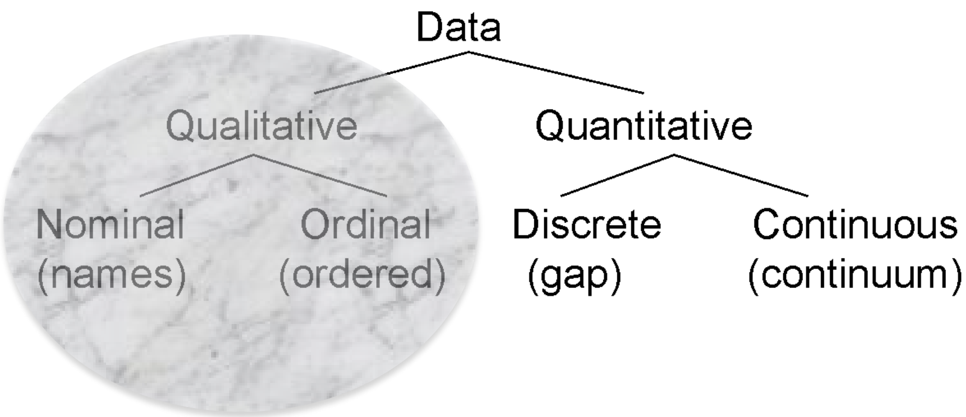
11 Applications of Chi-Square

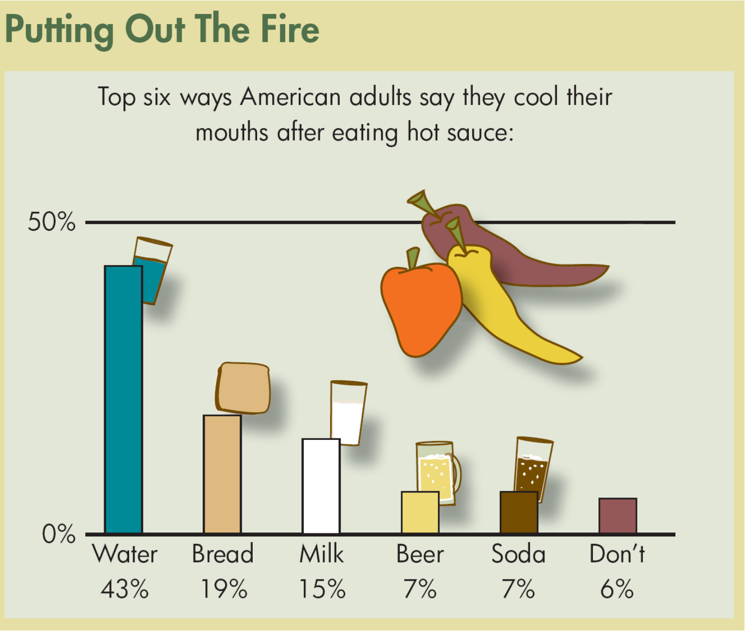


Cooling a Great Hot Taste

If you like hot foods, you probably have a favorite hot sauce and preferred way to "cool" your mouth after eating a mind-blowing spicy morsel.

Some of the more common methods used by people are drinking water, milk, soda, or beer or eating bread or other food.

There are even a few who prefer not to cool their mouth on such occasions and therefore do nothing.



Source: Data from Anne R. Carey and Suzy Parker, © 1995 *USA Today*.

Cooling a Great Hot Taste

Recently a sample of two hundred adults professing to love hot spicy food were asked to name their favorite way to cool their mouth after eating food with hot sauce.

The table summarizes the responses. **[EX11- 01]**

Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Number	73	29	35	19	20	13	11

Count data like these are often referred to as enumerative data.

Data Set-Up

- The **observed frequencies** in each cell are denoted by $O_1, O_2, O_3, \ldots, O_k$

	k Categories					
	1st	2nd	3rd	. . .	*k*th	Total
Observed frequencies	O_1	O_2	O_3	. . .	O_k	n
Expected frequencies	E_1	E_2	E_3	. . .	E_k	n

- Note that the sum of all the observed frequencies is

$$O_1 + O_2 + \ldots + O_k = n$$

where n is the sample size.

- We would like to compare the observed frequencies with **expected frequencies**, denoted by $E_1, E_2, E_3, \ldots, E_k$

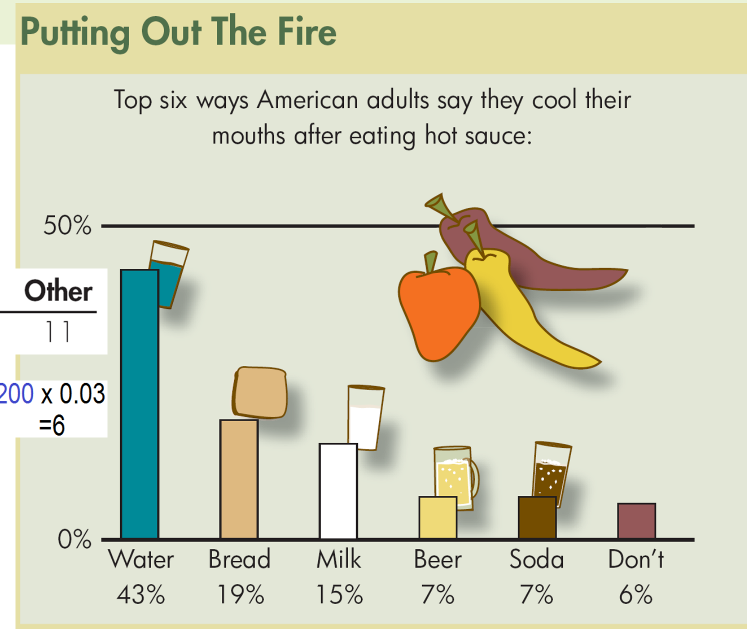
- Again, the sum of these expected frequencies is exactly n:

$$E_1 + E_2 + \ldots + E_k = n$$

Data Set-Up

- For a sample of 200 adults, Observed counts are:

Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Observed	73	29	35	19	20	13	11
Expected	200 x 0.43 =86	200 x 0.19 =38	200 x 0.15 =30	200 x 0.07 =14	200 x 0.07 =14	200 x 0.06 =12	200 x 0.03 =6



- Expected counts are:
  - $E_i = n\, p_i$

- We will then decide whether the observed frequencies seem to agree or disagree with the expected frequencies.

- We will do this by using a **hypothesis test** with **chi-square**, χ^2 ("ki-square"; that's "ki" as in *kite;* χ is the Greek lowercase letter chi).

Outline of Test Procedure

Outline of Test Procedure

Test Statistic for Chi-Square

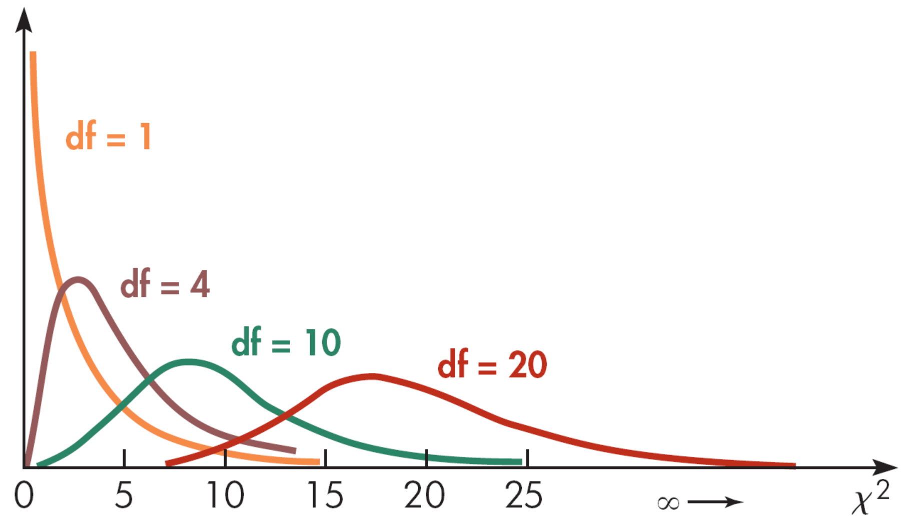
$$\chi^2\star = \sum_{\text{all cells}} \frac{(O - E)^2}{E} \quad (11.1)$$

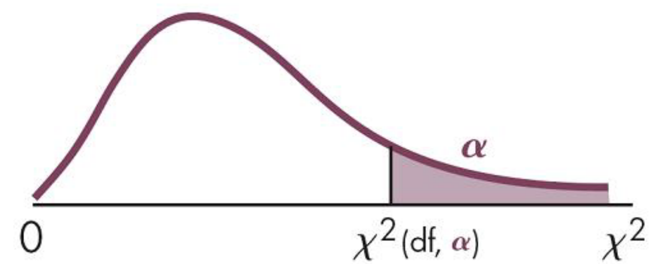
- In repeated sampling, the calculated value of $\chi^2\star$ in formula (11.1) will have a sampling distribution that can be approximated by the chi-square probability distribution when *n* is large.

- This approximation is generally considered adequate when all the expected frequencies are equal to or greater than 5.

Chi-Squared (χ^2) Distribution

- Right skewed distribution
- Defined over positive numbers
- Parameter: df
- How to write:
  - $\chi^2(\mathrm{df})$
- How to find probabilities?
  - Chi-Squared Calculator
  - χ^2-table (“D2L > Useful Links > Z, T and χ^2 Tables”)
    - $P(\chi^2 \geq c_\alpha)$
  - χ^2-table (from book, next slide)



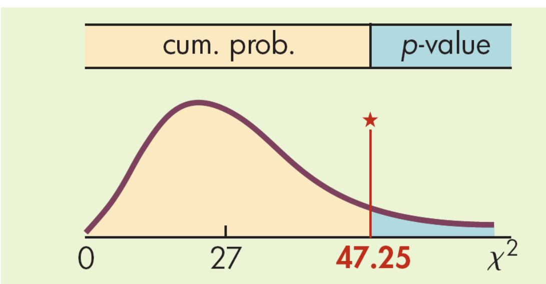


χ^2 Table

- $\chi^2(\text{df}, \text{area to right})$
  - $df = 28$, area to right=0.9 $\rightarrow$ area to left=0.1
  - $\chi^2(df = 28, \text{area to right=0.9}) = 18.9$

- Example in a Hypothesis Test:

- Lets say $\chi^2_* = 47.25$ and you want to find:
- $\text{p} - \text{value} = P(\chi^2 > \chi^2_*, \text{with df} = 27)$



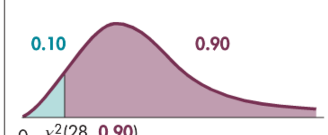
- $0.005 < \text{p} - \text{value} < 0.01$
- So if $\alpha = 0.05$, we have
- $\text{p} - \text{value} < \alpha$, so we Reject H_0

a) Area to the Right

b) Area to the Left (the Cumulative Area)

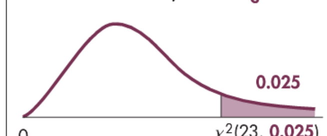
df	0.005	0.01	0.025	0.05	0.10	0.25	0.50 (Median)	0.75	0.90	0.95	0.975	0.99	0.995
a) Area to the Right	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.34	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.34	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.34	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.34	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.34	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.34	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.34	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.34	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.34	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.34	27.1	32.0	35.2	$\boxed{38.1}$	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.34	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.34	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.34	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.34	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	$\boxed{18.9}$	22.7	27.34	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.34	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.34	34.8	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	33.7	39.34	45.6	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	42.9	49.33	56.3	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	52.3	59.33	67.0	74.4	79.1	83.3	88.4	92.0
70	43.3	45.4	48.8	51.7	55.3	61.7	69.33	77.6	85.5	90.5	95.0	100.4	104.2
80	51.2	53.5	57.2	60.4	64.3	71.1	79.33	88.1	96.6	101.9	106.6	112.3	116.3
90	59.2	61.8	65.6	69.1	73.3	80.6	89.33	98.6	107.6	113.1	118.1	124.1	128.3
100	67.3	70.1	74.2	77.9	82.4	90.1	99.33	109.1	118.5	124.3	129.6	135.8	140.2

Left-tail example:
Find χ^2 with df = 28; area in left-tail = 0.10.



χ^2(df, area to right) = χ^2(28, 0.90) = 18.9

Right-tail example:
Find χ^2 with df = 23; area in right-tail = 0.025



χ^2(df, area to right) = χ^2(23, 0.025) = 38.1

11.2 Inferences Concerning Multinomial Experiments

Inferences Concerning Multinomial Experiments

Multinomial experiment A multinomial experiment has the following characteristics:

1. It consists of n identical independent trials.

2. The outcome of each trial fits into exactly one of k possible cells.

3. There is a probability associated with each particular cell, and these individual probabilities remain constant during the experiment. (It must be the case that $p_1 + p_2 + \ldots + p_k = 1$.)

Inferences Concerning Multinomial Experiments

4. The experiment will result in a set of k observed frequencies, O_1, O_2,…, O_k where each O_i is the number of times a trial outcome falls into that particular cell. (It must be the case that $O_1 + O_2 + \ldots + O_k = n$.)

Degrees of Freedom for Multinomial Experiments

$$\text{df} = k - 1 \tag{11.2}$$

Expected Value for Multinomial Experiments

$$E_i = n \cdot p_i \tag{11.3}$$

Inferences Concerning Multinomial Experiments

The preceding die problem is a good illustration of a **multinomial experiment**. Let's consider this problem again.

Suppose that we want to test this die (at $\alpha = 0.05$) and decide whether to fail to reject or reject the claim "This die is fair." (The probability of each number is $\frac{1}{6}$.) The die is rolled from a cup onto a smooth, flat surface 60 times, with the following observed frequencies:

Number	1	2	3	4	5	6
Observed frequency	7	12	10	12	8	11

Inferences Concerning Multinomial Experiments

Number	1	2	3	4	5	6
Observed frequency	7	12	10	12	8	11

The null hypothesis that the die is fair is assumed to be true. This allows us to calculate the expected frequencies. If the die is fair, we certainly expect 10 occurrences of each number.

Now let's calculate an observed value of χ^2. These calculations are shown in next Table.

Number	Observed (O)	Expected (E)	$O - E$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
1	7	10	−3	9	0.9
2	12	10	2	4	0.4
3	10	10	0	0	0.0
4	12	10	2	4	0.4
5	8	10	−2	4	0.4
6	11	10	1	1	0.1
Total	60	60	0 ck		**2.2**

The calculated value is $\chi^{2\bigstar} = 2.2$

Inferences Concerning Multinomial Experiments

Now let's use our familiar hypothesis-testing format.

Step 1 **a. Parameter of interest:** The probability with which each side faces up: $P(1)$, $P(2)$, $P(3)$, $P(4)$, $P(5)$, $P(6)$

b. Statement of hypotheses:

H_o: The die is fair (each $p = \frac{1}{6}$).

H_a: The die is not fair (at least one p is different from the others).

Inferences Concerning Multinomial Experiments

Step 2 **a. Assumptions:** The data were collected in a random manner, and each outcome is one of the six numbers.

b. Test statistic: The chi-square distribution and formula $\chi^2\star = \sum_{\text{all cells}} \frac{(O - E)^2}{E}$, with df = k – 1 = 6 – 1 = 5

In a multinomial experiment, df = k – 1, where k is the number of cells.

c. Level of significance: $\alpha = 0.05$

Inferences Concerning Multinomial Experiments

Step 3 **a. Sample information:** See Table 11.2.

Number	Observed (O)	Expected (E)	$O - E$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
1	7	10	−3	9	0.9
2	12	10	2	4	0.4
3	10	10	0	0	0.0
4	12	10	2	4	0.4
5	8	10	−2	4	0.4
6	11	10	1	1	0.1
Total	60	60	0 ck		**2.2**

Computations for Calculating χ^2

Table 11.2

b. Calculated test statistic: Using formula

$$\chi^2\bigstar = \sum_{\text{all cells}} \frac{(O-E)^2}{E}$$ we have

$\chi^{2\bigstar}$ = **2.2** (calculations are shown in Table 11.2)

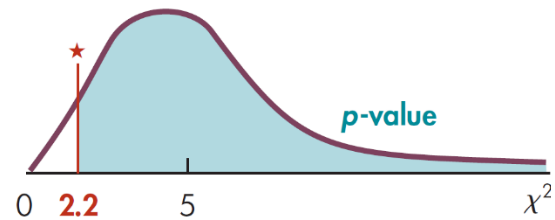
Inferences Concerning Multinomial Experiments

Step 4 Probability Distribution:

***p*-Value:**

a. Use the right-hand tail because "larger" values of chi-square disagree with the null hypothesis:

$$p\text{-value} = P(\chi^{2\bigstar} > 2.2 \mid \text{df} = 5)$$



p-value:

0.75 < P < 0.90.

a) Area to the Right

	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
b) Area to the Left (the Cumulative Area)							Median						
df	**0.005**	**0.01**	**0.025**	**0.05**	**0.10**	**0.25**	**0.50**	**0.75**	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7

b. The *p*-value is not smaller than the level of significance, α.

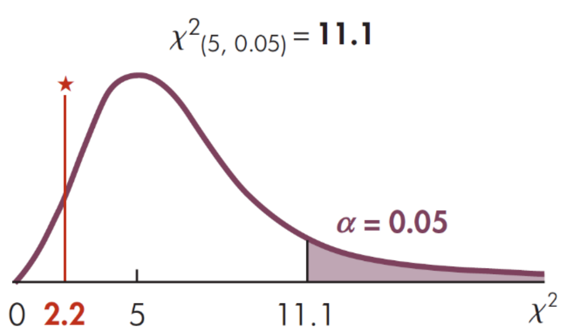
Inferences Concerning Multinomial Experiments

Step 4 Probability Distribution:

Classical:

a. The critical region is the right-hand tail because "larger" values of chi-square disagree with the null hypothesis.

The critical value is obtained at the intersection of row df = 5 and column $\alpha = 0.05$:



a) Area to the Right

b) Area to the Left (the Cumulative Area)

a) Area to the Right	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
							Median						
df	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975	0.99	0.995
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7

b. $\chi^2\star$ is not in the critical region, as shown in **red** in the figure.

Inferences Concerning Multinomial Experiments

Step 5 **a. Decision:** Fail to reject H_o.

b. Conclusion: At the 0.05 level of significance, the observed frequencies are not significantly different from those expected of a fair die.

11.3 Inferences Concerning Contingency Tables

Inferences Concerning Contingency Tables

A **contingency table** is an arrangement of data in a two-way classification. The data are sorted into cells, and the count for each cell is reported.

The contingency table involves two factors (or variables), and a common question concerning such tables is whether the data indicate that the two variables are independent or dependent.

Two different tests use the contingency table format. The first one we will look at is the *test of independence*.

Example 5 – *Hypothesis Test for Independence*

Each person in a group of 300 students was identified as male or female and then asked whether he or she prefers taking liberal arts courses in the area of math–science, social science, or humanities. Table 11.5 is a contingency table that shows the frequencies found for these categories.

	Favorite Subject Area			
Gender	Math–Science (MS)	Social Science (SS)	Humanities (H)	Total
Male (M)	37	41	44	122
Female (F)	35	72	71	178
Total	72	113	115	300

Sample Results for Gender and Subject Preference

Table 11.5

Test of Independence

In general, the ***r*** × ***c*** **contingency table** (*r* is the number of **rows**; *c* is the number of **columns**) is used to test the independence of the row factor and the column factor. The number of **degrees of freedom** is determined by

Degrees of Freedom for Contingency Tables

$$\text{df} = (r - 1) \cdot (c - 1) \tag{11.4}$$

where *r* and *c* are both greater than 1

	A_1	$\cdots$	A_{c-1}	A_c	**Total**
B_1	O_{11}	$\cdots$	$O_{1\,c-1}$		R_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
B_{r-1}	$O_{r-1\,1}$	$\cdots$	$O_{r-1\,c-1}$		R_{r-1}
B_r					R_r
Total	C_1	$\cdots$	C_{c-1}	C_c	n

Test of Independence

- **Observed Counts**

	A_1	A_2	$\cdots$	A_c	**Total**
B_1	O_{11}	O_{12}	$\cdots$	O_{1c}	R_1
B_2	O_{21}	O_{22}	$\cdots$	O_{2c}	R_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	R_r
Total	C_1	C_2	$\cdots$	C_c	n

- **Expected Counts**

	A_1	A_2	$\cdots$	A_c	**Total**
B_1	$E_{11} = \frac{R_1 \times C_1}{n}$	$E_{12} = \frac{R_1 \times C_2}{n}$	$\cdots$	$E_{1c} = \frac{R_1 \times C_c}{n}$	R_1
B_2	$E_{21} = \frac{R_2 \times C_1}{n}$	$E_{22} = \frac{R_2 \times C_2}{n}$	$\cdots$	$E_{2c} = \frac{R_2 \times C_c}{n}$	R_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_r	$E_{r1} = \frac{R_r \times C_1}{n}$	$E_{r2} = \frac{R_r \times C_2}{n}$	$\cdots$	$E_{rc} = \frac{R_r \times C_c}{n}$	R_r
Total	C_1	C_2	$\cdots$	C_c	n

- **If A_i and B_j are independent** $P(A_i \text{ and } B_j) = P(A_i) \times P(B_j)$

Test of Independence

In general, the expected frequency at the intersection of the *i*th row and the *j*th column is given by

Expected Frequencies for Contingency Tables

$$E_{ij} = \frac{\text{row total} \times \text{column total}}{\text{grand total}} = \frac{R_i \times C_j}{n} \qquad \textbf{(11.5)}$$

We should again observe the previously mentioned guideline: Each $E_{i,j}$ should be at least 5.

Example 5 – *Hypothesis Test for Independence* cont'd

Does this sample present sufficient evidence to reject the null hypothesis "Preference for math–science, social science, or humanities is independent of the gender of a college student"? Complete the **hypothesis** test using the 0.05 level of significance.

Solution:

Step 1 **a. Parameter of interest:** Determining the independence of the variables "gender" and "favorite subject area" requires us to discuss the probability of the various cases and the effect that answers about one variable have on the probability of answers about the other variable.

Example 5 – *Solution*

cont'd

Independence, as defined in Chapter 4, requires $P(MS \mid M) = P(MS \mid F) = P(MS)$; that is, gender has no effect on the probability of a person's choice of subject area.

b. Statement of hypotheses:

H_o: Preference for math–science, social science, or humanities is independent of the gender of a college student.

H_a: Subject area preference is not independent of the gender of the student.

Example 5 – *Solution*

cont'd

Step 2 **a. Assumptions:** The sample information is obtained using one random sample drawn from one population, with each individual then classified according to gender and favorite subject area.

b. Test statistic: In the case of contingency tables, the number of degrees of freedom is exactly the same as the number of cells in the table that may be filled in freely when you are given the marginal totals.

Example 5 – *Solution*

cont'd

The totals in this example are shown in the following table:

			122
			178
72	113	115	300

Given these totals, you can fill in only two cells before the others are all determined. (The totals must, of course, remain the same.) For example, once we pick two arbitrary values (say, 50 and 60) for the first two cells of the first row, the other four cell values are fixed (see the following table):

50	60	C	122
D	E	F	178
72	113	115	300

Example 5 – *Solution*

50	60	*C*	122
D	*E*	*F*	178
72	113	115	300

The values have to be $C = 12$, $D = 22$, $E = 53$, and $F = 103$. Otherwise, the totals will not be correct. Therefore, for this problem there are two free choices. Each free choice corresponds to 1 degree of freedom. Hence, the number of degrees of freedom for our example is 2 (df = 2).

The chi-square distribution will be used along with formula (11.1), with df = 2.

$$\chi^2\star = \sum_{\text{all cells}} \frac{(O - E)^2}{E} \qquad \textbf{(11.1)}$$

c. Level of significance: $\alpha = 0.05$

Example 5 – *Solution*

cont'd

Step 3 **a. Sample information:** See Table 11.5.

Gender	Favorite Subject Area			Total
	Math–Science (MS)	Social Science (SS)	Humanities (H)	
Male (M)	37	41	44	122
Female (F)	35	72	71	178
Total	72	113	115	300

b. Calculated test statistic: Before we can calculate the value of chi-square, we need to determine the expected values, *E*, for each cell.

	MS	SS	H	**Total**
M	$E_{11} = 29.28$	$E_{12} = 45.95$	$E_{13} = 46.77$	122
F	$E_{21} = 42.72$	$E_{22} = 67.05$	$E_{23} = 68.23$	178
Total	72	113	115	300

Example 5 – *Solution*

cont'd

Typically, the contingency table is written so that it contains all this information (see Table 11.7).

	Favorite Subject Area			
Gender	**MS**	**SS**	**H**	**Total**
Male	37 (29.28)	41 (45.95)	44 (46.77)	122
Female	35 (42.72)	72 (67.05)	71 (68.23)	178
Total	72	113	115	300

The calculated chi-square is

$$\chi^2\bigstar = \sum_{\text{all cells}} \frac{(O-E)^2}{E}: \chi^2\bigstar = \frac{(37-29.28)^2}{29.28} + \frac{(41-45.95)^2}{45.95} + \frac{(44-46.77)^2}{46.77} + \frac{(35-42.72)^2}{42.72} + \frac{(72-67.05)^2}{67.05} + \frac{(71-68.23)^2}{68.23}$$

$= 2.035 + 0.533 + 0.164 + 1.395 + 0.365 + 0.112$

$=$ **4.604**

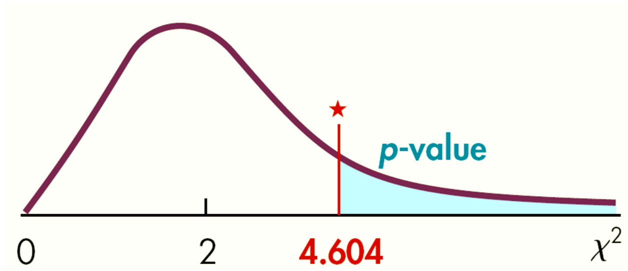
Example 5 – *Solution*

cont'd

Step 4 Probability Distribution:

***p*-Value:**

a. Use the right-hand tail because "larger" values of chi-square disagree with the null hypothesis: **P** = $\boldsymbol{P(\chi^{2}\star > 4.604 \mid df = 2)}$, as shown in the figure.

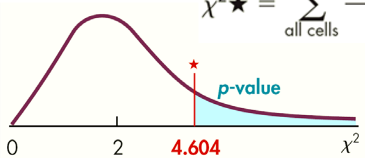


To find the p-value, you have two options:

1. Use Table 8 (Appendix B) to place bounds on the p-value: **0.10 < P < 0.25.**

Example 5 – *Solution*

cont'd



$$\chi^2\star = \sum_{\text{all cells}} \frac{(O-E)^2}{E}: \ \chi^2\star = \mathbf{4.604}$$

p-value: **0.10 < P < 0.25**

a) Area to the Right	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
b) Area to the Left (the Cumulative Area)							Median						
df	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975	0.99	0.995
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7

2. Use a computer or calculator to find the p-value: **P = 0.1001.** DistCalc

b. The p-value is not smaller than α .

Classical:

a. The critical region is the right-hand tail because "larger" values of chi-square disagree with the null hypothesis. The critical value is obtained from Table 8, at the intersection of row df = 2 and column $\alpha = 0.05$:

Example 5 – *Solution*

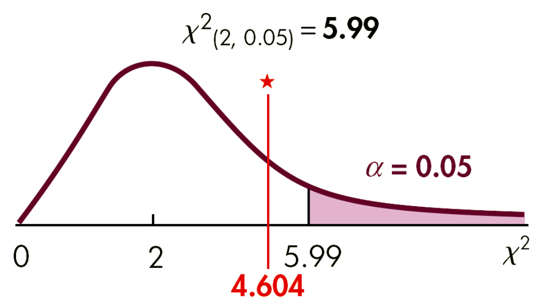
cont'd

a) Area to the Right

b) Area to the Left (the Cumulative Area)

df	0.005	0.01	0.025	0.05	0.10	0.25	0.50 (Median)	0.75	0.90	0.95	0.975	0.99	0.995
(Area to the Right)	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7

$$\chi^2\bigstar = \sum_{\text{all cells}} \frac{(O-E)^2}{E}: \ \chi^2\bigstar = \mathbf{4.604}$$



b. $\chi^2\bigstar$ is not in the critical region, as shown in **red** in the figure.

Step 5 **a. Decision:** Fail to reject H_o.

b. Conclusion: At the 0.05 level of significance, the evidence does not allow us to reject independence between the gender of a student and the student's preferred academic subject area.

Test of Homogeneity

Test of Homogeneity

The second type of contingency table problem is called a *test of homogeneity*. This test is used when one of the two variables is controlled by the experimenter so that the row or column totals are predetermined.

For example, suppose that we want to poll registered voters about a piece of legislation proposed by the governor. In the poll, 200 urban, 200 suburban, and 100 rural residents are randomly selected and asked whether they favor or oppose the governor's proposal.

Test of Homogeneity

That is, a simple random sample is taken for each of these three groups.

A total of 500 voters are polled. But notice that it has been predetermined (before the sample is taken) just how many are to fall within each row category, as shown in Table 11.9, and each category is sampled separately.

	Governor's Proposal		
Residence	Favor	Oppose	Total
Urban			200
Suburban			200
Rural			100
Total			500

Registered Voter Poll with Predetermined Row Totals

Table 11.9

Example 6 – *Hypothesis Test for Homogeneity*

Each person in a random sample of 500 registered voters (200 urban, 200 suburban, and 100 rural residents) was asked his or her opinion about the governor's proposed legislation. Does the sample evidence shown in Table 11.10 support the hypothesis "Voters within the different residence groups have different opinions about the governor's proposal"? Use $\alpha = 0.05$.

	Governor's Proposal		
Residence	Favor	Oppose	Total
Urban	143	57	200
Suburban	98	102	200
Rural	13	87	100
Total	254	246	500

Sample Results for Residence and Opinion

Table 11.10

Example 6 – *Solution*

cont'd

Step 1 **a. Parameter of interest:** The proportion of voters who favor or oppose (i.e., the proportion of urban voters who favor, the proportion of suburban voters who favor, the proportion of rural voters who favor, and the proportion of all three groups, separately, who oppose)

b. Statement of hypotheses:

H_o: The proportion of voters who favor the proposed legislation is the same in all three residence groups.

Example 6 – *Solution*

cont'd

H_a: The proportion of voters who favor the proposed legislation is not the same in all three groups. (That is, in at least one group, the proportion is different from the others.)

Step 2 **a. Assumptions:** The sample information is obtained using three random samples drawn from three separate populations in which each individual is classified according to his or her opinion.

Example 6 – *Solution*

cont'd

b. Test statistic: The chi-square distribution and formula (11.1), with
df = $(r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$

c. Level of significance: $\alpha = 0.05$

Step 3 **a. Sample information:** See Table 11.10.

Residence	Governor's Proposal: Favor	Governor's Proposal: Oppose	Total
Urban	143	57	200
Suburban	98	102	200
Rural	13	87	100
Total	254	246	500

Sample Results for Residence and Opinion

Table 11.10

Example 6 – *Solution*

cont'd

b. Calculated test statistic: The expected values are found by using formula (11.5) and are given in Table 11.11.

Residence	Governor's Proposal		Total
	Favor	Oppose	
Urban	143 (101.6)	57 (98.4)	200
Suburban	98 (101.6)	102 (98.4)	200
Rural	13 (50.8)	87 (49.2)	100
Total	254	246	500

Sample Results and Expected Values

Table 11.11

Example 6 – *Solution*

Residence	Governor's Proposal: Favor	Governor's Proposal: Oppose	Total
Urban	143 (101.6)	57 (98.4)	200
Suburban	98 (101.6)	102 (98.4)	200
Rural	13 (50.8)	87 (49.2)	100
Total	254	246	500

The calculated chi-square is

$$\chi^2\bigstar = \sum_{\text{all cells}} \frac{(O-E)^2}{E}: \ \chi^2\bigstar = \frac{(143-101.6)^2}{101.6} + \frac{(57-98.4)^2}{98.4} + \frac{(98-101.6)^2}{101.6} + \frac{(102-98.4)^2}{98.4} + \frac{(13-50.8)^2}{50.8} + \frac{(87-49.2)^2}{49.2}$$

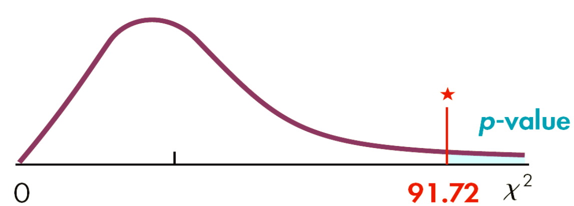
$$= 16.87 + 17.42 + 0.13 + 0.13 + 28.13 + 29.04$$

$$= \mathbf{91.72}$$

Step 4 Probability Distribution:

***p*-Value:**

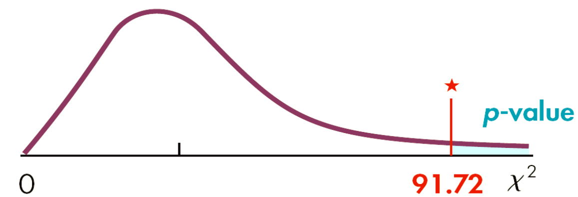
a. Use the right-hand tail because "larger" values of chi-square disagree with the null hypothesis: **P = *P*(χ^2 > 91.72 | df = 2),** as shown in the figure.



Example 6 – *Solution*

cont'd

$$\chi^2\star = \sum_{\text{all cells}} \frac{(O - E)^2}{E} : \chi^2\star = \mathbf{91.72}$$



To find the p-value, you have two options:

1. Use Table 8 (Appendix B) to place bounds on the p-value:

a) Area to the Right

b) Area to the Left (the Cumulative Area)

	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	
							Median							
df	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975	0.99	0.995	
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88	
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	91.72
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9	
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7	

$P < 0.005$.

2. Use a computer or calculator to find the p-value: **$P = 0.000+$.** DistCalc

b. The p-value is not smaller than α .

Example 6 – *Solution*

cont'd

Classical:

a. The critical region is the right-hand tail because "larger" values of chi-square disagree with the null hypothesis. The critical value is obtained from Table 8, at the intersection of row df = 2 and column α = 0.05:

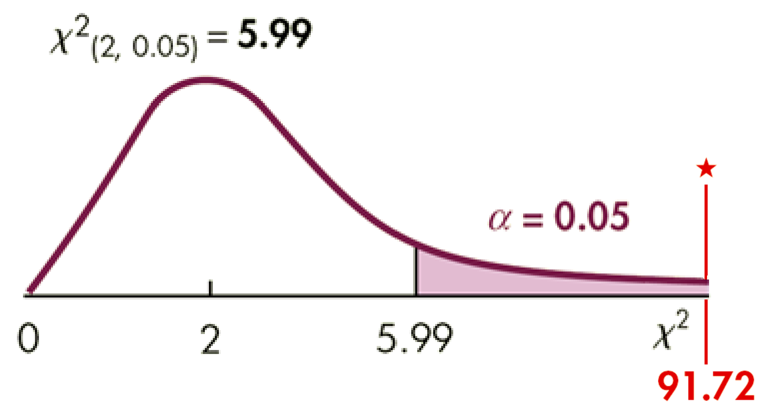
a) Area to the Right

b) Area to the Left (the Cumulative Area) — Median

Area to the Right	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
df	**0.005**	**0.01**	**0.025**	**0.05**	**0.10**	**0.25**	**0.50**	**0.75**	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7

b. $\chi^{2\bigstar}$ is not in the critical region, as shown in **red** in the figure.

$$\chi^2\bigstar = \sum_{\text{all cells}} \frac{(O-E)^2}{E}:\ \chi^2\bigstar = \mathbf{91.72}$$



Example 6 – *Solution*

cont'd

Step 5 **a. Decision:** Reject H_o.

b. Conclusion: The three groups of voters do not all have the same proportions favoring the proposed legislation.