

Answers may vary due to rounding.

Goodness of fit test.

Computer instruction

Enter observed values in column 1 and expected values in column 2.

Click on Calc<Calculator.

Type c3 in the box Store Results.

Type the formula $(c1-c2)**2 / c2$ in the large box. Hit enter.

Click on Calc<column statistics

Click on sum.

Type c3 in the input variable box. The value printed is the χ^2 calculated.

The χ^2 table on MINITAB works the same way as the t - table. MINITAB always reads to the left.

1. An official of a plastics industry claimed that the industry employed 30% white women, 5% minority women, 50% white men, and 15% minority men. To test the claim, an affirmative action committee randomly sampled 150 employees and obtained the following information:

Category	observed
White females	40
Minority females	15
White males	80
Minority males	15

Test the official's claim at a 5% level of significance [$\chi^2 = 10.889$, $df=3$, p -value = 0.0123, Rho]

2. A computer science major claimed to have written a program that would randomly generate integers from 1 to 100. The program generated the following data.
[$\chi^2 = 6.40$, $df=9$, p -value = .6993, FtRho]

Integers	Observed
1-10	6
11-20	6
21-30	13
31-40	9
41-50	13
51-60	11
61-70	8
71-80	12
81-90	10
91-100	12

3. Given below are the frequencies observed from 310 tosses of a die. Do these data cast doubt on the fairness of the die at the 5% significance level? [$\chi^2 = 13.225$, $df = 5$, p -value = .0214, Rho]

Face No.	1	2	3	4	5	6
Frequency	38	61	54	65	55	37

4. A shipment of assorted nuts is labeled as having 45% walnuts, 20% hazelnuts, 20% almonds, and 15% pistachios. By randomly picking several scoops of nuts from this shipment, an inspector find the following counts.

	Walnuts	Hazelnuts	Almonds	Pistachios	Total
Counts	92	69	32	42	235

Could these findings be a strong basis for an accusation of mislabeling? Test at the 5% significance level. [$\chi^2 = 18.165$, $df = 3$, $p\text{-value} = 0.0004$, Rho]

Chi-Square Tests of Independence

Minitab will perform all necessary calculations for chi-square tests on contingency tables, presenting the expected values, the value of the test statistic, degrees of freedom and the p-value.

1. Enter the observed frequencies into rows and columns just as they are given in the contingency table.
 2. Go to the main header and click on **STAT>Tables>Chi-Square Test**
 3. In **Columns containing the tables**, enter the columns at which your contingency table is contained.
 4. Click on **OK**.
1. In a genetic study of chromosome structures, 132 individuals are classified according to the type of structural chromosome aberration and carriers in their parents. The following counts are obtained.

Type of Aberration	Carrier		Total
	One Parent	Neither Parent	
Presumably innocuous	28	19	47
Substantially unbalanced	35	50	85
Total	63	69	132

- (a) Test the null hypothesis that type of aberration is independent of parental carrier. [$\chi^2 = 4.106$, $df = 1$, $p\text{-value} = 0.043$, Rho]
 - (b) Perform the directional test
 - (c) Construct a 95% confidence interval for the difference in proportion for presumably innocuous for one parent and neither parent. [.0047, .3243]
2. A personnel administrator provided the following data as an example of hiring to fill 12 positions from among 40 male and 40 female applicants.

Applicant	Selected	Not Selected	Total
Male	7	33	40
Female	5	35	40

- (a) Does this sample indicate a selection bias in favour of males? [$\chi^2 = 0.392$, $df = 1$, $p\text{-value} = 0.531$, FtRho]
- (b) Construct a 95% confidence interval for the difference in proportions of selected males and females. [-.1115, .2067] Interpret the confidence interval.

3. Over the years pollsters have found that the public's confidence in big business has been closely tied to the economic climate of the country. When businesses are growing and employment is increasing public confidence is high. When the opposite occurs, public confidence is low. In one study, Harvey Kahalas (1981) explored the relationship between confidence in big business and job satisfaction. He hypothesized that there is a relationship between the level of confidence and job satisfaction and that this relationship holds true for both union and nonunion workers. To test his hypotheses he used the sample data given in the tables below:

Union Members

Job Satisfaction

Confidence in Major Corporations	Very Satisfied	Moderately Satisfied	Little dissatisfied	Very Dissatisfied
A great deal	30	19	6	6
Only some	99	77	20	9
Hardly any	38	32	14	15

NonUnion Members

Job Satisfaction

Confidence in Major Corporations	Very Satisfied	Moderately Satisfied	Little dissatisfied	Very Dissatisfied
A great deal	111	52	13	5
Only some	246	142	37	18
Hardly any	73	51	19	9

Perform a hypothesis test on each of these data sets at the 5% significance level. Does the data support Kahalas's theory? [$\chi^2 = 13.359$, $df = 6$, $p\text{-value} = 0.033$, Rho] [$\chi^2 = 8.298$, $df = 6$, $p\text{-value} = 0.2171$, FtRho]

4. Applicants for public assistance are allowed an appeals process when they feel unfairly treated. At such a hearing, the applicant may choose self-representation or representation by an attorney. The appeal may result in an increase, decrease, or no

change of the aid recommendation. Court records of 320 appeals cases provided the following data.

Amount of Aid

Type of Representation	Increased	Unchanged	Decreased
Self	59	108	17
Attorney	70	63	3

- (a) Are the patterns of the appeals decision significantly different between the two types of representation? [$\chi^2 = 15.734$, $df = 2$, $p\text{-value} < .0001$, Rho]
- (b) Construct a 95% confidence interval for the difference in proportion for increased aid for self representation and attorney representation. [.0848, .299] Interpret.

5. A survey was conducted by sampling 400 persons who were questioned regarding union membership and attitude toward decreased national spending on social welfare programs. The cross-tabulated frequency counts are presented.

	Support	Indifferent	Opposed
Union	112	36	28
NonUnion	84	68	72
Total	196	104	100

Can these observed differences be explained by chance or are there real differences of attitude between the populations of members and non-members at the 5% significance level? [$\chi^2 = 27.847$, $df = 2$, $p\text{-value} < .0001$, Rho]

6. In a Study of possible genetic influence of parental hand preference, a sample of 400 children was classified according to each child's handedness and the handedness of the biological parents. Do these findings demonstrate an association between the handedness of parents and their biological offspring at the 5% significance level? [$\chi^2 = 10.653$, $df = 2$, $p\text{-value} = 0.005$, Rho]

Handedness of Biological Offspring

Parents' Handedness Father x Mother	Right	Left	Total
Right x Right	303	37	340
Right x Left	29	10	39
Left x Right	16	6	22
Total	348	52	401

One-Way ANOVA

A firm developing a new citrus-flavoured soft drink conducted an experiment to study customer preferences for the colour of the drink. Four colours were considered: colourless, pink, orange, and lime green. Twenty test localities, which were similar in sales potential and representative of the target market

for this product, were selected. Each colour was randomly assigned to five localities for test marketing. The number of cases sold per 1000 population during the test period are recorded below:

Colourless	Pink	Orange	Lime Green
26.5	31.2	27.9	30.8
28.7	28.3	25.1	29.6
25.1	30.8	28.5	32.4
29.1	27.9	24.2	31.7
27.2	29.6	26.5	32.8

Are the sales the same for all colours of the drink? Test at the 5% significance level.

Minitab

Method 1

1. Enter the data into the MINITAB worksheet one sample per column- *ie.* The 5 values from the 'colourless' sample are entered in c1, the 5 values of the 'pink' sample in c2, and so forth.
2. From the MENU BAR, select STAT>ANOVA>ONEWAY(UNSTACKED) When the dialog box appears, list all the columns in which you entered data.
3. Note the table and other details are printed in the SESSION window.

Method 2

1. Enter all the sample values into one column of the worksheet. In a corresponding position in a second column, enter the number of the sample (1,2,3, etc.) from which each value came. For instance, if you entered all the 'colourless' observations, then pink, then orange, then lime green in column c1, c1 would contain 26.5, 28.7, 25.1, 29.1, 27.2, 31.2, 28.3, ... 32.8. You would enter into c2: 1, 1, 1, 1, 1, 2, 2, ..., 4. NOTE: These entries in c2 are most efficiently accomplished with CALC>MAKE PATTERNED DATA>SIMPLE SET OF NUMBERS. In the dialog box, store patterned data in c2. FROM 1st VALUE box type **1**, TO LAST VALUE box type **4**, LIST EACH VALUE box type **5**, LIST THE WHOLE SEQUENCE box type **1**.
 2. From the MENU BAR, select STAT>ANOVA>ONEWAY As the RESPONSE, specify the column in which you entered the observations, as the FACTOR specify the column in which you placed the sample number.
 3. The output appears in the SESSION window.
1. A large consulting firm hires a West Coast university to provide an M.B.A program for its employees. The basic statistics course is taught at four locations of the firm. After completion of the course, standardized tests are given to the participating employees at each location. The results are:

Location	A	B	C	D
Observation 1	96	65	66	60
2	88	74	90	72
3	92	77	88	66

4	75	82	73	75
5	81	70	85	78
6	62	78	87	68
7	86	84	94	
8	86		74	
9	90			
10	85			

Can it be concluded that there is no difference in test results between location? Use $\alpha = 0.05$

$$\{F_{\text{calc}} = 4.130 > 2.9680, \text{Rho}\}$$

2. The sales research division of a large corporation is conducting research on sales methods for selling one of the products of the corporation. The division has designed a completely randomized one-factor analysis of variance model to investigate the efficiency of three sales methods. The responses are measured in units of \$100 sales, and are listed below.

Response	Sales Method		
	A	B	C
1	20	13	31
2	23	18	28
3	22	16	48
4	22	29	28
5	36	14	30
6	45		25
7	21		
8	26		

Are there differences between the three different sales methods? Use $\alpha = 0.10$. What assumptions did you make? $\{F_{\text{calc}} = 3.899 > 2.668, \text{Rho}, \text{Sales are normally distributed and share a common variance}\}$

3. A county employs 3 assessors who are responsible for determining the values of residential property in the county. To see whether or not these assessors differ in their appraisal, 5 houses were selected and each assessor determined the market value of each house. The data (assessors are the treatment since your main concern is to see if there is difference in appraisals) was then analyzed using a two-way ANOVA routine giving the following (partially completed) ANOVA table.

(a) Complete the ANOVA table

Source	SS	DF	MS	F
Treatment	45.9			
Block	141.7			
<u>Error</u>				
Total	250.8			

- (b) Is there any indication of a difference between appraisors? Use a 5% significance level for testing.

$$\{F_{\text{calc}} = 2.905 < 4.459, \text{Fail to Rho}\}$$

Two-Way ANOVA

The data in the following table represent the milliequivalents of sodium excreted by six subjects 2 hours after treatment with one of four diuretics assigned at random by a clinician over a 6-day period. Using significance level of 0.05, analyze the data to determine whether or not there are any differences between patients and any differences in the effectiveness of diuretics.

Subjects	Treatments (Diuretics)			
	A	B	C	D
1	3.9	30.6	25.2	4.4
2	5.6	30.1	33.5	7.9
3	5.8	16.9	25.5	4.0
4	4.3	23.2	18.9	4.4
5	5.9	26.7	20.5	4.2
6	4.3	10.9	26.7	4.4

$$\{F_{\text{Bl}} = 1.5758 < 2.90, \text{Aho depending on rounding}\}$$

$$\{F_{\text{Tr}} = 37.153 > 3.287, \text{Rho}\}$$

Method

1. Enter all of the data into one column. You should first enter all the values for treatment A, then treatment B, and so forth.
2. In the second column, enter from which treatment each value came from (*ie.* You would enter A, A, A, A, A, A, B, B,..., D) These entries are most efficiently accomplished with `CALC>MAKE PATTERNED DATA > TEXT VALUES`
 Store patterned data in **C2**
 In the dialog box, type A B C D (Note: leave a space between each letter)
 In the list each value box, type **6**
 In the list the whole sequence box, type **1**.
3. In the third column, we want to enter the subject number (*ie.* 1,2,3,4,5,6,1,2,3,4,5,6,...)
 Go to `CALC>MAKE PATTERNED DATA>SIMPLE SET OF NUMBERS`
 - (a) Store patterned data in **C3**
 - (b) From 1st value, type **1**
 - (a) To last value, type **6**
 - (b) In steps of , type **1**
 - (c) List each value, type **1**
 - (d) List the whole sequence, type **4**

This will assign the numbers 1 through 6 four times.

4. From MENU BAR select `STAT>ANOVA>TWO-WAY`
 - (a) Response, type **C1**

- (b) Row factor , type **C2** (MINITAB takes row factor to represent treatment effect))
- (c) Column factor, type **C3** (MINITAB takes column factor to represent block effect)

Note: do not take “row factor” and “column factor” literally

5. Output appears in the SESSION window.
4. For the following data, present the ANOVA table. What conclusions can you draw from the two F tests? Use an $\alpha = 0.05$

	Block			
<u>Treatment</u>	1	2	3	4
1	14	8	10	4
2	7	7	4	2
3	12	6	16	6

$$\{F_{Bl} = 3.462 < 4.7571, \text{ Fail to RHo}\}$$

$$\{F_{Tr} = 3.2299 < 5.1433, \text{ Fail to RHo}\}$$

5. Three loaves of bread, each made according to a different recipe, are baked in one oven at the same time. Because of possible uncontrolled variations in oven performance, each baking is treated as a block. This procedure is repeated five times and the following measurements of density are obtained.

	Block				
<u>Recipe</u>	1	2	3	4	5
1	.95	.86	.71	.72	.74
2	.71	.85	.62	.72	.64
3	.69	.88	.51	.73	.44

- (a) How should the three oven positions of the three loaves be selected for each trial?
{Randomize the position of the loaves in the oven}
- (b) Perform an analysis of variance for these data using a 5% significance level
 $\{F_{Bl} = 5.31 > 3.837, \text{ Rho}\}$
 $\{F_{Tr} = 3.92 < 4.46, \text{ Fail to RHo}\}$
- (c) Find the p-value for the above Fcalculated value for Treatment. {.065}

F distribution and Simple Linear Regression

Note: there may be slight differences in the answers due to rounding.

1. Here is a set of data showing the historic yearly rates of return in seven randomly selected years, for Stock Y and the New York Stock Exchange Index (the predictor variable).

Year	1	2	3	4	5	6	7
Stock Y	2.0%	7.9%	-6.0%	-9.5%	13.5%	7.5%	1.2%
NYSE Index	4.9%	13.0%	-2.5%	-10.6%	11.0%	14.5%	4.3%

- (a) Write down the linear regression model expressing the yearly rate of return on Stock Y as a linear function of the yearly rate of return of the NYSE Index.

- (b) Estimate the intercept and slope term in the model. (Note: the slope term is referred to as Stock Y's "beta", or β . This is a measure that stock analysts use to evaluate the past performance of a stock. Stocks possessing β 's greater than 1 tend to have larger expected rates of return compared to stocks with smaller β 's. [$\beta_0 = -1.7470, \beta_1 = 0.8332$]

Minitab instructions

1. Enter Stock Y data in column C1
2. Enter NYSE Index data in column C2
3. Click on **STAT>Regression>Regression**
4. Enter C1 in the **Response** box
5. Enter C2 in the **Predictor** box
6. Click on **Graphs**, click on **residuals versus fits**, click **OK**
7. Click **OK**

You will now get a graph of the residuals vs fits for the data. There will also be a printout of the regression equation and the ANOVA table on the screen. If you want to see a scatter plot of the data with the fitted line,

Click on **STAT>Regression>Fitted Line Plot**

Enter C1 in **Response (Y)**

Enter C2 in **Predictor (X)**

Highlight **Linear** for **Type of Regression**.

Click **OK**.

- (c) Construct a 95% confidence interval estimate for Stock Y's β (β_1). Interpret the meaning of this interval. [$0.4560 \leq \beta_1 \leq 1.2105$]
- (d) Is the rate of return on Stock Y positively related to the rate of return on the NYSE Index? Test at a level of significance of 0.05. [$T_{calc} = 5.676, p\text{-value} = .002, RHo$]
- (e) Construct an analysis of variance table for the above regression. In addition, perform the same test in (d) using a different test. (again, $\alpha = 0.05$). Are the results in (c) and (d) consistent? [$F_{calc} = 32.23, p\text{-value} = .002, RHo, \text{yes}$]
- (f) Find the standard error of the regression and interpret its significance. [$s = 3.250$]
- (g) Find the coefficient of determination and interpret its meaning. [$r^2 = 0.8656$]
- (h) Find a 99% confidence interval estimate for this year's rate of return on Stock Y if the New York Stock Exchange Index has a rate of return of 8.1%. (or a 99% prediction interval). Interpret this interval. Would you invest in this stock, based on your interval? [$-9.1316, 19.1354$]
- (i) Find the coefficient of correlation between the rate of return on Stock Y and the rate of return on the NYSE [$r = +0.930$]
2. The Director of Management Information Systems at a conglomerate must prepare his long-range forecasts for the company's 3-year budget. In particular, he must develop staffing ratios to predict the number of managers and project leaders based on the number of programmers. The results of a sample of the electronic data processing staffs of 10 companies within the industry are displayed below.

# of applications Programmers	15	7	20	12	16	20	10	9	18	15
# of Managers and Project leaders	6	2	10	4	7	8	4	6	7	9

- Find the regression coefficients. State the least squares linear regression equation. [$\hat{y} = -0.0885 + 0.45x$]
- Interpret the meaning of the slope and intercept.
- Compute Se and interpret this value. [$s = 1.42$]
- Compute the coefficient of determination and interpret its meaning in this problem. [$r^2 = 0.7018$]
- At the 0.05 level of significance, is there a linear relationship between the number of managers and the number of application programmers? Use T-test [T=4.339, p-value = .0024, RHo]
- At the 0.05 level of significance, test for the appropriateness of the simple linear regression model. Use F-test [F=18.834 p-value = .0024, RHo]
- Set up a 95% confidence interval estimate of the true population slope. [0.2109 $\leq \beta_1 \leq$ 0.6891]
- Set up a 95% confidence interval estimate of the true population intercept. [-3.6374 $\leq \beta_0 \leq$ 3.4604]
- Set up a 95% confidence interval estimate of the average number of managers at companies where there are 10 programmers. [2.9711, 5.8559]
- Set up a 95% prediction interval estimate of the number of managers for a particular company in which there are 10 programmers. [0.8354, 8.1716]
- Construct a residual plot of the above data. What can you conclude from this residual plot? Does the linear model seem appropriate? Explain.

3. High salaries for presidents and high executives of charitable organizations have been in the news from time to time. Consider the information in the table below for the United Way in 10 major cities in Canada.

City	Salary of President	Money Raised (per capita)
Ottawa	\$161,396	\$17.35
Montreal	\$189,808	\$15.81
Toronto	\$201,490	\$16.74
Winnipeg	\$171,798	\$31.49
Halifax	\$108,364	\$15.51
St. John's	\$126,002	\$23.87
Regina	\$146,641	\$15.89
Saskatoon	\$155,192	\$9.32
Edmonton	\$169,999	\$29.84
Vancouver	\$143,025	\$24.19

- Find the least-squares regression equation that expresses the presidents' annual salary as a linear function of the amount of money raised (per capita). Interpret the meaning of the slope term in the context of the question. [$\hat{y} = 152657 + 235.71x$]
- This past year, the City of Lethbridge (with a population of approximately 70,000) raised a total of 1.9 million dollars. Estimate the salary of the president of the United Way Lethbridge Chapter. [$x=27.14, \hat{y} = \$159,024.95$]

- (c) Find the ANOVA table. What percentage of the variation in presidents' salary is explained by the fact that some raised more money per capita than others?
[$r^2 = 0.0035$, very small]
- (d) Is there a significant linear relationship between the president's salary and per capita money raised? Use the p-value approach and interpret the p-value in the context of the question. [t=0.1677, p-value = .871, Fail to RHo]
- (e) Does there appear to be a significant linear relationship between the amount of money raised per capita and the presidents' salary? Conduct this test using both the t-test and the F-test. What is your conclusion?
[F=0.0281, p-value = .871, Fail to RHo]
- (f) This past year the United Way in Calgary raised \$34.94 per capita. Construct a 99% confidence interval for the mean salary of the president of the United Way in Calgary. [\$107,862.93, \$213,922.71]
- (g) Construct a 95% prediction interval for the salary of the president of the United Way in Calgary. [\$74,227.12, \$247,588.52]
- (h) Estimate, with 90% level of reliability, the average (mean) salary of a United Way president who raised \$29.00 per capita. [\$130,189.84, \$188,795.56]
4. The following is a MINITAB output for a random sample of 8 employees at Tackey Toy Manufacturing Company. The company wanted to see if there was a relationship between aptitude test results and output (dozens of units produced).

The regression equation is
Output = 1.03 + 5.14 test results

Predictor	Coef	St.Dev	T	P
Constant		2.070		
Test result		0.2831		

S=1.695 R-sq= R-sq(adj)=97.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression					
Residual Error					
Total		968.00			

- (a) Fill in the above tables and find R-sq.
- (b) State the least squares regression equation. [$\hat{y} = 1.03 + 5.14x$]
- (c) What percentage of the variation in output is explained by the fact that some had higher test results on the aptitude tests than others? [$r^2 = 0.982$, very large]
- (d) Is there a significant linear relationship between output and aptitude test results? Use the p-value approach and interpret the p-value in the context of the question. [t=18.16, RHo, p-value~0]

Does there appear to be a significant linear relationship between the aptitude test results and output? Conduct this test using both the t-test and the F-test. What is your conclusion? [t=18.16, F=329.61, RHo]