Click Here

For this specific case, I like the suggestion in my other answer better. But this one addresses the question asked: How to do a test of association on a table of nominal counts with more than two dimensions. ~ There are a couple of relatively easy approaches to determine if there's an association on a contingency table with three dimensions. The CochranMantelHaenszel test can be used for 3-dimensions only. One dimension needs to be identified as the stratifying dimension, so it tests the independence of the variables of the m x n tables within k strata. Some versions of the test allow for only 2 x 2 tables within the strata. In your case it makes sense to test for the association of gender and numbofStars, and treat country as the stratifying variable. Another approach is to use log-linear models, which test for association among nominal variables. They can test for mutual independence,conditional independence, or partial independence. The references listed here may be helpful in understanding these options. See the links listed in the code for ideas for post-hoc testing and testing assumptions. ### These examples adopted from### !require(MASS)){install.packages("MASS")} I'll need to simulate some data for these examples. Note that in constructing the table, country is the stratifying variable, and is listed third in the xtabs function. set.seed(12345)gender = c(rep("Male", 36), rep("Female", 36))country = rep(c("Belgium", "Germany", "Netherlands"), 24)numbofStars = c(sample(c(0:4), 36, replace=TRUE), sample(c(1:5), 36, replace=T))Data = data.frame(country, gender, numbofStars)Data$numbofStars = factor(Data$numbofStars)levels(Data$numbofStars)Table = xtabs(~ gender + numbofStars + country, data=Data)ftable(Table) CochranMantelHaenszel test: mantelhaen.test(Table)### Cochran-Mantel-Haenszel test###### data: Table### Cochran-Mantel-Haenszel M^2 = 19.765, df = 5, p-value = 0.001383 Log-linear model: library(MASS)loglm(~ gender + numbofStars + country, Table)### Call:### loglm(formula = ~gender + numbofStars + country, data = Table)###### Statistics:### X^2 df P(> X^2)### Likelihood Ratio 48.69355 27 0.006417156### Pearson 40.27253 27 0.048305047The distribution of a categorical variable in a sample often needs to be compared with the distribution of a categorical variable in another sample. For example, over a period of 2 years a psychiatrist has classified by socioeconomic class the women aged 20-64 admitted to her unit suffering from self poisoning sample A. At the same time she has likewise classified the women of similar age admitted to a gastroenterological unit in the same hospital sample B. She has employed the Registrar Generals five socioeconomic classes, and generally classified the women by reference to their fathers or husbands occupation. The results are set out in table 8.1.The psychiatrist wants to investigate whether the distribution of the patients by social class differed in these two units. She therefore erects the null hypothesis that there is no difference between the two distributions. This is what is tested by the chi squared () test (pronounced with a hard ch as in sky). By default, all tests are two sided.It is important to emphasise here that tests may be carried out for this purpose only on the actual numbers of occurrences, not on percentages, proportions, means of observations, or other derived statistics. Note, we distinguish here the Greek () for the test and the distribution and the Roman (x) for the calculated statistic, which is what is obtained from the test.The test is carried out in the following steps:For each observed number (0) in the table find an expected number (E); this procedure is discussed below.To calculate the expected number for each cell of the table consider the null hypothesis, which in this case is that the numbers in each cell are proportionately the same in sample A as they are in sample B. We therefore construct a parallel table in which the proportions are exactly the same for both samples. This has been done in columns (2) and (3) of table 8.2 . The proportions are obtained from the totals column in table 8.1 and are applied to the totals row. For instance, in table 8.2 , column (2), 11.80 = (22/289) x 155; 24.67 = (46/289) x 155; in column (3) 10.20 = (22/289) x 134; 21.33 = (46/289) x 134 and so on.Thus by simple proportions from the totals we find an expected number to match each observed number. The sum of the expected numbers for each sample must equal the sum of the observed numbers for each sample, which is a useful check. We now subtract each expected number from its corresponding observed number.The results are given in columns (4) and (5) of table 8.2 . Here two points may be noted.The sum of these differences always equals zero in each column.Each difference for sample A is matched by the same figure, but with opposite sign, for sample B.Again these are useful checks.The figures in columns (4) and (5) are then each squared and divided by the corresponding expected numbers in columns (2) and (3). The results are given in columns (6) and (7). Finally these results, (O-E)/E are added. The sum of them is xA helpful technical procedure in calculating the expected numbers may be noted here. Most electronic calculators allow successive multiplication by a constant multiplier by a short cut of some kind. To calculate the expected numbers a constant multiplier for each sample is obtained by dividing the total of the sample by the grand total for both samples. In table 8.1 for sample A this is 155/289 = 0.5363. This fraction is then successively multiplied by 22, 46, 73, 91, and 57. For sample B the fraction is 134/289 = 0.4636. This too is successively multiplied by 22, 46, 73, 91, and 57.The results are shown in table 8.2 , columns (2) and (3).Having obtained a value for we look up in a table of distribution the probability attached to it (Appendix Table C.pdf). Just as with the t table, we must enter this table at a certain number of degrees of freedom. To ascertain these requires some care.When a comparison is made between one sample and another, as in table 8.1 , a simple rule is that the degrees of freedom equal (number of columns minus one) x (number of rows minus one) (not counting the row and column containing the totals). For the data in table 8.1 this gives (2 1) x (5 1) = 4. Another way of looking at this is to ask for the minimum number of figures that must be supplied in table 8.1 , in addition to all the totals, to allow us to complete the whole table. Four numbers disposed anyhow in samples A and B provided they are in separate rows will suffice.Entering Table C at four degrees of freedom and reading along the row we find that the value of x(7.147) lies between 3.357 and 7.779. The corresponding probability is: 0.10The above method of calculating x illustrates the nature of the statistic clearly and is often used in practice. A quicker method, similar to the quick method for calculating the standard deviation, is particularly suitable for use with electronic calculators.(1)The data are set out as in table 8.1 . Take the left hand column of figures (Sample A) and call each observation a. Their total, which is 155, is then .Let p = the proportion formed when each observation a is divided by the corresponding figure in the total column. Thus here p in turn equals 17/22, 25/46 32/57.Let = the proportion formed when the total of the observations in the left hand column, , is divided by the total of all the observations.Here = 155/289. Let = 1 , which is the same as 134/289.ThenWorking with the figures in table 8.1, we use this formula on an electronic calculator (Casio fx-350) in the following way:Withdraw result from memory on to display screenMR (1.7769764)We now have to divide this by Here = 155/289 and = 134/289.This gives us x= 7.146.The calculation naturally gives the same result if the figures for sample B are used instead of those for sample A. Owing to rounding off of the numbers the two methods for calculating x may lead to trivially different results.A special form of the test is particularly common in practice and quick to calculate. It is applicable when the results of an investigation can be set out in a fourfold table or 2 x 2 contingency table.For example, the practitioner whose data we displayed in believed that the wives of the printers and farmers should be encouraged to breast feed their babies. She has records for her practice going back over 10 years, in which she has noted whether the mother breast fed the baby for at least 3 months or not, and these records show whether the husband was a printer or a sheep farmer (or some other occupation less well represented in her practice). The figures from her records are set out in table 8.3The disparity seems considerable, for, although 28% of the printers wives breast fed their babies for three months or more, as many as 45% of the farmers wives did so. What is its significance? The null hypothesis is set up that there is no difference between printers wives and farmers wives in the period for which they breast fed their babies. The test on a fourfold table may be carried out by a formula that provides a short cut to the conclusion. If a, b, c, and d are the numbers in the cells of the fourfold table as shown in table 8.4 (in this case Variable 1 is breast feeding (< 3 months 0, 3 months 1) and Variable 2 is husbands occupation (Printer (0) or Farmer (1)), xis calculated from the following formula:With a fourfold table there is one degree of freedom in accordance with the rule given earlier.As many electronic calculators have a capacity limited to eight digits, it is advisable not to do all the multiplication or all the division in one series of operations, lest the number become too big for the display.Multiply a by d and store in memoryMultiply b by c and subtract from memoryEntering the table with one degree of freedom we read along the row and find that 3.418 lies between 2.706 and 3.84 1. Therefore 0.05We now calculate a confidence interval of the differences between the two proportions, as described in Chapter 6 In this case we use the standard error based on the observed data, not the null hypothesis. We could calculate the confidence interval on either the rows or the columns and it is important that we compare proportions of the outcome variable, that is, breast feeding.The 95% confidence interval is0.17 1.96 x 0.0924 to 0.17 + 1.96 x 0.0924 = -0.011 to 0.351Thus the 95% confidence interval is wide, and includes zero, as one might expect because the test was not significant at the 5% level.It can be shown mathematically that if is a Normally distributed variable, mean zero and variance 1, then x has a distribution with one degree of freedom. The converse also holds true and we can use this fact to improve the precision of our P values. In the above example we have = 3.418, with one degree of freedom. Thus X = 1.85, and from we find P to be about 0.065. However, we do need the tables for more than one degree of freedom.When the numbers in a 2 x 2 contingency table are small, the approximation becomes poor. The following recommendations may be regarded as a sound guide. (2) In fourfold tables a test is inappropriate if the total of the table is less than 20, or if the total lies between 20 and 40 and the smallest expected (not observed) value is less than 5; in contingency tables with more than one degree of freedom it is inappropriate if more than about one fifth of the cells have expected values less than 5 or any cell an expected value of less than 1. An alternative to the test for fourfold tables is known as Fishers Exact test and is described in Chapter 9When the values in a fourfold table are fairly small a correction for continuity known as the Yates correction may be applied (3). Although there is no precise rule defining the circumstances in which to use Yates correction, a common practice is to incorporate it into calculations on tables with a total of under 100 or with any cell containing a value less than 10. The test on a fourfold table is then modified as follows:The vertical bars on either side of ad bc mean that the smaller of those two products is taken from the larger. Half the total of the four values is then subtracted from that the difference to provide Yates correction. The effect of the correction is to reduce the value of x.Applying it to the figures in table 8.3 gives the following result:In this case x=2.711 falls within the same range of P values as the x= 3.418 we got without Yates correction (0.05Earlier in this chapter we compared two samples by the test to answer the question Are the distributions of the members of these two samples between five classes significantly different? Another way of putting this is to ask Are the relative proportions of the two samples the same in each class?For example, an industrial medical officer of a large factory wants to immunise the employees against influenza. Five vaccines of various types based on the current viruses are available, but nobody knows which is preferable. From the work force 1350 employees agree to be immunised with one of the vaccines in the first week of December, 50 the medical officer divides the total into five approximately equal groups. Disparities occur between their total numbers owing to the layout of the factory complex. In the first week of the following March he examines the records he has been keeping to see how many employees got influenza and how many did not. These records are classified by the type of vaccine used (table 8.5).In table 8.6 the figures are analysed by the test. For this we have to determine the expected values. The null hypothesis is that there is no difference between vaccines in their efficacy against influenza. We therefore assume that the proportion of employees contracting influenza is the same for each vaccine as it is for all combined. This proportion is derived from the total who got influenza, and is 225/1350. To find the expected number in each vaccine group who would contract the disease we multiply the actual numbers in the Total column of table 8.5 by this proportion. Thus 280 x (225/1350) = 46.7; 250 x (225/1350) = 41.7; and so on. Likewise the proportion who did not get influenza is 1125/1350.The expected numbers of those who would avoid the disease are calculated in the same way from the totals in table 8.5, so that 280 x (1125/1350) = 233.3; 250 x (1250/1350) = 208.3; and so on.The procedure is thus the same as shown in table 8.1 and table 8.2 .The calculations made in table 8.6 show that with four degrees of freedom is 16.564, and 0.001Inspection of table 8.6 shows that the largest contribution to the total x comes from the figures for vaccine III. They are 8.889 and 1.778, which together equal 10.667. If this figure is subtracted from the total x, 16.564 10.667 5.897. This gives an approximate figure for x for the remainder of the table with three degrees of freedom (by removing the vaccine III we have reduced the table to four rows and two columns). We then find that 0.1But this is not quite the end of the story. Before concluding from these figures that vaccine III is superior to the others we ought to carry out a check on other possible explanations for the disparity. The process of randomisation in the choice of the persons to receive each of the vaccines should have balanced out any differences between the groups, but some may have remained by chance. The sort of questions worth examining now are: Were the people receiving vaccine III as likely to be exposed to infection as those receiving the other vaccines? Could they have had a higher level of immunity from previous infection? Were they of comparable socioeconomic status? Of similar age on average? Were the sexes comparably distributed? Although some of these characteristics could have been more or less balanced by stratified randomisation, it is as well to check that they have in fact been equalised before attributing the numeral discrepancy in the result to the potency of the vaccine.Table 8.1 is a 5 x 2 table, because there are five socioeconomic classes and two samples. Socioeconomic groupings may be thought of as an example of an ordered categorical variable, as there are some outcomes (for example, mortality) in which it is sensible to state that (say) social class II is between social class I and social class III. The test described at that stage did not make use of this information; if we had interchanged any of the rows the value of x would have been exactly the same. Looking at the proportions p in table 8.1 we can see that there is no real ordering by social class in the proportions of self poisoning; social class V is between social classes I and II. However in many cases, when the outcome variable is an ordered categorical variable, a more powerful test can be devised which uses this information.Consider a randomised controlled trial of health promotion in general practice to change peoples eating habits.(5) Table 8.7 gives the results from a review at 2 years, to look at the change in the proportion eating poultry.If we give each category a score x the test for trend is calculated in the following way:ThusThis has one degree of freedom because the linear scoring means that when one expected value is given all the others are fixed, and we find p = 0.02. The usual test gives a value of = 5.51; d.f. = 2; 0.05Note that this is another way of splitting the overall x statistic. The overall x will always be greater than the for trend, but because the latter uses only one degree of freedom, it is often associated with a smaller probability. Although one is often counselled not to decide on a statistical test after having looked at the data, it is obviously sensible to look at the proportions to see if they are plausibly monotonic (go steadily up or down) with the ordered variable, especially if the overall test is nonsignificant.Comparison of an observed and a theoretical distributionIn the cases so far discussed the observed values in one sample have been compared with the observed values in another. But sometimes we want to compare the observed values in one sample with a theoretical distribution.For example, a geneticist has a breeding population of mice in his laboratory. Some are entirely white, some have a small patch of brown hairs on the skin, and others have a large patch. According to the genetic theory for the inheritance of these coloured patches of hair the population of mice should include 51.0% entirely white, 40.8% with a small brown patch, and 8.2% with a large brown patch. In fact, among the 784 mice in the laboratory 380 are entirely white, 330 have a small brown patch, and 74 have a large brown patch. Do the proportions differ from those expected?The data are set out in table 8.8 . The expected numbers are calculated by applying the theoretical proportions to the total, namely 0.510 x 784, 0.408 x 784, and 0.082 x 784. The degrees of freedom are calculated from the fact that the only constraint is that the total for the expected cases must equal the total for the observed cases, and so the degrees of freedom are the number of rows minus one. Thereafter the procedure is the same as in previous calculations of x. In this case it comes to 2.875. The x table is entered at two degrees of freedom. We find that 0.2McNemars test for paired nominal data was described in , using a Normal approximation. In view of the relationship between the Normal distribution and the distribution with one degree of freedom, we can recast the McNemar test as a variant of a test. The results are often expressed as in table 8.9. From appendix-table-c.pdf we find that for both values 0.02Extensions of the testIf the outcome variable in a study is nominal, the test can be extended to look at the effect of more than one input variable, for example to allow for confounding variables. This is most easily done using multiple logistic regression , a generalisation of multiple regression , which is described in Chapter 11. If the data are matched, then a further technique (conditional logistic regression) should be employed. This is described in advanced textbooks and will not be discussed further here.I have matched data, but the matching criteria were very weak. Should I use McNemars test?The general principle is that if the data are matched in any way, the analysis should take account of it. If the matching is weak then the matched analysis and the unmatched analysis should agree. In some cases when there are a large number of pairs with the same outcome, it would appear that the McNemars test is discarding a lot of information, and so is losing power. However, imagine we are trying to decide which of two high jumpers is the better. They each jump over a bar at a fixed height, and then the height is increased. It is only when one fails to jump a given height and the other succeeds that a winner can be announced. It does not matter how many jumps both have cleared.Snedecor GW, Cochran WG. In: Statistical Methods , 7th ed. Iowa: Iowa State University Press, 191,0:47.Cochran WG. Some methods for strengthening the common tests. Biometrics 1956; l0 :4l7-5l.Yates F. Contingency tables involving small numbers and the test. J Roy Stat Soc Suppl 1934; 1:217-3.Capples ME, McKnight A. Randomised controlled trial of health promotions in general practice for patients at high cardiovascular risk. BMJ l994;3O9:993-6.8.1 In a trial of new drug against a standard drug for the treatment of depression the new drug caused some improvement in 56% of 73 patients and the standard drug some improvement in 41% of 70 patients. The results were assessed in five categories as follows:What is the value of x which takes no account of the ordered value of data, what is the value of the x test for trend, and the P value? How many degrees of freedom are there? What is the value of P in each case?8.2 An outbreak of pediculosis capitis is being investigated in a girls school containing 291 pupils. Of 130 children who live in a nearby housing estate 18 were infested and of 161 who live elsewhere 37 were infested. What is the x value of the difference, and what is its significance? Find the difference in infestation rates and a 95% confidence interval for the difference.8.3 The 55 affected girls were divided at random into two groups of 29 and 26. The first group received a standard local application and the second group a new local application. The efficacy of each was measured by clearance of the infestation after one application. By this measure the standard application failed in ten cases and the new application in five. What is the value of the difference (with Yates correction), and what is its significance? What is the difference in clearance rates and an approximate 95% confidence interval?8.4 A general practitioner reviewed all patient notes in four practices for 1 year. Newly diagnosed cases of asthma were noted, and whether or not the case was referred to hospital. The following referrals were found (total cases in parentheses): practice A, 14 (103); practice B, 11 (92); practice C, 39 (166); practice D, 31 (221). What are the x and P values for the distribution of the referrals in these practices? Do they suggest that any one practice has significantly more referrals than others?For this specific case, I like the suggestion in my other answer better. But this one addresses the question asked: How to do a test of association on a table of nominal counts with more than two dimensions. ~ There are a couple of relatively easy approaches to determine if there's an association on a contingency table with three dimensions. The CochranMantelHaenszel test can be used for 3-dimensions only. One dimension needs to be identified as the stratifying dimension, so it tests the independence of the variables of the m x n tables within k strata. Some versions of the test allow for only 2 x 2 tables within the strata. In your case it makes sense to test for the association of gender and numbofStars, and treat country as the stratifying variable. Another approach is to use log-linear models, which test for association among nominal variables. They can test for mutual independence,conditional independence, or partial independence. The references listed here may be helpful in understanding these options. See the links listed in the code for ideas for post-hoc testing and testing assumptions. ### These examples adopted from### !require(MASS)){install.packages("MASS")} I'll need to simulate some data for these examples. Note that in constructing the table, country is the stratifying variable, and is listed third in the xtabs function. set.seed(12345)gender = c(rep("Male", 36), rep("Female", 36))country = rep(c("Belgium", "Germany", "Netherlands"), 24)numbofStars = c(sample(c(0:4), 36, replace=TRUE), sample(c(1:5), 36, replace=T))Data = data.frame(country, gender, numbofStars)Data$numbofStars = factor(Data$numbofStars)levels(Data$numbofStars)Table = xtabs(~ gender + numbofStars + country, data=Data)ftable(Table) CochranMantelHaenszel test: mantelhaen.test(Table)### Cochran-Mantel-Haenszel test###### data: Table### Cochran-Mantel-Haenszel M^2 = 19.765, df = 5, p-value = 0.001383 Log-linear model: library(MASS)loglm(~ gender + numbofStars + country, Table)### Call:### loglm(formula = ~gender + numbofStars + country, data = Table)###### Statistics:### X^2 df P(> X^2)### Likelihood Ratio 48.69355 27 0.006417156### Pearson 40.27253 27 0.048305047 Share copy and redistribute the material in any medium or format for any purpose, even commercially. Adapt remix, transform, and build upon the material for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the license terms. Attribution You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. ShareAlike If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. No additional restrictions You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits. You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation . No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material. I have just enrolled in a Data Science course on Udemy and I learned good stuff.In this article, we will do a Chi-square test with more than 2 categories. We will use the A/B test Country which has 3 categories which corresponds to 3 countries: German, Spain and France. Select Gender Actual tab, make a copy with a right-click and select Duplicate.Name the tab Gender Actual (2) by Country Actual. In Dimensions, move the variable Geography over Gender in Columns to replace Gender with Geography. Heres how to do an A/B statistical test when there are 3 categories. Well start with the classic method and then Ill show you another way to do Chi-square test with any number of categories. Lets start with the classical method. In this case, there are 3 categories so we cant use the online tool of the previous article. In the previous article we used an online tool with only 2 categories Sample1 and Sample2. Thats why were going to use another online tool, click here. In this online tool, we can enter the values without using the total values. That is, we enter only the number of observations in each category. We simply need to enter the values that are on our A/B test. And Im going to show you how to turn our A/B test into a table. In this way, it will be easier to enter the values in the online tool without making any mistakes. Go to the Show me tool at the top right. Click on text tablesClick on Swap Rows ans Columns button.Cool, now you have a table arranged in exactly the same way as the online tool.In the online tool, we will select 2 rows and 3 columns. As we have 3 categories and 2 possible results, we enter our values exactly as in the table we just created on Tableau. Perfect, our table is ready. You can click on the Calculate button. As you can see, we observe the same thing as the other online tool. There is our indicator p value which is less than 5%. Which means there is a meaning. This statistical significance means that these results are valid for the total number of the banks clients and not just for the sample of 10 000 clients. We observe similar differences with A/B test Country whose results are based solely on the sample of 10 000 clients. We can conclude that in the total number of the banks clients, its the clients in Germany who are more likely to leave the bank. This is how we do things cleanly. You saw, this online tool limited by 5 by 5 tables so you cant use this tool when you have 6 categories or more. But fortunately its possible to do Chi-square test with any number of categories. Its a special method and for you to understand that, Ill give you a theoretical explanation. Here we have 3 countries: German, Spain and France. What were trying to compare is the clients number leaving the bank in each of these countries. With our basic A/B test based on a sample of 10 000 clients, we obtained 16% for France, 32% for Germany and 17% for Spain. Now the question is: Do we observe the same results on the total clients number of the bank?, it means: In general, does the country have a significant effect on the clients number leaving bank?. Germany has the largest number of clients leaving the bank so the idea is: Why would we need to compare the 3 countries at the same time?.If we do an A/B test statistical test with Germany and France and we get a significant difference in the clients number leaving the bank between these 2 countries, then that would mean that in general, the country has a significant effect on the clients number who bank. Indeed, if we find by comparing Germany and France that the Germans are more likely to leave the bank than the French, we can consider that Spain will not change anything. Germans will always be more likely to leave the bank than the French. Maybe there will be a different relationship between Germany and Spain but there will always be a statistically significant difference between France and Germany with a larger number of clients leaving the bank in Germany than France. Here is a way to confirm that this logic is true. There is a test and the participants of this test are German, Spanish and French. Imagine that this test was done without looking at what is happening in Spain. Now you get the result and you ask yourself the question: Would the results changed if you added Spain?. The answer is no because there is no interdependence between Germany, Spain and France. That is, the decision to leave the bank in France and Germany doesnt depend on Spain. And therefore, its quite correct to separate the categories by putting 1 aside to compare the 2 others. And as now we have 2 categories, we can do a Chi-square test with the online tool that we used in the previous article. So lets go back to our worksheet and put a country aside to compare only 2 countries. Select Country tab. What we observe is that the difference between Spain and France is very small, so it wouldnt be interesting to do a Chi-square test between Spain and France. Its more interesting to do a Chi-square test between Germany and France and to prove that there is a statistically significant difference between these 2 countries. This will be enough to conclude that the country has a statistically significant impact on the clients number who leave the bank. Selects Country Actual tab. We will use the online tool of the previous article, click here.We will make a copy of Country Actual to have a bar chart with absolute values. Select Country Actual, right-click and select Duplicate. In Show Me, select horizontal bars.Removes SUM (Number of Records) from Columns and removes Exited and Geography from Rows. In Dimensions, move Geography in Columns.In Measures, move Number of Records to Rows.In Measures, move SUM(Number of Records) in Label.In Dimensions, move Exited in Label.In Dimensions, move Exited in Colors.We also need total absolute values, which means the total number of men and women. There is a very fast way to get that. Right-click on the vertical axis and select Add Reference Line. Then in Value, click on the drop-down on the right and select Sum to have the total sum of the observations. And in Scope, you select Per Cell option to specifythat you want the total sums for each category, male and female. Now, we have the total sum at the top of the bars. We will modify labels to have the absolute values. In Label, we will change Computation to Value and click on the OK button. Heres how to enter the data:For Sample1 in #success, you enter 810 because there are 810 people who left the bank. For Sample1 in #trials, you enter 5014 because there are 5014 people in total.For Sample2 in #success, you enter 814 because there are 814 people who left the bank. For Sample2 in #trials, you enter 2509 because there are 2509 people in total. Here is the verdict: Sample2 is more successful. Sample2 corresponds to Germans clients and #success is:yes, the client left the bank. This verdict means that of all the clients from German are more likely to leave the bank than clients from France. And look, there is something important, its p X^2)### Likelihood Ratio 48.69355 27 0.006417156### Pearson 40.27253 27 0.048305047 Online MPH and Teaching Public Health Modules. Looking for an affordable Online MPH program from top ranked Boston University without leaving home?Learn more about all of Boston Universitys degree programs, including the Online Master of Public Health that prepares working professionals like you to solve real-world public health challenges.Sign up for degree information: I acknowledge that by clicking the Submit button above, I am giving consent for representatives of Boston University to contact me about educational opportunities via email, text, or phone, including my mobile phone at the phone number above. I understand that these calls may be placed using an automatic dialer or prerecorded messages and I am not required to provide this consent in order to enroll. Message and data rates may apply. I may withdraw my consent at any time.Or are you an educator looking for resources and innovation in teaching public health?If you were a previous user of BUs online modules for public health education, BUSPH will be launching a new platform in 2025 called Teaching Public Health with these resources, and many more for educators and students.Notify me when TPH launches: Questions? Please contact spheducation@bu.edu The Chi-Square test is a statistical method used to determine if there is a significant association between categorical variables. When dealing with multiple columns, you can use Chi-Square tests to explore relationships across several categorical variables simultaneously. In this article, well go through the steps to conduct Chi-Square tests for multiple columns in R, including data preparation, executing the tests, and interpreting the results using R Programming Language.Understanding the Chi-Square TestThe Chi-Square test examines whether the distribution of categorical variables differs from what would be expected by chance. It is often used in two primary forms:Chi-Square Test of Independence: Tests whether two categorical variables are independent.Chi-Square Goodness of Fit Test: Tests whether the observed frequencies match expected frequencies.Now we will discuss step by step implementation of Chi-Square Tests for Multiple Columns in R.Step 1: Install and Load Required PackagesFor conducting Chi-Square tests, we can use base R functions. However, it's beneficial to have dplyr and ggplot2 for data manipulation and visualization. R install.packages("dplyr")install.packages("ggplot2") library(dplyr)library(ggplot2)Step 2: Create a Sample DatasetLets create a sample dataset that contains multiple categorical variables. R # Create a sample datasetset.seed(123)data