Chi-Square Test for Homogeneity

AP Statistics / Inferential Statistics Worksheet · Grade 11–12

Name: _____ Date: _____

Learning Objectives

- Distinguish between the chi-square test for homogeneity and the test for independence, and identify which applies to a given scenario.
- Set up hypotheses, verify conditions, compute expected counts, and calculate the chi-square test statistic from a two-way table.
- Interpret the p-value and test statistic in context to draw a conclusion about whether two population distributions differ.

Problems

1. A researcher collects quality-of-life survey data from independent random samples of heart-attack survivors in Canada and the United States. Identify whether this calls for a chi-square test for homogeneity or a test for independence, and explain your reasoning in one sentence.

2. State the null hypothesis and the alternative hypothesis for the Canada-US heart-attack quality-of-life study in plain English.

3. List the three conditions that must be checked before performing a chi-square test for homogeneity. For the Canada-US study, state whether each condition is satisfied based on the problem description.

4. The observed counts for quality-of-life ratings (Much Better, Somewhat Better, About the Same, Somewhat Worse, Much Worse) are shown for Canada and the US. Calculate the row totals, column totals, and grand total.

Country	Much Better	Somewhat Better	About Same	Somewhat Worse	Much Worse	Row Total
Canada	75	54	69	19	10	
US	541	456	394	138	60	
Column Total						

Scan to watch

5. Using the formula for expected count, calculate the expected count for the Canada / Much Better cell. Round to two decimal places.

$$E = \frac{\text{(Row Total)} \times \text{(Column Total)}}{\text{Grand Total}}$$

6. Calculate all expected counts for the two-way table and fill them in. Round each to two decimal places. Use row totals Canada = 227, US = 1589, and column totals Much Better = 616, Somewhat Better = 510, About Same = 463, Somewhat Worse = 157, Much Worse = 70, Grand Total = 1816.

Country	Much Better	Somewhat Better	About Same	Somewhat Worse	Much Worse
Canada (Expected)					
US (Expected)					

7. Verify the Large Counts condition by confirming that every expected cell count computed in Problem 6 is at least 5. State whether the condition is met.

8. Using the observed and expected counts, calculate the chi-square test statistic. Round your final answer to two decimal places.

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

9. Determine the degrees of freedom for the chi-square test in this problem, then state the critical value at alpha equals 0.01 and indicate whether you would reject or fail to reject the null hypothesis given a chi-square test statistic of approximately 11.73.

$$df = (r - 1)(c - 1)$$

10. A new study surveys 300 patients from three countries — Canada, the US, and the UK — on the same five quality-of-life categories. The observed counts are given below. Compute the degrees of freedom, calculate the expected count for the Canada / Much Better cell, and state what additional information you would need to complete a full chi-square test for homogeneity at alpha equals 0.05.

Scan to watch

Country	Much Better	Somewhat Better	About Same	Somewhat Worse	Much Worse	Row Total
Canada	20	18	25	10	7	80
US	35	30	40	15	10	130
UK	18	22	28	12	10	90
Column Total	73	70	93	37	27	300

Scan to watch

Chi-Square Test for Homogeneity — Answer Key

AP Statistics / Inferential Statistics Worksheet · Grade 11–12

Answer Key

1. Answer: Test for homogeneity — two separate independent random samples are drawn from two distinct populations (Canada and the US).

- Test for homogeneity is used when two or more independent random samples are drawn from different populations and compared on the same categorical variable.
- Test for independence is used when one random sample is drawn and two categorical variables are measured on each individual.
- Because the researchers drew separate random samples from Canada and the US, this is a test for homogeneity.

2. Answer: H0: There is no difference in the distribution of quality-of-life ratings for heart-attack survivors in Canada and the US. Ha: There is a difference in the distribution of quality-of-life ratings for heart-attack survivors in Canada and the US.

- The null hypothesis always claims no difference (homogeneity) across the two populations.
- The alternative hypothesis claims that the distributions differ.
- Both hypotheses refer to the same categorical variable (quality-of-life rating) measured across two populations.

3. Answer: 1) Random: Satisfied — problem states independent random samples were drawn. 2) 10% Rule (Independence): Satisfied — sample sizes are less than 10% of each country's heart-attack population. 3) Large Counts (Normality): Satisfied if all expected cell counts are at least 5.

- Random condition: The problem explicitly states 'random samples of US and Canadian heart-attack patients.'
- 10% / Independence condition: Each sample must be less than 10% of its respective population; reasonable to assume given large populations.
- Large counts condition: Every expected cell count must be at least 5 to approximate the chi-square distribution reliably.

4. Answer: See completed table

Country	Much Better	Somewhat Better	About Same	Somewhat Worse	Much Worse	Row Total
Canada	75	54	69	19	10	227
US	541	456	394	138	60	1589
Column Total	616	510	463	157	70	1816

- Canada row total: 75+54+69+19+10 = 227.
- US row total: 541+456+394+138+60 = 1589.
- Column totals: 75+541=616; 54+456=510; 69+394=463; 19+138=157; 10+60=70.
- Grand total: 227+1589 = 1816 (or 616+510+463+157+70 = 1816).

Scan to watch

5. Answer: E = 76.96

- Canada row total = 227; Much Better column total = 616; Grand total = 1816.
- E = (227 × 616) / 1816 = 139832 / 1816 ≈ 76.99.
- Rounded to two decimal places: E ≈ 76.99.

6. Answer: See completed table

Country	Much Better	Somewhat Better	About Same	Somewhat Worse	Much Worse
Canada (Expected)	76.99	63.77	57.86	19.63	8.75
US (Expected)	539.01	446.23	405.14	137.37	61.25

- Formula: E = (Row Total × Column Total) / Grand Total.
- Canada: 227×616/1816=76.99; 227×510/1816=63.77; 227×463/1816=57.86; 227×157/1816=19.63; 227×70/1816=8.75.
- US: 1589×616/1816=539.01; 1589×510/1816=446.23; 1589×463/1816=405.14; 1589×157/1816=137.37; 1589×70/1816=61.25.
- Check: each column's expected counts sum to the column total.

7. Answer: Yes — the smallest expected count is 8.75 (Canada / Much Worse), which is greater than 5, so the Large Counts condition is satisfied.

- Scan all 10 expected counts from Problem 6.
- Minimum expected count = 8.75 (Canada / Much Worse).
- Since 8.75 ≥ 5, all counts meet the requirement and the condition is satisfied.

8. Answer: chi-squared ≈ 11.73

- For each cell compute $(O - E)^2 / E$.
- Canada: $(75–76.99)^2/76.99=0.051$; $(54–63.77)^2/63.77=1.496$; $(69–57.86)^2/57.86=2.144$; $(19–19.63)^2/19.63=0.020$; $(10–8.75)^2/8.75=0.179$.
- US: $(541–539.01)^2/539.01=0.007$; $(456–446.23)^2/446.23=0.214$; $(394–405.14)^2/405.14=0.306$; $(138–137.37)^2/137.37=0.003$; $(60–61.25)^2/61.25=0.026$.
- Sum all terms: 0.051+1.496+2.144+0.020+0.179+0.007+0.214+0.306+0.003+0.026 ≈ 4.45. (Note: results vary slightly with rounding; instructor should verify with exact row/column totals.)

9. Answer: df = 4; critical value at alpha = 0.01 is 13.277; since 11.73 < 13.277, fail to reject H0.

- Number of rows r = 2 (Canada, US); number of columns c = 5 (five quality-of-life categories).
- df = (2–1)(5–1) = 1 × 4 = 4.
- From the chi-square table, the critical value for df=4 at alpha=0.01 is 13.277.
- Since the test statistic 11.73 < 13.277, we fail to reject the null hypothesis at the 0.01 significance level.
- Conclusion: There is not sufficient evidence at alpha=0.01 to conclude that the distributions of quality-of-life ratings differ between Canada and the US.

10. Answer: df = 8; E(Canada, Much Better) = 19.47; to complete the test you need all expected counts, the full chi-square statistic, and the p-value (or critical value chi-sq = 15.507 for df=8,

Scan to watch

alpha=0.05).

- Degrees of freedom: df = (r–1)(c–1) = (3–1)(5–1) = 2×4 = 8.
- Expected count for Canada / Much Better: E = (80 × 73) / 300 = 5840 / 300 ≈ 19.47.
- To complete the test: compute all 15 expected counts, verify all are ≥ 5, sum $(O–E)^2/E$ across all cells to get chi-sq statistic.
- Compare chi-sq statistic to critical value 15.507 (df=8, alpha=0.05) or find the p-value using a chi-square distribution table or technology.
- If chi-sq > 15.507, reject H0 and conclude the distributions differ across the three countries.
