Topic: Chi-square tests

Question: A university is graduating $5,000$ seniors and wants to know if their graduation rate is affected by student involvement in extracurriculars. They randomly sampled seniors as they graduated (or failed to graduate), and asked them about their extracurricular activities. What can the university conclude using a chi-square test at 95% confidence?

	Number of extracurriculars			
	0-2	3-5	6+	Totals
Graduating	221	118	41	380
Not graduating	11	75	34	120
Totals	232	193	75	500

Answer choices:

A Number of extracurriculars doesn't affect graduation rate

B Number of extracurriculars affects graduation rate

C The university can't use a chi-square test because their sample doesn't meet the large counts condition

D The university can't use a chi-square test because their sample doesn't meet the independence condition

Solution: B

Start by computing expected values.

Expected Graduating/$0-2$: $(380 \cdot 232)/500 = 176.32$

Expected Graduating/$3-5$: $(380 \cdot 193)/500 = 146.68$

Expected Graduating/$6+$: $(380 \cdot 75)/500 = 57.00$

Expected Not graduating/$0-2$: $(120 \cdot 232)/500 = 55.68$

Expected Not graduating/$3-5$: $(120 \cdot 193)/500 = 46.32$

Expected Not graduating/$6+$: $(120 \cdot 75)/500 = 18$

Then fill in the table.

	Number of extracurriculars			
	0-2	**3-5**	**6+**	**Totals**
Graduating	221 (176.32)	118 (146.68)	41 (57.00)	380
Not graduating	11 (55.68)	75 (46.32)	34 (18.00)	120
Totals	232	193	75	500

Now we'll check our sampling conditions. The problem told us that we took a random sample, and all of our expected values are at least 5, so we've met the random sampling and large counts conditions. And even though we're sampling without replacement, there are $5{,}000$ students in the graduating class, and we're sampling 10% of them, so we've met the independence condition as well.

We'll state the null hypothesis.

H_0: Graduation rate is not affected by number of extracurriculars.

H_a: Graduation rate is affected by number of extracurriculars.

Calculate χ^2.

$$\chi^2 = \frac{(221 - 176.32)^2}{176.32} + \frac{(118 - 146.68)^2}{146.68} + \frac{(41 - 57)^2}{57}$$

$$+ \frac{(11 - 55.68)^2}{55.68} + \frac{(75 - 46.32)^2}{46.32} + \frac{(34 - 18)^2}{18}$$

$$\chi^2 \approx 11.32 + 5.61 + 4.49 + 35.85 + 17.76 + 14.22$$

$$\chi^2 \approx 89.25$$

The degrees of freedom are

df $= ($number of rows $- 1)($number of columns $- 1)$

df $= (2 - 1)(3 - 1)$

df $= (1)(2)$

df $= 2$

With df $= 2$ and $\chi^2 \approx 89.25$, the χ^2-table gives

	Upper-tail probability p											
df	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.0025**	**0.001**	**0.0005**
1	1.32	1.64	2.07	2.71	3.81	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

We're off the chart on the right, which means we will definitely exceed the alpha level $\alpha = 0.05$. Therefore, the university will reject the null hypothesis, and conclude that number of extracurricular activities affects graduation rate.

Topic: Chi-square tests

Question: A restaurant wants to know whether a diner's choice to order dessert is affected by whether or not they ordered an appetizer. On an evening when they served $2{,}000$ diners, they randomly sampled diners as they finished their meals, and recorded whether they had ordered an appetizer and/or dessert. What can the restaurant conclude using a chi-square test at 95% confidence?

	Dessert	No dessert	Totals
Appetizer	70	42	112
No appetizer	50	38	88
Totals	120	80	200

Answer choices:

A Ordering an appetizer doesn't affect the dessert order

B Ordering an appetizer affects the dessert order

C The restaurant can't use a chi-square test because their sample doesn't meet the large counts condition

D The restaurant can't use a chi-square test because their sample doesn't meet the independence condition

Solution: A

Start by computing expected values.

Expected Appetizer/Dessert: $(112 \cdot 120)/200 = 67.2$

Expected Appetizer/No dessert: $(112 \cdot 80)/200 = 44.8$

Expected No appetizer/Dessert: $(88 \cdot 120)/200 = 52.8$

Expected No appetizer/No dessert: $(88 \cdot 80)/200 = 35.2$

Then fill in the table.

	Dessert	No dessert	Totals
Appetizer	70 (67.2)	42 (44.8)	112
No appetizer	50 (52.8)	38 (35.2)	88
Totals	120	80	200

Now we'll check our sampling conditions. The problem told us that we took a random sample, and all of our expected values are at least 5, so we've met the random sampling and large counts conditions. And even though we're sampling without replacement, there are $2{,}000$ diners, and we're sampling 10% of them, so we've met the independence condition as well.

We'll state the null hypothesis.

H_0: Whether or not a diner orders dessert is not affected by whether or not they ordered an appetizer.

H_a: Whether or not a diner orders dessert is affected by whether or not they ordered an appetizer.

Calculate χ^2.

$$\chi^2 = \frac{(70-67.2)^2}{67.2} + \frac{(42-44.8)^2}{44.8} + \frac{(50-52.8)^2}{52.8} + \frac{(38-35.2)^2}{35.2}$$

$$\chi^2 \approx 0.12 + 0.18 + 0.15 + 0.22$$

$$\chi^2 \approx 0.67$$

The degrees of freedom are

$$\text{df} = (\text{number of rows} - 1)(\text{number of columns} - 1)$$

$$\text{df} = (2-1)(2-1)$$

$$\text{df} = (1)(1)$$

$$\text{df} = 1$$

With df $= 1$ and $\chi^2 \approx 0.67$, the χ^2-table gives

	Upper-tail probability p											
df	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.0025**	**0.001**	**0.0005**
1	**1.32**	1.64	2.07	2.71	3.81	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

We're off the chart on the left, which means we will definitely not exceed the alpha level $\alpha = 0.05$. Therefore, the restaurant will fail to reject the null hypothesis, and conclude whether or not a diner orders an appetizer does not affect whether or not they order dessert.

Topic: Chi-square tests

Question: A company wants to know if the number of sick days taken by its employees is affected by quarter. They recorded sick days taken each quarter. What can the company conclude using a chi-square test at 95% confidence?

Quarter	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Total
Sick days	44	49	45	42	180

Answer choices:

A Sick days taken is not affected by quarter

B Sick days taken is affected by quarter

C The company can't use a chi-square test because their sample doesn't meet the large counts condition

D The company can't use a chi-square test because their sample doesn't meet the independence condition

Solution: A

With 180 total sick days, the expected number of sick days in each quarter would be $180/4 = 45$.

Quarter	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Total
Sick days	44	49	45	42	180
Expected	45	45	45	45	180

We'll state the null hypothesis.

H_0: Number of sick days taken is not affected by quarter.

H_a: Number of sick days taken is affected by quarter.

Calculate χ^2.

$$\chi^2 = \frac{(44-45)^2}{45} + \frac{(49-45)^2}{45} + \frac{(45-45)^2}{45} + \frac{(42-45)^2}{45}$$

$$\chi^2 = \frac{1}{45} + \frac{16}{45} + \frac{0}{45} + \frac{9}{45}$$

$$\chi^2 \approx 0.02 + 0.36 + 0.00 + 0.20$$

$$\chi^2 \approx 0.58$$

The degrees of freedom are $n - 1 = 4 - 1 = 3$. With df $= 3$ and $\chi^2 = 0.58$, the χ^2-table gives

	Upper-tail probability p											
df	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.0025**	**0.001**	**0.0005**
1	1.32	1.64	2.07	2.71	3.81	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

We're off the chart on the left, which means we will definitely not exceed the alpha level $\alpha = 0.05$. Therefore, the company will fail to reject the null hypothesis, and conclude that number of sick days taken is not affected by quarter.