$$e_i \sim NID(0, \sigma^2),$$

where NID stands for "normally and independently distributed." The quantitiesin parentheses denote the mean and the variance, respectively, of thenormal distribution.

Least Squares Estimation

The simple linear model has two parametersβ_0 and β_1, which are to beestimated from the data. If there were no random error in Y_i, any two datapoints could be used to solve explicitly for

the values of the parameters.The random variation in *Y*, however, causes each pair of observed datapoints to give different results. (All estimates would be identical only if theobserved data fell exactly on the straight line.) A method is needed thatwill combine all the information to give one solution which is "best" bysome criterion.

The **least squares estimation procedure** uses the criterion that the solution must give the smallest possible sum of squared deviations of the observed Y_i from the estimates of their true means provided by the solution. Let β_0 and β_1 be numerical estimates of the parameters β_0 and β_1, respectively, and let

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

be the estimated mean of Y for each X_i, $i = 1, \ldots, n$. Note that $\hat{Y}_i$ is obtained by substituting the estimates for the parameters in the functional form of the model relating $E(Yi)$ to X_i. The least squares principlechooses $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squares of the residuals,SS(Res):

$$SS(\text{Res}) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$= \sum e_i^2$$

where$e_i = (Y_i - \hat{Y}_i)$ is the observed residual for the *ith* observation. Thesummation indicated by $\sum$ is over all observations in the data set as indicated by the index of summation, $i = 1$ to n. (The index of summation is omitted when the limits of summation are clear from the context.)

The estimators for β_0 and β_1 are obtained by using calculus to find thevalues that minimize SS(Res). The derivatives of SS(Res) with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ in turn are set equal to zero. This gives two equations in twounknowns called the **normal equations**:

$$n(\hat{\beta}_0) + (\sum X_i)\hat{\beta}_1 = \sum Y_i$$

$$(\sum X_i)\hat{\beta}_0 + (\sum X_i^2)\hat{\beta}_1 = \sum X_i Y_i.$$

Solving the normal equations simultaneously for $\hat{\beta}_0$ and $\hat{\beta}_1$ gives the estimates

ofβ_0 and $\beta_1$as

$$\hat{\beta}_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}.$$

Note that $x_i = (X_i - \overline{X})$ and $y_i = (Y_i - \overline{Y})$ denote observations expressedas deviations from their sample means $\overline{X}$ and $\overline{Y}$, respectively. The moreconvenient forms for hand computation of sums of squares and sums ofproducts are

$$\sum x_i^2 = \sum X_i^2 - \frac{(\sum X_i)2}{n}$$

$$\sum x_i y_i = \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}$$

Thus, the computational formula for the slope is

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

These estimates of the parameters give the regression equation

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

ANALYSIS OF VARIANCE (ANOVA)

Analysis of Variance (ANOVA) was introduced by Sir Ronald Fisher and is essentially an arithmetic process for partitioning a total sum of squares into components associated with recognized source of variation. It has been used to advantage in all fields of research where data are measured quantitatively. Suppose in an industrial experiment that an engineer is interested in how the mean absorption of moisture in concrete varies among 5 different concrete aggregate.

The samples are exposed to moisture for 48 hours. It decided that 6 tested. The data are presented in Table 1.

The model for this situation is considered as follows. There are 6 observations taken from each of 5 populations with means $\mu_1, \mu_2, ..., \mu_5$, respectively. We may wish to test

$$H_0 : \mu_1 = \mu_2 = ... = \mu_5,$$

H_1 : At least two of the means are not equal.

Aggregate:	1	2	3	4	5	
	551	595	639	417	563	
	457	580	615	449	631	
	450	508	511	517	522	
	731	583	573	438	613	
	499	633	648	415	656	
	632	517	677	555	679	
Total	3320	3416	3663	2791	3664	16,854
Mean	553.33	569.33	610.50	465.17	610.67	561.80

In addition, we may be interested in making individual comparisons among these 5 population means.

Two Sources of Variability in the Data

In the ANOVA procedure, it is assumed that whatever variation exist between the aggregate average is attributed to (1) variation in absorption among observations within aggregate types, and (2) variation due to aggregate type, that is, dueto differences in the chemical composition of the aggregates. The within aggregate variation is, of course, brought about by various causes. Perhaps humidity and temperature conditions were kept entirely constant throughout the experiment. It is possible that there was a certain amount of heterogeneity in the batches of raw materials that were used. At any rate, we shall consider the within sample variation to be chance or random variation, and part of the goal of the ANOVA is to determine if the differences among the 5 sample means are what we would expect due to random variation alone.

Many pointed questions appear at this stage concerning the preceding problem. For example, how many samples must be tested for each aggregate? This is a question that continually haunts

the practitioner. In addition, what if the within sample variation is so large that it is difficult for a statistical procedure to detect the systematic differences? Can we systematically control extraneous sources of variation and thus remove them from portion we call random variation? We shall attempt to answer these and other questions in this course.

Completely Randomized Design (One-Way ANOVA)

Random samples of size *n* are selected from each of *k*populations. The *k*different populations are classified on the basis of a single criterion such as different treatments or groups. Today the term treatment is used generally to refer to the various classifications, whether they are different aggregates, different analysts, different fertilizers, or different regions of the country.

Assumptions and Hypotheses in One-Way ANOVA

It is assumed that the *k*populations are independent and normally distributed with means $\mu_1, \mu_2, ..., \mu_k$ and common variance σ^2. These assumptions are made more palatable by randomization. We wish to derive appropriate methods for testing the hypothesis

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

H_1 : At least two of the means are not equal.

Let y_{ij} denote the j^{th} observation from the i^{th} treatment and arrange the data as in Table 2. Here, $Y_{i.}$ is the total of all observations in the sample from the i^{th} treatment, $\bar{y}_{i.}$ is the mean of all observations in the sample from the i^{th} treatment, $Y_{..}$ is the total of all nk observations, and $\bar{y}_{..}$ is the mean of all nk observations.

Model for One-Way ANOVA

Each observation may be written in the form

$$Y_{ij} = \mu_i + \varepsilon_{ij},$$

where ε_{ij} measures the deviation of the j^{th} observation of the i^{th} sample from the corresponding treatment mean. The ε_{ij}-term represents random error and plays the same role as the error terms in the regression models. An alternative and

Table 2. *k*Random Samples

Treatment	**1**	**2**	**.**	**.**	**.**	***i***	**.**	**.**	**.**	***k***	
	y_{11}	y_{21}	.	.	.	y_{i1}	.	.	.	y_{k1}	
	y_{12}	y_{22}	.	.	.	y_{i2}	.	.	.	y_{k2}	
	.	.								.	
	.	.								.	
	.	.								.	
	y_{1n}	y_{2n}	.	.	.	y_{in}	.	.	.	y_{kn}	
Total	$Y_{1.}$	$Y_{2.}$	.	.	.	$Y_{i.}$	.	.	.	$Y_{k.}$	$Y_{..}$
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	.	.	.	$\bar{y}_{i.}$	.	.	.	$\bar{y}_{k.}$	$\bar{y}_{..}$

Preferred form of this equationis obtained by substituting $\mu_i = \mu + \alpha_i$, subject to the constraint $\sum_{i=1}^{k} \alpha_i = 0$. Hence we may write

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

Where μ is just the grand mean of all the μ_i's; that is,

$$\mu = \frac{1}{k}\sum_{i=1}^{k} \mu_i,$$

and α_i is called the effect of the i^{th} treatment.

The null hypothesis that the *k*population means are equal against the alternative that at least two of the means are unequal may now be replaced by the equivalent hypothesis.

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

H_1 : At least two of the α_i's are not equal zero.

Resolution of Total Variability into Components

Our test will be based on a comparison of two independent estimates of the common population variance σ^2. These estimates will be obtained by partitioning the total variability of our data, designated by the double summation

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

into two components.

Theorem 13.1: Sum-of-squares Identity

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

It will be convenient in what follows to identify the terms of the sum-of-squares identify by the following notation:

Three Important Measures of Variability

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = \text{total sum of squares,}$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 = \text{treatment sum of squares,}$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 = \text{error sum of squares,}$$

The sum-of squares identity can then be represented symbolically by the equation

$$SST = SSA + SSE.$$

F-Ratio for Testing Equality of Means

When H_0 is true, the ratio $f = \frac{s_1^2}{s^2}$ is a value of the random variable *F* 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 distribution.

The null hypothesis H_0 is rejected at the $\alpha - level$ of significance when

$$f > f_{\alpha}[k-1, k(n-1)].$$

Another approach, the P-value approach, suggests that the evidence in favour of or against H_0 is

$$P = P[\, f[k-1, k(n-1)] > f\,].$$

The computations for ANOVA problem are usually summarized tabular form as shown in Table 3.

ANOVA for the One-Way ANOVA

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	SSA	k-1	$s_1^2 = \frac{SSA}{k-1}$	$\frac{s_1^2}{s^2}$
Error	SSE	k(n-1)	$s^2 = \frac{SSE}{k(n-1)}$	
Total	SST	Kn-1		

Example 1. Test the hypothesis $\mu_1 = \mu_2 = ... = \mu_5$ at the 0.05 level of significance for the data of Table 1 on absorption of moisture by various types of cement aggregates.

Solution: $H_0 : \mu_1 = \mu_2 = ... = \mu_5,$

H_1 : At least two of the means are not equal.

$\alpha = 0.05$ Critical region: $f > 2.76$ with $v_1 = 4$ and $v_2 = 25$ degrees of freedom. The sum of squares computations give

SST=209,377 SSA=85,356 SSE=124,021.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	85356.47	4	21339.12	4.30
Error	124020.33	25	4960.81	
Total	209376.80	29		

Decision: Reject H_0 and conclude that the aggregates do not have the same mean absorption.

Randomized Complete Block Designs

A typical layout for the randomized complete block design (RCB) using 3 measurements in 4 blocks is as follows:

Block 1 Block 2 Block 3 Block 4

t_2	t_1	t_3	t_2
t_1	t_3	t_2	t_1
t_3	t_2	t_1	T_3

The t's denote the assignment to blocks of each of the 3 treatments. Of course, the true allocation of treatments to units within blocks is done at random. Once the experiment has been completed, the data can be recorded as in the following

Treatment Block	1	2	3	4
1	y_{11}	y_{12}	y_{13}	y_{14}
2	y_{21}	y_{22}	y_{23}	y_{24}
3	y_{31}	y_{32}	y_{33}	y_{34}

where y_{11} represents the response obtained by using treatment 1 in block 1, y_{12} represents the response obtained by using treatment 1 in block 2, . . . and y_{34} represents the response obtained by using treatment 3 in block4.

Let us now generalize and consider the case of *k*treatments assigned to *b*blocks. The data may summarized as shown in the $k \times b$ rectangular array of Table 4. It will be assumed that the y_{ij}, *i=1,2,…,k* and *j=1,2,…,b,* are values of independent random variables having normal distributions with means μ_{ij} and common variance σ^2.

Table 4. $k \times b$ Array for the RCB Design

Block

Treatment	1	2	.	.	.	j	.	.	.	B	Total	Mean
1	y_{11}	y_{12}	.	.	.	y_{1j}	.	.	.	y_{1b}	$T_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	.	.	.	y_{2j}	.	.	.	y_{2b}	$T_{2.}$	$\bar{y}_{2.}$
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
I	y_{i1}	y_{i2}	.	.	.	y_{ij}	.	.	.	y_{ib}	$T_{i.}$	$\bar{y}_{i.}$
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
K	y_{k1}	y_{k2}	.	.	.	y_{kj}	.	.	.	y_{kb}	$T_{k.}$	$\bar{y}_{k.}$
Total	$T_{.1}$	$T_{.2}$	.	.	.	$T_{.j}$	.	.	.	$T_{.b}$	$T_{..}$	
Mean	$\bar{y}_{.1}$	$\bar{y}_{.2}$	.	.	.	$\bar{y}_{.j}$	.	.	.	$\bar{y}_{.b}$		$\bar{y}_{..}$

Let $\mu_{i.}$ represent the average (rather than the total) of the *b* population means for the *ith*treatment. That is,

$$\mu_{i.} = \frac{1}{b}\sum_{j=1}^{b}\mu_{ij}.$$

Similarly, the average of the population means for the *jth* block, $\mu_{.j}$, is defined by

$$\mu_{.j} = \frac{1}{k}\sum_{i=1}^{k}\mu_{ij},$$

and the average of the *bk*population means μ is defined by

$$\mu = \frac{1}{bk}\sum_{i=1}^{k}\sum_{j=1}^{b}\mu_{ij}.$$

To determine if part of the variation is our observations is due to differences among the treatments, we consider the test

$H_0 = \mu_{1.} = \mu_{2.} = ... = \mu,$

H_1 : The $\mu_{i.}$'s are not all equal

Model for the RCB Design

Each observation may be written in the form

$$y_{ij} = \mu_{ij} + \varepsilon_{ij},$$

where ε_{ij} measure the deviation of the observed value y_{ij} from the population mean μ_{ij}. The preferred form of this equation is obtained by substituting

$$\mu_{ij} = \mu + \alpha_i + \beta_j,$$

where α_i is, as before, the effect of the *ith* treatment and β_j is the effect of the *jth* block. It is assumed that the treatment and block effects are additive. Hence we may write

$$\mathbf{y}_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}.$$

The basic concept is much like that of the one way classification except that we must account in the analysis for the additional effect due to blocks, since we are now systematically controlling variation in two directions.

ANOVA for the Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	SSA	k-1	$s_1^2 = \frac{SSA}{k-1}$	$f_1 = \frac{s_1^2}{s^2}$
Blocks	SSB	b-1	$s_2^2 = \frac{SSA}{b-1}$	
Error	SSE	(k-1)(b-1)	$s^2 = \frac{SSE}{(k-1)(b-1)}$	
Total	SST	kb-1		

Example 2.Four different machines, M_1, M_1, M_3 and M_4, are being considered for the assembling of a particular product. It is decided that 6 different operators are to be used in a randomized block experiment to compare the machines. The machines are assigned in a random order to each operator. The operation of the machines requires physical dexterity, and it is anticipated that there will be a difference among the operators in the speed with which they operate the machines (Table 5). The amount of time (seconds) were recorded for assembling the product:

Test the hypothesis H_0, at the 0.05 level significance, that the machines perform at the same mean rate of speed.

Table 4: Time, in Seconds, to Assemble Product

Operator

Machine	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	42.5	39.3	39.6	39.9	42.9	43.6	247.8
2	39.8	40.1	40.5	42.3	42.5	43.1	248.3

3	40.2	40.5	41.3	43.4	44.9	45.1	255.4
4	41.3	42.2	43.5	44.2	45.9	42.3	259.4
Total	163.8	162.1	164.9	169.8	176.2	174.1	1010.9

Solution: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ (machine effects are zero),

H_1 : At least one of the $\alpha_i's$ is not equal to zero

Table5. ANOVA Table for Table 4

Source of variation	**Sum of Squares**	**Degree of Freedom**	**Mean square**	**Computed f**
Machines	15.93	3	5.31	3.34
Operators	42.09	5	8.42	
Error	23.84	15	1.59	
Total	81.86	23		

using 5% as at least an approximate yardstick, we conclude that the machines do not perform at the same mean rate of speed.

Latin Squares

The randomized block design is very effective for reducing experimental error by removing one source of variation.Another designthat is particular useful in controlling two sources of variation, while reducing the required number of treatment combinations, is called the Latin square.Suppose that we are interested in the yields of 4 varieties of wheat using 4 different fertilizers over a period of 4 years. The total number of treatment combinations for a completely randomized design would be 64. By selecting the same number of categories for all three criteria of classification, we may select a Latin square design and perform the analysis of variance using the results of only 16 treatment combinations. A typical Latin square, selected at random from all possible 4×4 squares, is the following:

Column

Row	1	2	3	4
1		A	B	C	D
2		D	A	B	C
3		C	D	A	B
4		B	C	D	A

The four letters, A, B,C, and D, represent the 4 varieties of wheat that are referred to as the treatments. The rows and columns, represented by the 4 fertilizers and the years, respectively, are the two sources of variation that we wish to control. We now see that each treatment occurs exactly once in each row and each column. With such a balanced arrangement the analysis of variance enables one to separate the variation due to the different fertilizers and different years from the error sum of squares and thereby obtain a more accurate test for differences in the yielding capabilities of the 4 varieties of wheat. When there is interaction present between any of the sources of variation, the f-values in the analysis of variance are no longer valid. In that case, the Latin square design would be inappropriate.

Generalization to the Latin Square

We now generalized and consider an $r \times r$ Latin squarewhere y_{ijk} denotes an observation in the ith row and j^{th} column corresponding to kth letter. Note that once i and j are specified for a particular Latin square, we automatically know the letter given by k. For example, when $i = 2$ and $j = 3$ in the 4×4 Latin square, we have $k = B$. Hence k is a function of i and j. If α_i and β_j are the effects of the ith row and j^{th} column, τ_k the effect of the kth treatment, the μ the grand mean, and ε_{ijk} the random error, then we can write

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk},$$

where we impose the restrictions

$$\sum_i \alpha_i = \sum_j \beta_i = \sum_k \tau_k = 0.$$

As before, the y_{ijk} are assumed to be values of independent random variables having normal distributions with means

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \tau_k$$

And common variance σ^2. The hypothesis to be tested is as follows:

$$H_0 : \tau_1 = \tau_2 = ... = \tau_r = 0$$

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

The ANOVA (Table 6) indicates the appropriate F-test for treatments.

Table 4. ANOVA for an $r \times r$ Latin Square

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed** f
Rows	SSR	r-1	$s_1^2 = \frac{SSR}{r-1}$	
Columns	SSC	r-1	$s_2^2 = \frac{SSC}{r-1}$	
Treatments	SSTr	r-1	$s_3^2 = \frac{SSTr}{r-1}$	$f = \frac{s_3^2}{s^2}$
Error	SSE	(r-1)(r-2)	$s^2 = \frac{SSE}{(r-1)(r-2)}$	
Total	SST	r^2-1		

To illustrate the analysis of a Latin square design, let us return to the experiment where the letters A, B, C and D represent 4 represent 4 varieties of wheat; the rows represent 4 different fertilizers; and the columns account for 4 different years. The data in Table 5 are the yields for the 4 varieties of wheat, measured in kilograms per plot. It is assumed that the various sources variation do not interact. Using a 0.05 level of significance, test the hypothesis H_0: There is no difference in the average yields of the 4 varieties of wheat.

Table7. Yields of Wheat (kilograms per plot)

Fertilizer Treatment	**1981**	**1982**	**1983**	**1984**
t_1	A:70	B:75	C:68	D:81
t_2	D:66	A:59	B:55	C:63
t_3	C:59	D:66	A:39	B:42
t_4	B:41	C:57	D:39	D:55

Solution:

$$H_0 : \tau_1 = \tau_2 = ... = \tau_r = 0$$

H_1 : At least one of the $\tau_i's$ is not equal to zero.

Table 8. ANOVA for the Data of Table 7

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed** f
Fertilizer	1557	3	519.00	
Year	418	3	139.33	
Treatments	264	3	88.00	2.02
Error	261	6	43.50	
Total	2500	15		

We therefore, conclude that wheat varieties significantly affect wheat yield.