# Case C→C

**CO-4:** Distinguish among different measurement scales, choose the appropriate descriptive and inferential statistical methods based on these distinctions, and interpret the results.

Learning Objectives

**LO 4.35:** For a data analysis situation involving two variables, choose the appropriate inferential method for examining the relationship between the variables and justify the choice.

Learning Objectives

**LO 4.36:** For a data analysis situation involving two variables, carry out the appropriate inferential method for examining relationships between the variables and draw the correct conclusions in context.

**CO-5:** Determine preferred methodological alternatives to commonly used statistical methods when assumptions are not met.

**Review:** Unit 1 Case C-C

Video

**Video:** Case C→C (47:09)

**Related SAS Tutorials**

- 6A – (3:07) Two-Way (Contingency) Tables – EDA
- 6B – (9:41) Two-Way (Contingency) Tables – Inference

**Related SPSS Tutorials**

- 6A – (7:57) Two-Way (Contingency) Tables – EDA
- 6B – (9:19) Two-Way (Contingency) Tables – Inference

## Introduction

The last procedures we studied (two-sample t, paired t, ANOVA, and their non-parametric alternatives) all involve the relationship between a categorical explanatory variable and a quantitative response variable (case C→Q).In all of these procedures, the result is a comparison of the quantitative response variable (Y) among the groups defined by the categorical explanatory variable (X).The standard tests result in a comparison of the population means of Y within each group defined by X.

Next, we will consider inferences about the relationships between two categorical variables, corresponding to case C→C.

| | | Response | |
|---|---|---|---|
| | | Categorical | Quantitative |
| Explanatory | Categorical | C→C | ✓C→Q |
| | Quantitative | Q→C | Q→Q |

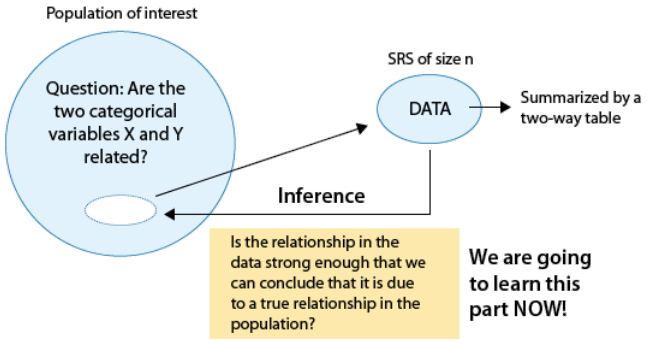
For case C→C, we will learn the following tests:

| Independent Samples (Only Emphasis) | Dependent Samples (Not Discussed) |
|---|---|
| **Standard Tests**<br>• **Continuity Corrected Chi-square Test for Independence (2×2 case)**<br>• **Chi-square Test for Independence (RxC case)**<br>Non-Parametric Test<br>• Fisher's exact test | *Standard Test*<br>• *McNemar's Test – 2×2 Case* |

In the Exploratory Data Analysis unit of the course, we summarized the relationship between two categorical variables for a given data set (using a two-way table and conditional percents), without trying to generalize beyond the sample data.

Now we will perform statistical inference for two categorical variables, using the sample data to draw conclusions about whether or not we have evidence that the variables are related in the larger population from which the sample was drawn.

In other words, we would like to **assess whether the relationship between X and Y that we observed in the data is due to a real relationship between X and Y in the population, or if it is something that could have happened just by chance due to sampling variability.**




Before moving into the statistical tests, let's look at a few (fake) examples.

## RxC Tables

Suppose our explanatory variable X has r levels and our response variable Y has c levels. We usually arrange our table with the explanatory variable in the rows and the response variable in the columns.

### EXAMPLE: RxC Table

Suppose we have the following partial (fake) data summarized in a two-way table using X = BMI category (r = 4 levels) and Y = Diabetes Status (c = 3 levels).

| | No Diabetes | Pre-Diabetes | Diabetes | Total |
|---|---|---|---|---|
| Underweight | | | | 100 |
| Normal | | | | 400 |
| Overweight | | | | 300 |
| Obese | | | | 200 |
| Total | 700 | 200 | 100 | 1000 |

From our study of probability we can determine:

- P(No Diabetes) = 700/1000 = 0.7
- P(Pre-Diabetes) = 200/1000 = 0.20
- P(Diabetes) = 100/1000 = 0.10

In the test we are going to use, our **null hypothesis** will be:

**Ho:** There is no relationship between X and Y.

Which in this case would be:

**Ho:** There is no relationship between BMI category (X) and diabetes status (Y).

If there were no relationship between X and Y, this would imply that the distribution of diabetes status is the same for each BMI category.

In this case (C→C), the distribution of diabetes status consists of the probability of each diabetes status group and the null hypothesis becomes:

**Ho:** BMI category (X) and diabetes status (Y) are **INDEPENDENT**.

Since the probability of "No Diabetes" is 0.7 in the entire dataset, if there were no differences in the distribution of diabetes status between BMI categories, we would obtain the same proportion in each row. Using the row totals we can find the **EXPECTED** counts as follows.

Notice the formula used below is simply the formula for the mean or expected value of a binomial random variable with n "trials" and probability of "success" p which was $\mu = E(X) = np$ where X = number of successes for a sample of size n.

| | No Diabetes | Pre-Diabetes | Diabetes | Total |
|---|---|---|---|---|
| Underweight | 100(0.7) = 70 | | | 100 |
| Normal | 400(0.7) = 280 | | | 400 |
| Overweight | 300(0.7) = 210 | | | 300 |
| Obese | 200(0.7) = 140 | | | 200 |
| Total | 700 | 200 | 100 | 1000 |

Notice that these do indeed add to 700.

Similarly we can determine the **EXPECTED** counts for the remaining two columns since 20% of our sample were classified as having pre-diabetes and 10% were classified as having diabetes.

| | No Diabetes | Pre-Diabetes | Diabetes | Total |
|---|---|---|---|---|
| Underweight | 70 | 100(0.2) = 20 | 100(0.1) = 10 | 100 |
| Normal | 280 | 400(0.2) = 80 | 400(0.1) = 40 | 400 |
| Overweight | 210 | 300(0.2) = 60 | 300(0.1) = 30 | 300 |
| Obese | 140 | 200(0.2) = 40 | 200(0.1) = 20 | 200 |
| Total | 700 | 200 | 100 | 1000 |

What we have created, using only the row totals, column totals, and column percents, is a table of what we would expect to happen if the null hypothesis of no relationship between X and Y were true. Here is the final result.

| | No Diabetes | Pre-Diabetes | Diabetes | Total |
|---|---|---|---|---|
| Underweight | **70** | **20** | **10** | 100 |
| Normal | **280** | **80** | **40** | 400 |
| Overweight | **210** | **60** | **30** | 300 |
| Obese | **140** | **40** | **20** | 200 |
| Total | 700 | 200 | 100 | 1000 |

Suppose we gather data and find the following (expected counts are in parentheses for easy comparison):

| | No Diabetes | Pre-Diabetes | Diabetes | Total |
|---|---|---|---|---|
| Underweight | **65** (*70*) | **22** (*20*) | **13** (*10*) | 100 |
| Normal | **285** (*280*) | **78** (*80*) | **37** (*40*) | 400 |
| Overweight | **216** (*210*) | **53** (*60*) | **31** (*30*) | 300 |
| Obese | **134** (*140*) | **47** (*40*) | **19** (*20*) | 200 |
| Total | 700 | 200 | 100 | 1000 |

If we compare our counts to the expected counts they are fairly close. This data would not give much evidence of a difference in the distribution of diabetes status among the levels of BMI categories. In other words, this data would not give much evidence of a relationship (or association) between BMI categories and diabetes status.

The standard test we will learn in case C→C is based upon comparing the **OBSERVED** cell counts (our data) to the **EXPECTED** cell counts (using the method discussed above).

We want you to see how the expected cell counts are created so that you will understand what kind of evidence is being used to reject the null hypothesis in case C→C.

Suppose instead that we gather data and we obtain the following counts (expected counts are in parentheses and row percentages are provided):

| | No Diabetes | Pre-Diabetes | Diabetes | Total |
|---|---|---|---|---|
| Underweight | **90** (*70*)<br>90% | **7** (*20*)<br>7% | **3** (*10*)<br>3% | 100 |
| Normal | **340** (*280*)<br>85% | **40** (*80*)<br>10% | **20** (*40*)<br>5% | 400 |
| Overweight | **180** (*210*)<br>60% | **90** (*60*)<br>30% | **30** (*30*)<br>10% | 300 |
| Obese | **90** (*140*)<br>45% | **63** (*40*)<br>31.5% | **47** (*20*)<br>23.5% | 200 |
| Total | 700 | 200 | 100 | 1000 |

In this case, most of the differences are drastic and there seems to be clear evidence that the distribution of diabetes status is not the same among the four BMI categories.

Although this data is entirely fabricated, it illustrates the kind of evidence we need to reject the null hypothesis in case C→C.

## 2×2 Tables

One special case occurs when we have two categorical variables where both of these variables have two levels. Two-level categorical variables are often called **binary** variables or **dichotomous** variables and when possible are usually coded as 1 for "Yes" or "Success" and 0 for "No" or "Failure."

Here is another (fake) example.

### EXAMPLE: 2x2 Table

Suppose we have the following partial (fake) data summarized in a two-way table using X = treatment and Y = significant improvement in symptoms.

| | No Improvement | Improvement | Total |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Control | | | 100 |
| Treatment | | | 100 |
| Total | 120 | 80 | 200 |

From our study of probability we can determine:

- P(No Improvement) = 120/200 = 0.6
- P(Improvement) = 80/200 = 0.4

Since the probability of "No Improvement" is 0.6 in the entire dataset and the probability for "Improvement" is 0.4, if there was no difference we would obtain the same proportion in each row. Using the row totals we can find the EXPECTED counts as follows.

| | No Improvement | Improvement | Total |
|---|---|---|---|
| Control | **100(0.6) = 60** | **100(0.4) = 40** | 100 |
| Treatment | **100(0.6) = 60** | **100(0.4) = 40** | 100 |
| Total | 120 | 80 | 200 |

Suppose we obtain the following data:

| | No Improvement | Improvement | Total |
|---|---|---|---|
| Control | **80** | **20** | 100 |
| Treatment | **40** | **60** | 100 |
| Total | 120 | 80 | 200 |

In this example we are interested in the probability of improvement and the above data seem to indicate the treatment provides a greater chance for improvement than the control.

We use this example to mention two ways of comparing probability (sometimes "risk") in 2×2 tables. Many of you may remember these topics from Epidemiology or may see these topics again in Epidemiology courses in the future!

**Risk Difference:**

For this data, a larger proportion of subjects in the treatment group showed improvement compared to the control group. In fact, the estimated probability of improvement is 0.4 higher for the treatment group than the control group.

This value (0.4) is called a **risk-difference** and is one common measure in 2×2 tables. Estimates and confidence intervals can be obtained.

For a fixed sample size, the larger this difference, the more evidence against our null hypothesis (no relationship between X and Y).

The population risk-difference is often denoted $p_1 - p_2$, and is the difference between two population proportions. We estimate these proportions in the same manner as Unit 1, once for each sample.

For the current example, we obtain

$$\hat{p}_1 = \hat{p}_{\text{TRT}} = \frac{60}{100} = 0.60$$

and

$$\hat{p}_2 = \hat{p}_{\text{Control}} = \frac{20}{100} = 0.20$$

from which we find the risk difference

$$\hat{p}_{\text{TRT}} - \hat{p}_{\text{Control}} = 0.60 - 0.20 = 0.40$$

**Odds Ratio:**

Another common measure in 2×2 tables is the odds ratio, which is defined as the odds of the event occurring in one group divided by the odds of the event occurring in another group.

In this case, the odds of improvement in the treatment group is

$$\text{ODDS}_{\text{TRT}} = \frac{P(\text{ Improvement } \mid \text{TRT})}{P(\text{ No Improvement } \mid \text{TRT})} = \frac{0.6}{0.4} = 1.5$$

and the odds of improvement in the control group is

$$\text{ODDS}_{\text{Control}} = \frac{P(\text{ Improvement } \mid \text{ Control })}{P(\text{ No Improvement } \mid \text{ Control })} = \frac{0.2}{0.8} = 0.25$$

so the odds ratio to compare the treatment group to the control group is

$$\text{Odds Ratio } = \frac{\text{ODDS }_{\text{TRT}}}{\text{ODDS}_{\text{Control}}} = \frac{1.5}{0.25} = 6$$

**This value means that the odds of improvement are 6 times higher in the treatment group than in the control group.**

Properties of Odds Ratios:

- The odds ratio is always larger than 0.
- An odds ratio of 1 implies the odds are equal in the two groups.
- Values much larger than 1 indicate the event is more likely in the treatment group (numerator group) than the control group (denominator group). This would give evidence that our null hypothesis is false.
- Values much smaller than 1 (closer to zero) would indicate the event is much less likely in the treatment group than the control group. This would also give evidence that our null hypothesis is false.
    - **Notice:** if we compared control to treatment (instead of treatment to control) we would obtain an odds ratio of 1/6 which would say that the odds of improvement in the control group is 1/6 the odds of improvement in the treatment group which leads us to exactly the same conclusion, worded in an opposite manner.

## Chi-square Test for Independence

### Learning Objectives

**LO 4.43:** In a given context, determine the appropriate standard method for examining the relationship between two categorical variables. Given the appropriate software output choose the correct p-value and provide the correct conclusions in context.

### Learning Objectives

**LO 4.44:** In a given context, set up the appropriate null and alternative hypotheses for examining the relationship between two categorical variables.

**Step 1: State the hypotheses**The hypotheses are:

**Ho**: There is no relationship between the two categorical variables. (They are independent.)

**Ha**: There is a relationship between the two categorical variables. (They are not independent.)

Note: for 2×2 tables, these hypotheses can be formulated the same as for population means except using population proportions. This can be done for RxC tables as well but is not common as it requires more notation to compare multiple group proportions.

- **Ho:** $p_1 - p_2 = 0$ (which is the same as $p_1 = p_2$)
- **Ha:** $p_1 - p_2 \neq 0$ (which is the same as $p_1 \neq p_2$) **(two-sided)**

**Step 2: Obtain data, check conditions, and summarize data**

(i) The sample should be random with independent observations (all observations are independent of all other observations).

(ii) In general, the larger the sample, the more precise and reliable the test results are. There are different versions of what the conditions are that will ensure reliable use of the test, all of which involve the expected counts. One version of the conditions says that all expected counts need to be greater than 1, and at least 80% of expected counts need to be greater than 5. A more conservative version requires that all expected counts are larger than 5. Some software packages will provide a warning if the sample size is "too small."

**Test Statistic of the Chi-square Test for Independence**:

The single number that summarizes the overall difference between observed and expected counts is the chi-square statistic, which tells us in a standardized way how far what we observed (data) is from what would be expected if Ho were true.

Here it is:

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{ observed count } - \text{ expected count })^2}{\text{expected count}}$$

**Step 3: Find the p-value of the test by using the test statistic as follows**We will rely on software to obtain this value for us. We can also request the expected counts using software.

The p-values are calculated using a chi-square distribution with (r-1)(c-1) degrees of freedom (where r = number of levels of the row variable and c = number of levels of the column variable). We will rely on software to obtain the p-value for this test.

**IMPORTANT NOTE**

- **Use Continuity Correction for 2×2 Tables:** For 2×2 tables, a continuity correction is used to improve the approximation of the p-value. This value will only be calculated by the software for 2×2 tables where both variables are binary – have only two levels.

**Step 4: Conclusion**

As usual, we use the magnitude of the p-value to draw our conclusions. A small p-value indicates that the evidence provided by the data is strong enough to reject Ho and conclude (beyond a reasonable doubt) that the two variables are related. In particular, if a significance level of 0.05 is used, we will reject Ho if the p-value is less than 0.05.

## Non-Parametric Alternative: Fisher's Exact Test

**Learning Objectives**

**LO 5.1:** For a data analysis situation involving two variables, determine the appropriate alternative (non-parametric) method when assumptions of our standard methods are not met.

We will look at one non-parametric test in case C→C. Fisher's exact test is an exact method of obtaining a p-value for the hypotheses tested in a standard chi-square test for independence. This test is often used when the sample size requirement of the chi-square test is not satisfied and can be used for 2×2 and RxC tables.

**Step 1: State the hypotheses**The hypotheses are:

**Ho**: There is no relationship between the two categorical variables. (They are independent.)

**Ha**: There is a relationship between the two categorical variables. (They are not independent, they are dependent.)

**Step 2: Obtain data, check conditions, and summarize data**

The sample should be random with independent observations (all observations are independent of all other observations).

**Step 3: Find the p-value of the test by using the test statistic as follows**

The p-values are calculated using a distribution specific to this test. We will rely on software to obtain the p-value for this test. The p-value measures the chance of obtaining a table as or more extreme (against the null hypothesis) than our table.

**Step 4: Conclusion**

As usual, we use the magnitude of the p-value to draw our conclusions. A small p-value indicates that the evidence provided by the data is strong enough to reject Ho and conclude (beyond a reasonable doubt) that the two variables are related. In particular, if a significance level of 0.05 is used, we will reject Ho if the p-value is less than 0.05.

Now let's look at a some examples with real data.

## EXAMPLE: Risk Factor for Low Birth Weight

Low birth weight is an outcome of concern due to the fact that infant mortality rates and birth defect rates are very high for babies with low birth weight. A woman's behavior during pregnancy (including diet, smoking habits, and obtaining prenatal care) can greatly alter her chances of carrying the baby to term and, consequently, of delivering a baby of normal birth weight.

In this example, we will use a 1986 study (Hosmer and Lemeshow (2000), Applied Logistic Regression: Second Edition) in which data were collected from 189 women (of whom 59 had low birth weight infants) at the Baystate Medical Center in Springfield, MA. The goal of the study was to identify risk factors associated with giving birth to a low birth weight baby.

Data: SPSS format, SAS format, Excel format

**Response Variable:**

- LOW – Low birth weight
  - 0=No (birth weight >= 2500 g)
  - 1=Yes (birth weight < 2500 g)

**Possible Explanatory Variables** (variables we will use in this example are in bold):

- **RACE** – Race of mother (1=White, 2=Black, 3=Other)
- **SMOKE** – Smoking status during pregnancy (0=No, 1=Yes)
- **PTL** – History of premature labor (0=None, 1=One, etc.)
- **HT** – History of hypertension (0=No, 1=Yes)
- **UI** – Presence of uterine irritability (0=No, 1=Yes)
- FTV – Number of physician visits during the first trimester
- BWT – The actual birth weight (in grams)
- AGE – Age of mother (in years)
- LWT – Weight of mother at the last menstrual period (in pounds)

**Results:**

**Step 1: State the hypotheses**

The hypotheses are:

**Ho**: There is no relationship between the categorical explanatory variable and presence of low birth weight. (They are independent.)

**Ha**: There is a relationship between the categorical explanatory variable and presence of low birth weight.(They are not independent, they are dependent.)

**Steps 2 & 3: Obtain data, check conditions, summarize data**, and **find the p-value**

| Explanatory Variable | Which Test is Appropriate? | P-value | Decision |
|---|---|---|---|
| RACE | Min. Expected Count = **8.12**<br>**3×2 table**<br>Use Pearson Chi-square (since RxC) | **0.0819** (Chi-square – SAS)<br>**0.082** (Chi-square – SPSS) | Fail to Reject Ho |

| | | | |
|---|---|---|---|
| SMOKE | Min. Expected Count = **23.1**<br>**2×2 table**<br>Use Continuity Correction (since 2×2) | **0.040** (Continuity Correction – SPSS)<br>**0.0396** (Continuity Adj – SAS) | Reject Ho |
| PTL | Min. Expected Count = **0.31**<br>**4×2 table**<br>Fisher's Exact test is more appropriate | **3.106 E-04 = 0.0003106** (Fisher's – SAS)<br>**0.000** (Fisher's – SPSS)<br>0.0008 (Chi-square – SAS)<br>0.001 (Chi-square – SPSS) | Reject Ho |
| HT | Min. Expected Count = **3.75**<br>**2×2 table**<br>Fisher's Exact test may be more appropriate | **0.0516** (Fisher's – SAS)<br>**0.052** (Fisher's – SPSS) | Fail to Reject Ho (Barely) |
| UI | Min. Expected Count = **8.74**<br>**2×2 table**<br>Use Continuity Correction | **0.0355** (Continuity Adj. – SAS)<br>**0.035** (Continuity Correction – SPSS) | Reject Ho |

**Step 4: Conclusion**

When considered individually, presence of uterine irritability, history of premature labor, and smoking during pregnancy are all significantly associated (p-value $< 0.05$) with the presence/absence of a low birth weight infant whereas history of hypertension and race were only marginally significant ($0.05 \leq$ p-value $< 0.10$).

**Practical Significance:**

| Explanatory Variable | Comparison of Conditional Percentages of Low Birth Weight |
|---|---|
| RACE | Race = White: 23.96%<br>Race = Black: 42.31%<br>Race = Other: 37.31% |
| SMOKE | Smoke = No: 25.22%<br>Smoke = Yes: 40.54% |
| PTL | History of Premature Labor = 0: 25.79%<br>History of Premature Labor = 1: 66.67%<br>History of Premature Labor = 2: 40.00% (Note small sample size of 5 for this row)<br>History of Premature Labor = 3: 0.00% (Note small sample size of 1 for this row) |
| HT | Hypertension = No: 29.38%<br>Hypertension = Yes: 58.33% (Note small sample size of 12 for this row) |
| UI | Presence of uterine irritability = No: 27.95%<br>Presence of uterine irritability = Yes: 50.00% |

- Despite our failing to reject the null in two of the five tests, all of these results seem to have some practical significance although the small sample sizes for some portions of the results may be producing misleading information and likely would require further study to confirm the results seen here.

SPSS Output for tests

SAS Output, SAS Code

## EXAMPLE: 2x2 Table - Revisting "Looks vs. Personality" with Binary Categorized Response

If, instead of simply analyzing the “looks vs. personality” rating scale, we categorized the responses into groups then we would be in case C→C instead of case C→Q (see previous example in Case C-Q for Two Independent Samples).

Recall the rating score was from 1 to 25 with 1 = personality most important (looks not important at all) and 25 = looks most important (personality not important at all). A score of 13 would be equally important and scores around 13 should indicate looks and personality are nearly equal in importance.

**For our purposes we will use a rating of 16 or larger to indicate that looks were indeed more important than personality (by enough to matter).**

Data: SPSS format, SAS format

**Response Variable:**

- Looks – “Looks were (much) more important?”
  - 0=No (Less than 16 on the looks vs. personality rating)
  - 1=Yes (16 or higher on the looks vs. personality rating)

**Results:**

**Step 1: State the hypotheses**

The hypotheses are:

> **Ho:** The proportion of college students who find looks more important than personality **is the same** for males and females. (The two variables are independent)
>
> **Ha:** The proportion of college students who find looks more important than personality **is different** for males and females. (The two variables are dependent)

**Steps 2 & 3: Obtain data, check conditions, summarize data**, and f**ind the p-value**

The minimum expected cell count is 13.38. This is a 2×2 table so we will use the continuity corrected chi-square statistic.

The p-value is found to be 0.001 (SPSS) or 0.0007 (SAS).

**Step 4: Conclusion**

There is a significant association between gender and whether or not the individual rated looks more important than personality.

Among males, 27.1% rated looks higher than personality while among females this value was only 9.3%.

For fun: The odds ratio here is

$$\text{Odds Ratio} = \frac{0.271/(1-0.271)}{0.093/(1-0.093)} = \frac{0.37174}{0.10254} = 3.63$$

which means, based upon our data, we estimate that the odds of rating looks more important than personality is 3.6 times higher among males than among females.

**Practical Significance:**

It seems clear that the difference between 27.1% and 9.3% is practically significant as well as statistically significant. This difference is large and likely represents a meaningful difference in the views of males and females regarding the importance of looks compared to personality.

SPSS Output

SAS Output, SAS Code