Chi-Square Test

Categorical Variables

- Based on observations

- **Univariate** – single categorical variable
  - Example: Sample 100 people & ask if they agree or disagree with a question.
- **Bivariate** – uses two categorical variables
  - Example: Sample 100 people & ask if they are male/female and what political party they support.

One-Way Frequency Table - univariate

Data →

Democrat	Democrat	Democrat	Independent
Republican	Democrat	Republican	Independent
Republican	Republican	Republican	Republican

Vertical One-Way Table

	Freq.
Democrat	4
Republican	6
Independent	2

Horizontal One-Way Table

	Democrat	Republican	Independent
Freq.	4	6	2

Goodness of Fit Test χ^2

- Used to measure the extent to which the observed counts differ from the expected counts.
- K = # categories of a catagorical variable
- Df = k – 1
- Test Statistic: $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$

Degrees of freedom, often represented by v or df, is **the number of independent pieces of information used to calculate a statistic**. It's calculated as the sample size minus the number of restrictions.

Assumptions

- Observed Values are based on random Samples
- Sample size is large – each cell count is at least 5.

Hypotheses

- H_o: State each proportion’s hypothesized value.
- H_A: At least 1 of the proportions differ from the hypothesized value.

It uses the Chi-Square Chart

- Positively Skewed
- Uses d.f.
- On calculator!

Table entry for p is the point (χ^2) with probability p lying above it.

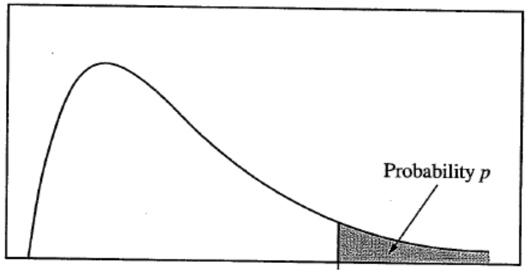


Table C χ^2 **critical values**

df	Tail probability p .25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

A school principal would like to know which days of the week students are most likely to be absent. The principal expects that students will be absent equally during the 5-day School week. The principal selects a random sample of 100 teachers asking them which day of this week they had the highest number of student absences. The observed and expected Results are shown in the table below. Based on these results, do the days for the highest number Of absences occur with equal frequencies? (use a 5% significance level)

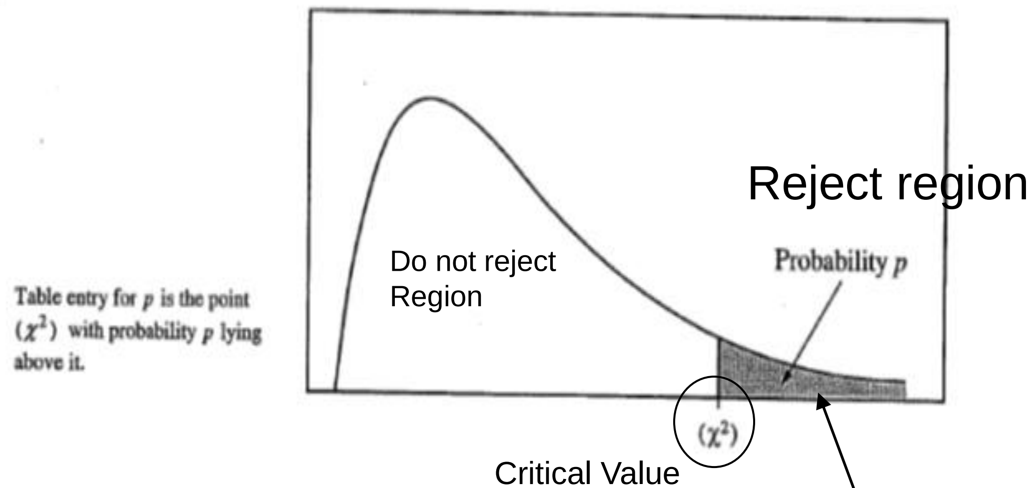
	Monday	Tuesday	Wednesday	Thursday	Friday
Observed absences	23	16	14	19	28
Expected absences	20	20	20	20	20

A school principal would like to know which days of the week students are most likely
to be absent. The principal expects that students will be absent equally during the 5-day School week. The principal selects a random sample of 100 teachers asking them which day of this week they had the highest number of student absences. The observed and expected Results are shown in the table below. Based on these results, do the days for the highest number Of absences occur with equal frequencies? (use a 5% significance level)

	Monday	Tuesday	Wednesday	Thursday	Friday
Observed absences	23	16	14	19	28
Expected absences	20	20	20	20	20

1st Hypothesis

$$H_0 = equal\ frequencies$$
$$H_a = unequal\ frequencies$$



5% significance level

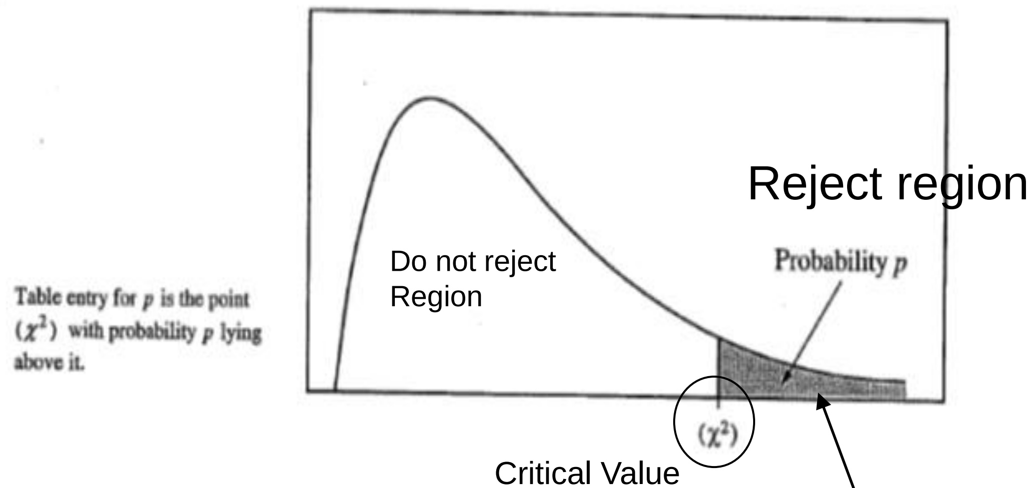
$$\alpha = 0.05$$

A school principal would like to know which days of the week students are most likely to be absent. The principal expects that students will be absent equally during the 5-day School week. The principal selects a random sample of 100 teachers asking them which day of this week they had the highest number of student absences. The observed and expected Results are shown in the table below. Based on these results, do the days for the highest number Of absences occur with equal frequencies? (use a 5% significance level)

	Monday	Tuesday	Wednesday	Thursday	Friday
Observed absences	23	16	14	19	28
Expected absences	20	20	20	20	20

1st Hypothesis

$$H_0 = equal\ frequencies$$
$$H_a = unequal\ frequencies$$



2nd Degree of freedom

Df = n-1 = 5-1 =4

5% significance level

$\alpha = 0.05$

Chi-Square (χ^2) Distribution

Area to the Right of Critical Value

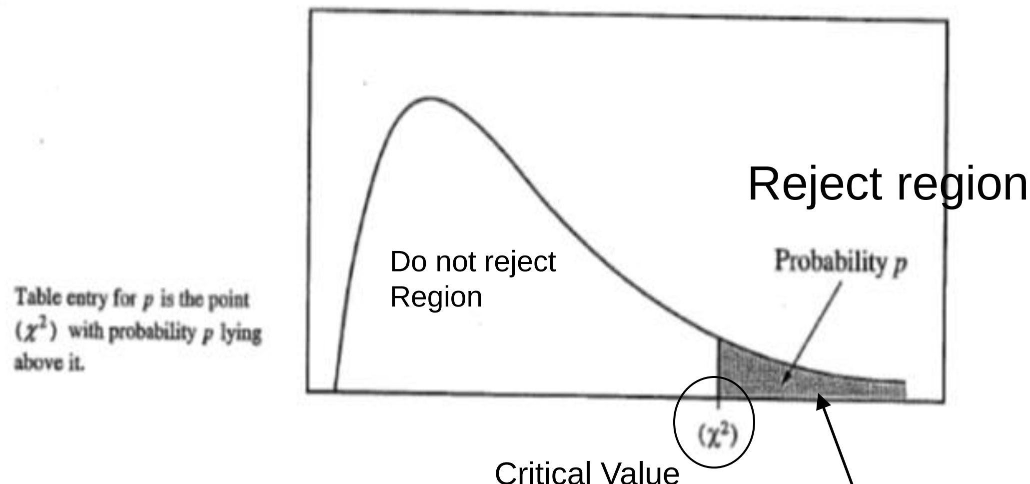
Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

A school principal would like to know which days of the week students are most likely
to be absent. The principal expects that students will be absent equally during the 5-day School week. The principal selects a random sample of 100 teachers asking them which day of this week they had the highest number of student absences. The observed and expected Results are shown in the table below. Based on these results, do the days for the highest number Of absences occur with equal frequencies? (use a 5% significance level)

	Monday	Tuesday	Wednesday	Thursday	Friday
Observed absences	23	16	14	19	28
Expected absences	20	20	20	20	20

1st Hypothesis

$$H_0 = equal\ frequencies$$
$$H_a = unequal\ frequencies$$



2nd Degree of freedom

Df = n-1 = 5-1 =4

Critical Value = 9.49

5% significance level

$\alpha = 0.05$

A school principal would like to know which days of the week students are most likely to be absent. The principal expects that students will be absent equally during the 5-day School week. The principal selects a random sample of 100 teachers asking them which day of this week they had the highest number of student absences. The observed and expected Results are shown in the table below. Based on these results, do the days for the highest number Of absences occur with equal frequencies? (use a 5% significance level)

	Monday	Tuesday	Wednesday	Thursday	Friday
Observed absences	23	16	14	19	28
Expected absences	20	20	20	20	20

1st Hypothesis

$$H_0 = equal\ frequencies$$
$$H_a = unequal\ frequencies$$

2nd Degree of freedom

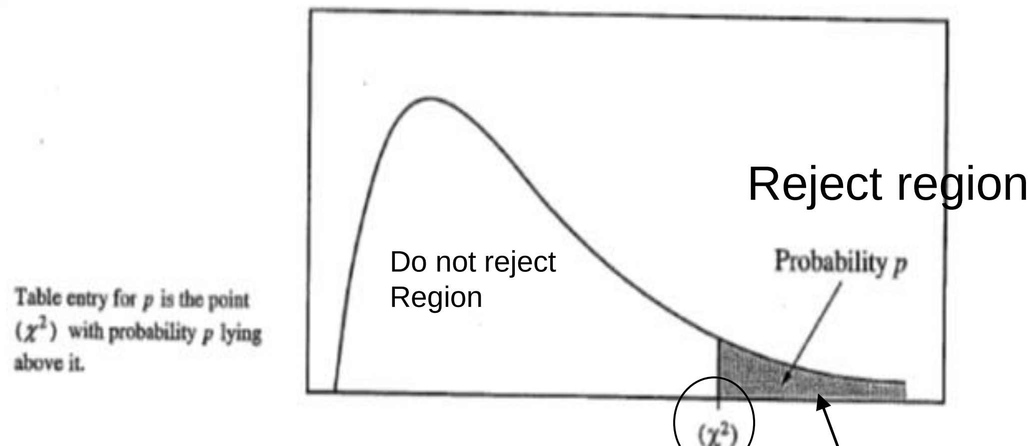
Df = n-1 = 5-1 =4

Critical Value = 9.49

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

Calculated value,

$$x^2 = \frac{3^2}{20} + \frac{(-4)^2}{20} + \frac{(-6)^2}{20} + \frac{(-1)^2}{20} + \frac{8^2}{20} = \frac{126}{20} = 6.3$$



Critical Value =9.49

5% significance level

$$\alpha = 0.05$$

A school principal would like to know which days of the week students are most likely to be absent. The principal expects that students will be absent equally during the 5-day School week. The principal selects a random sample of 100 teachers asking them which day of this week they had the highest number of student absences. The observed and expected Results are shown in the table below. Based on these results, do the days for the highest number Of absences occur with equal frequencies? (use a 5% significance level)

	Monday	Tuesday	Wednesday	Thursday	Friday
Observed absences	23	16	14	19	28
Expected absences	20	20	20	20	20

1st Hypothesis

$$H_0 = equal\ frequencies$$
$$H_a = unequal\ frequencies$$

2nd Degree of freedom

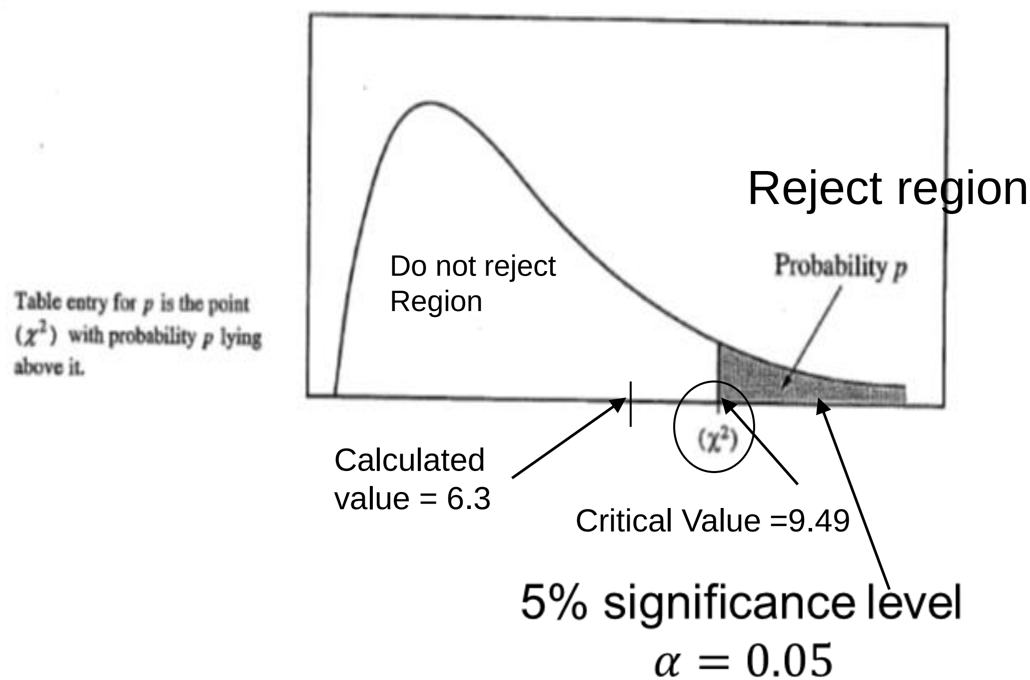
Df = n-1 = 5-1 =4

Critical Value = 9.49

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

Calculated value,

$$x^2 = \frac{3^2}{20} + \frac{(-4)^2}{20} + \frac{(-6)^2}{20} + \frac{(-1)^2}{20} + \frac{8^2}{20} = \frac{126}{20} = 6.3$$



Chi-square test of association

- Hypothetical outbreak of *Salmonella* on a cruise ship
  - Retrospective cohort study conducted
  - All 300 people on cruise ship interviewed, 60 had symptoms consistent with *Salmonella*
  - Questionnaires indicate many of the case-patients ate tomatoes from the salad bar

Chi-Square Test: Example (cont.)

Table a. Cohort study: Exposure to tomatoes and *Salmonella* infection

	Salmonella?		
	Yes	No	Total
Tomatoes	41	89	130
No Tomatoes	19	151	170
Total	60	240	300

- To see if there is a statistical difference in the amount of illness between those who ate tomatoes (41/130) and those who did not (19/170) we could conduct a chi-square test

Chi-Square Test: Example (cont.)

- To conduct a chi-square the following conditions must be met:
  - There must be at least a total of 30 observations (people) in the table
  - Each cell must contain a count of 5 or more
- To conduct a chi-square test we compare the observed data (from study results) with the data we would expect to see

Chi-Square Test: Example (cont.)

Table b. Row and column totals for tomatoes and *Salmonella* infection

	Salmonella?		
	Yes	No	Total
Tomatoes			130
No Tomatoes			170
Total	60	240	300

- Gives an overall distribution of people who ate tomatoes and became sick
- Based on these distributions we can fill in the empty cells with the **expected values**

Chi-Square Test: Example (cont.)

- Expected Value = $\dfrac{\text{Row Total x Column Total}}{\text{Grand Total}}$

- For the first cell, people who ate tomatoes and became ill:
  - Expected value = $\dfrac{130 \times 60}{300} = 26$
- Same formula can be used to calculate the expected values for each of the cells

Chi-Square Test: Example (cont.)

Table c. Expected values for exposure to tomatoes

Expected Value = $\frac{\text{Row Total x Column Total}}{\text{Grand Total}}$

	Salmonella? Yes	No	Total
Tomatoes	$\frac{130 \times 60}{300} = 26$	$\frac{130 \times 240}{300} = 104$	130
No Tomatoes	$\frac{170 \times 60}{300} = 34$	$\frac{170 \times 240}{300} = 136$	170
Total	60	240	300

- To calculate the chi-square statistic you use the observed values from Table a and the expected values from Table c
- Formula is [(Observed – Expected)^2/Expected] for each cell of the table

Chi-Square Test: Example (cont.)

Table d. Expected values for exposure to tomatoes

	Salmonella?		
	Yes	No	Total
Tomatoes	$\frac{(41-26)^2}{26} = 8.7$	$\frac{(89-104)^2}{104} = 2.2$	130
No Tomatoes	$\frac{(19-34)^2}{34} = 6.6$	$\frac{(151-136)^2}{136} = 1.7$	170
Total	60	240	300

- The chi-square (χ2) for this example is 19.2
  - 8.7 + 2.2 + 6.6 + 1.7 = 19.2

Chi-Square Test

- What does the chi-square tell you?
- In general, the higher the chi-square value, the greater the likelihood there is a statistically significant difference between the two groups you are comparing
- To know for sure, you need to look up the **p-value** in a chi-square table

We will discuss p-values after a discussion of different types of chi-square tests

Types of Chi-Square Tests

- Many computer programs give different types of chi-square tests
- Each test is best suited to certain situations
- Most commonly calculated chi-square test is Pearson's chi-square
  - Use Pearson's chi-square for a fairly large sample (>100)

P-Values for Chi-Square Tests

- Using our hypothetical cruise ship *Salmonella* outbreak:
  - 32% of people who ate tomatoes got *Salmonella* as compared with 11% of people who did not eat tomatoes
- How do we know whether the difference between 32% and 11% is a "real" difference?
  - In other words, how do we know that our chi-square value (calculated as 19.2) indicates a statistically significant difference?
- The p-value is our indicator

P-Values for Chi-Square Tests

- Many statistical tests give both a numeric result (e.g. a chi-square value) and a p-value
- The p-value ranges between 0 and 1
- What does the p-value tell you?
  - The p-value is the probability of getting the result you got, assuming that the two groups you are comparing are actually *the same*

P-Values for Chi-Square Tests

- DF (degree of freedom) = (row - 1) * (col - 1)
- Here DF = (2-1)*(2-1) = 1
- Statistical Significance $\alpha = 0.05$
- Critical Value 3.841
- Since out test statistic Chi-Square (19.2) is larger than the critical value (3.841), we reject the null hypothesis of our test. We have sufficient evidence to say that our results are statistically significant at alpha level 0.05.

Critical values of chi-square (right tail)

Degrees of freedom (*df*)	Significance level (α) .99	.975	.95	.9	.1	.05	.025	.01
1	-------	0.001	0.004	0.016	2.706	$\boxed{3.841}$	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000

P-Values for Chi-Square Tests

- Start by assuming there is no difference in outcomes between the groups
- Look at the test statistic and p-value to see if they indicate otherwise
  - A low p-value means that (assuming the groups are the same) the probability of observing these results by chance is very small
    - Difference between the two groups is statistically significant
  - A high p-value means that the two groups were not that different
  - A p-value of 1 means that there was no difference between the two groups

P-Values for Chi-Square Tests

- Generally, if the p-value is less than 0.05, the difference observed is considered statistically significant, ie. the difference did not happen by chance
- You may use a number of statistical tests to obtain the p-value
  - Test used depends on type of data you have

Chi-square test of association

Spring 2023 Question Solution

A university conducted a study to examine the relationship between students' academic majors and their preferred study locations on campus. A sample of 400 students was randomly selected, and they were classified into two groups based on their majors: Business, and Engineering. The students were also asked to indicate their preferred study location: Library, Cafeteria, Outdoor Spaces, and Study Rooms. The observed frequencies for each major-study location combination are as follows:

	Library	Cafeteria	Outdoor Spaces	Study Rooms
Business	60	70	40	20
Engineering	70	30	50	90

a) Calculate the expected frequencies for each major-study location combination assuming that there is no association between majors and study locations.
b) Perform a chi-square test to determine whether there is a significant relationship between students' academic majors and their preferred study locations. Use a significance level of 0.05.
c) State your conclusion based on the chi-square test result.

Chi-square test of association

Solution:

To analyze the relationship between students' academic majors and their preferred study locations, you can perform a chi-square test of independence. Here are the steps to carry out this analysis:

Set up the null and alternative hypotheses:

- Null Hypothesis (H0): There is no association between students' academic majors and their preferred study locations.
- Alternative Hypothesis (H1): There is an association between students' academic majors and their preferred study locations.

Chi-square test of association

Calculate the row and column totals (marginal totals):

	Library	Cafeteria	Outdoor Spaces	Study Rooms	Row Total
Business	60	70	40	20	190
Engineering	70	30	50	90	240
Column Total	130	100	90	110	430

To calculate the expected frequency for each cell, we used the formula:
Expected Frequency = (Row Total * Column Total) / Grand Total

The expected frequencies:

	Library	Cafeteria	Outdoor Spaces	Study Rooms
Business (Expected)	57.44	44.19	39.77	48.60
Engineering (Expected)	72.86	55.81	50.23	61.40

Chi-square test of association

Calculate the chi-square statistic:

The chi-square statistic is calculated using the formula:

Chi-Square = Σ [(Observed Frequency - Expected Frequency)^2 / Expected Frequency]

1. Business and Library:
    1. Observed Frequency (O) = 60
    2. Expected Frequency (E) = 57.44
    3. Chi-Square = [(60 - 57.44)^2 / 57.44] ≈ 0.1141
2. Business and Cafeteria:
    1. Observed Frequency (O) = 70
    2. Expected Frequency (E) = 44.19
    3. Chi-Square = [(70 - 44.19)^2 / 44.19] ≈ 15.074
3. Business and Outdoor Spaces:
    1. Observed Frequency (O) = 40
    2. Expected Frequency (E) = 39.77
    3. Chi-Square = [(40 - 39.77)^2 / 39.77] ≈ 0.0013
4. Business and Study Rooms:
    1. Observed Frequency (O) = 20
    2. Expected Frequency (E) = 48.60
    3. Chi-Square = [(20 - 48.60)^2 / 48.60] ≈ 16.83

1. Engineering and Library:
    1. Observed Frequency (O) = 70
    2. Expected Frequency (E) = 72.56
    3. Chi-Square = [(70 - 72.56)^2 / 72.56] ≈ 0.0903
2. Engineering and Cafeteria:
    1. Observed Frequency (O) = 30
    2. Expected Frequency (E) = 55.81
    3. Chi-Square = [(30 - 55.81)^2 / 55.81] ≈ 11.936
3. Engineering and Outdoor Spaces:
    1. Observed Frequency (O) = 50
    2. Expected Frequency (E) = 50.23
    3. Chi-Square = [(50 - 50.23)^2 / 50.23] ≈ 0.0011
4. Engineering and Study Rooms:
    1. Observed Frequency (O) = 90
    2. Expected Frequency (E) = 61.40
    3. Chi-Square = [(90 - 61.40)^2 / 61.40] ≈ 13.3218

Sum up all these individual chi-square values to get the total chi-square value:

Total Calculated Chi-Square = 0.1141 + 15.074 + 0.0013 + 16.83 + 0.0903 + 11.936 + 0.0011 + 13.3218 ≈ 57.256 (rounded to two decimal places)

P-Values for Chi-Square Tests

Determine the degrees of freedom (df):

Degrees of Freedom (df) = (Number of Rows - 1) * (Number of Columns - 1)

In this case, df = (2 - 1) * (4 - 1) = 1 * 3 = 3

Critical values of chi-square (right tail)

Degrees of freedom (*df*)	Significance level (α) .99	.975	.95	.9	.1	.05	.025	.01
1	-------	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	$\boxed{7.815}$	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000

P-Values for Chi-Square Tests

With df = 3 (as calculated previously) and a significance level of 0.05, you can compare this total calculated chi-square value to the critical chi-square value (approximately 7.815). Total Calculated Chi-Square (57.256) > Critical Chi-Square Value (7.815)
Since the total calculated chi-square value is greater than the critical chi-square value, you would reject the null hypothesis. This indicates that there is a significant association between students' academic majors and their preferred study locations.

Critical values of chi-square (right tail)

Degrees of freedom (*df*)	Significance level (α)							
	.99	.975	.95	.9	.1	.05	.025	.01
1	-------	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	$\boxed{7.815}$	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000