

Are people who are prone to sudden anger more likely to develop heart disease?

An observational study followed a random sample of 8474 people with normal blood pressure for about four years.

Degrees of Freedom - A different perspective.

How many expected values should we calculate with the formula

$$\frac{\text{row total} \times \text{column total}}{\text{Table Total}}$$

before we can simply subtract to get the rest?

	Mich.	Calif	Total
Country	14.2		16
Pop	37.6		29
Rap	57.1		43
Rock	14.3		17
Other			95
Total	100	100	200

Yesterday we did not address
that the calculator will not provide
you with contributions for a
follow up analysis

(like it did with χ^2 -G.O.F.)

- A spreadsheet could...••

Observed

Expected

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\frac{(L_1 - L_2)^2}{L_2}$$

L1	L2	3
84	58	---
51	69	---
69	69	---
L3 =		

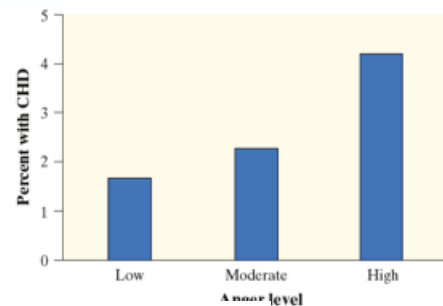
Are people who are prone to sudden anger more likely to develop heart disease?

An observational study followed a random sample of 8474 people with normal blood pressure for about four years.

Would the Chi-Square test for ~~Homogeneity~~ be appropriate to analyze this ?

Each person took the Spielberger Trait Anger Scale test. Researchers also recorded whether each individual developed coronary heart disease (CHD).

		Anger level			Total
		Low	Moderate	High	
CHD status	Yes	53	110	27	190
	No	3057	4621	606	8284
	Total	3110	4731	633	8474



Do these data provide convincing evidence of an association between the variables in the larger population?

needed: Chi-Square Test of Indep.

Tests for Independence: Stating Hypotheses

H_0 : There is **no association** between anger level and heart-disease status in the population of people with normal blood pressure.

H_a : There is **an association** between anger level and heart-disease status in the population of people with normal blood pressure.

An equivalent way to state the hypotheses is

H_0 : Anger and heart-disease status are **independent** in the population of people with normal blood pressure.

H_a : Anger and heart-disease status are **not independent** in the population of people with normal blood pressure.

CHD / Anger level ← single random sample
from a population with
normal blood pressure.

If instead they had taken 3 independent random
samples ●●●

← low level anger
med level anger
high level anger

We could then use a Chi-Square test
for Homogeneity.

Tests for Independence:

Conditions and Calculations

Conditions for Performing a Chi-Square Test for Independence

- **Random:** The data come from a random sample from the population of interest.
- **10%:** When sampling without replacement, $n < 0.10N$.
- **Large Counts:** All expected counts are at least 5.

← only one to check

Chi-Square Test for Independence

Suppose the conditions are met. To perform a test of

H_0 : There is no association between two categorical variables in the population of interest

compute the chi-square test statistic:

$$\chi^2 = \sum \frac{(\text{Observed count} - \text{Expected count})^2}{\text{Expected count}}$$

where the sum is over all cells (not including totals) in the two-way table.

The P -value is the area to the right of χ^2 under the chi-square density curve with degrees of freedom = (num. of rows - 1)(num. of columns - 1).

*We don't use
the words
"distribution".*

*the only difference
from Chi-Sq. test
for Homogeneity*

Chi-Square Test for Independence

Suppose the conditions are met. To perform a test of

H_0 : There is no association between two categorical variables in the population of interest

compute the chi-square test statistic:

$$\chi^2 = \sum \frac{(\text{Observed count} - \text{Expected count})^2}{\text{Expected count}}$$

where the sum is over all cells (not including totals) in the two-way table.

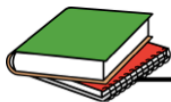
The P -value is the area to the right of χ^2 under the chi-square density curve with degrees of freedom = (num. of rows - 1)(num. of columns - 1).

TODAY →

Is there an association between gender and preference of English or Math?

Since our class is on the small side, data from a random sample of 114 from a certain high school in Michigan is provided.

Are gender and favorite class independent? Lesson 11.2 Day 2:



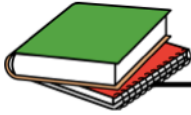
English

VS



Is there an association between gender and preference of English or math class? Below is the data for a random sample of 114 senior students. Do we have convincing evidence that gender and favorite class are associated?

Are gender and favorite class independent? Lesson 11.2 Day 2:



English

VS



Is there an association between gender and preference of English or math class? Below is the data for a random sample of 114 senior students. Do we have convincing evidence that gender and favorite class are associated?

1. Describe what it means for two events to be independent. (Chapter 5)

Is there an association between gender and preference of English or math class? Below is the data for a random sample of 114 senior students. Do we have convincing evidence that gender and favorite class are associated?

1. Describe what it means for two events to be independent. (Chapter 5)

Knowing if an event occurs does not change the probability of another event occurring.

Keep going...

2. Calculate the expected counts.

Observed:

	English	Math	Total
Female	43	22	65
Male	21	28	49
Total	64	50	114

Expected:

	English	Math	Total
Female			65
Male			49
Total	64	50	114

$\frac{65 \cdot 64}{114}$

3. Do the data provide significant evidence that there is an association between gender and preference of English or math class? Use $\alpha = 0.05$

STATE: Hypotheses:

Significance level:

2. Calculate the expected counts.

Observed:

	English	Math	Total
Female	43	22	65
Male	21	28	49
Total	64	50	114

Expected:

	English	Math	Total
Female	30.45	28.51	65
Male	27.51	21.49	49
Total	64	50	114

3. Do the data provide significant evidence that there is an association between gender and preference of English or math class? Use $\alpha = 0.05$

STATE: Hypotheses:

Significance level: $\alpha = 0.05$

H_0 : There is no association betw gender & favorite class.

H_a : There is an association betw. gender & fav. class

PLAN: Name of procedure: chi-square test for independence

Check the Large Counts Condition (assume the others are met)

All expected values ≥ 5 (See Table)

NOTE: You must always specify the complete name of the test

PLAN

Chi-Square Test !!!

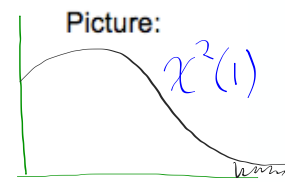
This person lost points :(

DO: Specific Formula: $\chi^2 = \sum \frac{(O-E)^2}{E}$

Work:

$$= \frac{(-)^2}{\dots} + \frac{(-)^2}{\dots} + \dots$$

$$df = (2-1)(2-1) = 1$$



Test statistic:

6.16

P-value:

= .013

DO: Specific Formula: $\chi^2 =$

Work:

$$= \frac{(43 - 36.45)^2}{36.45} + \frac{(22 - 28.51)^2}{28.51} + \dots$$

$$= 6.16$$

$$df = (2-1)(2-1) = 1$$



Test statistic:

6.16

P-value:

= .013

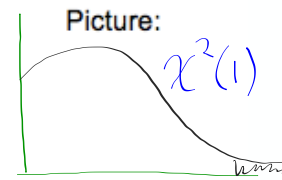
DO: Specific Formula: $\chi^2 =$

Work:

$$= \frac{(43 - 36.45)^2}{36.45} + \frac{(22 - 28.51)^2}{28.51} + \dots$$

$$= 6.16$$

$$df = (2-1)(2-1) = 1$$



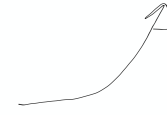
Test statistic:

6.16

P-value:

= .013

$$\chi^2_{CDF} [\underset{\text{Lower}}{6.16}, \underset{\text{Upper}}{10000}, \underset{\text{df}}{1}] =$$



AP® Exam Tip

When the P -value is very small, the calculator will report it using scientific notation. Remember that P -values are probabilities and must be between 0 and 1.

If your calculator reports the P -value with a number that appears to be greater than 1, look to the right, and you will see that the P -value is being expressed in scientific notation. If you claim that the P -value is 4.82, you will certainly lose credit.

< 000048

4.8232E-5

CONCLUDE:

Because the P-value = $0.013 < \alpha = 0.05$,
we reject H_0 .

\therefore We have convincing evidence of
an association between gender and
favorite class in the high school.

Chi-Square Test for Independence**Important ideas:**

H_0 : There is not an
association
between ___ and ___

H_a : There is an association
between ___ and ___

Which χ^2 ?

χ^2 GOF

1 sample, 1 variable

χ^2 for Homogeneity

2 samples, 1 variable

χ^2 for Independence

1 sample, 2 var

Using Chi-Square Tests Wisely

The **chi-square test for homogeneity** and the **chi-square test for independence** are very similar.

Using Chi-Square Tests Wisely

The **chi-square test for homogeneity** and the **chi-square test for independence** are very similar.

The best way to distinguish these two tests is to consider *how the data were produced*.

Using Chi-Square Tests Wisely

The **chi-square test for homogeneity** and the **chi-square test for independence** are very similar.

The best way to distinguish these two tests is to consider *how the data were produced*.

If the data come from two or more independent random samples or treatment groups in a randomized experiment, then do a **chi-square test for homogeneity**.

The **chi-square test for homogeneity** and the **chi-square test for independence** are very similar.

The best way to distinguish these two tests is to consider *how the data were produced*.

If the data come from two or more independent random samples or treatment groups in a randomized experiment, then do a **chi-square test for homogeneity**.

If the data come from a single random sample, with the individuals classified according to two categorical variables, use a **chi-square test for independence**.

NHL
Birthdays

Goodness of Fit

Gummy Bears
- Great Value
- Haribo

Homogeneity

English/math preference
Associated with

Independence

Selecting the Correct Chi-Square Test

- Work with ^{each other} each other to make a decision on the three situations.

1. Shopping at secondhand stores is becoming more popular and has even attracted the attention of business schools. A study of customers' attitudes toward secondhand stores interviewed separate random samples of shoppers at two secondhand stores of the same chain in different cities. The two-way table shows the breakdown of respondents by gender.

		Store		Total
		A	B	
Gender	Male	38	68	106
	Female	203	150	353
	Total	241	218	459

χ^2 for Homogeneity
 2 separate random samples
 1 variable (gender)

2. The General Social Survey (GSS) asked a random sample of adults their opinion about whether astrology is very scientific, sort of scientific, or not at all scientific. Here is a two-way table of counts for people in the sample who had three levels of higher education:

		Degree held			Total
		Associate's	Bachelor's	Master's	
Opinion about astrology	Not at all scientific	169	256	114	539
	Very or sort of scientific	65	65	18	148
	Total	234	321	132	687

χ^2 for Independence
 1 random sample
 2 variables (degree & opinion)

3. Casinos are required to verify that their games operate as advertised. American roulette wheels have 38 slots—18 red, 18 black, and 2 green. In one casino, managers record data from a random sample of 200 spins of one of their American roulette wheels. The table displays the results.

Color	Red	Black	Green
Count	85	99	16

χ^2 Goodness of fit
1 sample
1 variable (color)

Questions with Computer Output

On the AP Exam, you will see them.

Scary Movies and Fear

Are men and women equally likely to suffer lingering fear from watching scary movies as children? Researchers asked a random sample of 117 college students to write narrative accounts of their exposure to scary movies before the age of 13. More than one-fourth of the students said that some of the fright symptoms are still present when they are awake. The following table breaks down these results by gender.

		Gender		
		Male	Female	Total
Fright symptoms?	Yes	7	29	36
	No	31	50	81
	Total	38	79	117

Conditions for performing inference are met. Minitab output for a chi-square test using these

Chi-Square Test: Male, Female

Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

	Male	Female	Total
Yes	7	29	36
	11.69	24.31	
	1.883	0.906	
No	31	50	81
	26.31	54.69	
	0.837	0.403	
Total	38	79	117
Chi-Sq =	4.028,	DF = 1,	P-Value = 0.045

Should a test for independence or a chi-square test for homogeneity be used in this setting?

(a) Should a chi-square test for independence or a chi-square test for homogeneity be used in this setting?

The data was produced from a single random sample of college students who were then classified according to two variables, gender and whether or not they had lingering fright symptoms.

χ^2 Test for Homogeneity requires independ. rand. samples from each popul.

(c) Write the null hypothesis.

H_0 : There is no association between gender and whether or not college students have lingering fright symptoms.

(c) Which cell contributes most to the chi-square test statistic? In what way does this cell differ from what the null hypothesis suggests?

MALE FEMALE

Chi-Square Test: Male, Female

Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

	Male	Female	Total
Yes	7 11.69	29 24.31	36
No	31 26.31	50 54.69	81
Total	38 0.837	79 0.403	117

Chi-Sq = 4.028, DF = 1, P-Value = 0.045

largest component of χ^2 statistic

observed → 7
Expected → 11.69

→ 1.883

(c) Which cell contributes most to the chi-square test statistic? In what way does this cell differ from what the null hypothesis suggests?

Men who admit to having lingering fright symptoms account for the largest component of the chi-square test statistic (1.883).

Far fewer men in the sample admitted to lingering fright symptoms (7 men) than we would expect if H_0 were true (11.69 men)

LCQ

11.2....41, 43, 47, 49, 51, 55-60