

Chi-Square Tests for Two-Way Tables

Grade 12 – AP Statistics math · Test Answer Key · Standards: AP Statistics Unit 3 (2026-27)

The scoring guide shows how points are earned. Full credit needs hypotheses in context, the expected-count condition, a correct statistic with df and p-value, and a conclusion tied to the p-value.

Standards: AP Statistics Unit 3 (2026-27) · **Level:** Grade 12 AP Statistics

Question 1 Choosing the Procedure (2 points)

(a) A chi-square test for homogeneity (two independent samples, one categorical variable). (b) H_0 : the distribution of travel method is the same in both districts; H_a : the distribution is not the same. $df = (2 - 1)(4 - 1) = 3$.

Homogeneity; df = 3

Scoring: 1 point for the test, 1 point for the hypotheses and df.

Question 2 Expected Counts (2 points)

Row totals 100 and 100; column totals 72, 76, 52; total 200. Female and drama: $100(76)/200 = 38.0$. Male and action: $100(52)/200 = 26.0$.

38.0 and 26.0

Scoring: 1 point for each expected count.

Question 3 Test of Independence (3 points)

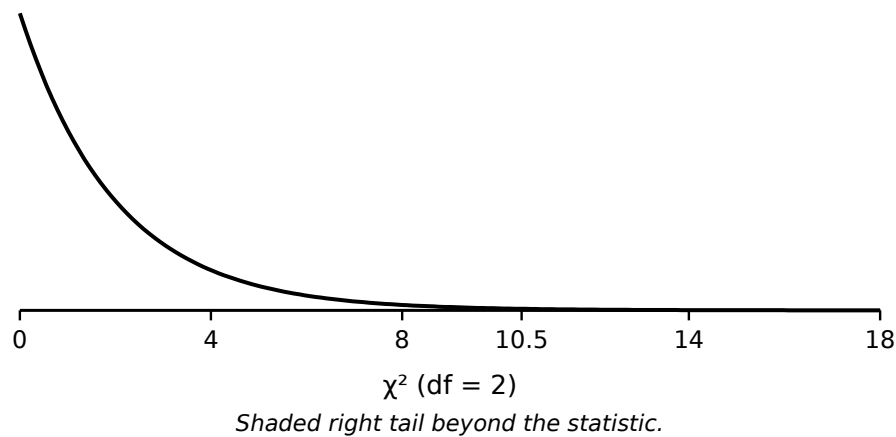
(a) H_0 : sex and favorite movie type are independent; H_a : they are associated. Random sample, 200 is under 10% of the population, and the expected counts 36.0, 38.0, 26.0, 36.0, 38.0, 26.0 are all at least 5. (b) $\chi^2 = 6.535$, $df = 2$, $p\text{-value} = 0.0381$. (c) Since $0.0381 < 0.05$, there is convincing evidence of an association between sex and favorite type of movie.

$\chi^2 \approx 6.54$, $df = 2$, $p\text{-value} \approx 0.0381$; reject H_0

Scoring: 1 point each part.

Question 4 Test of Homogeneity (3 points)

H_0 : the distribution of outcomes is the same for the three diets; H_a : it is not. Random assignment; expected counts 22.67, 27.33, 22.67, 27.33, 22.67, 27.33, all at least 5. $\chi^2 = 10.545$, $df = 2$, $p\text{-value} = 0.0051$. Since $0.0051 < 0.05$, reject H_0 : there is convincing evidence that the outcome distributions differ, and random assignment supports a causal conclusion for these volunteers.



$\chi^2 \approx 10.55$, $df = 2$, $p\text{-value} \approx 0.0051$; reject H_0

Scoring: 1 point for hypotheses and conditions, 1 point for the statistic, df and $p\text{-value}$, 1 point for the conclusion in context.

Question 5 Checking Conditions (2 points)

Being random is necessary but not sufficient. The condition on expected counts requires every one to be at least 5, and 4.0 is less than 5. The conditions are not met, so the chi-square test is not appropriate unless categories are combined or the sample size is increased.

Not met: one expected count (4.0) is below 5

Scoring: 1 point for checking expected counts, 1 point for the conclusion.

Question 6 Interpreting a p-Value (2 points)

(a) $p\text{-value} = P(\chi^2 \geq 8.2) = 0.0166$. (b) If the distribution of favorite season were the same in the two towns, the probability of getting samples whose distributions differ as much as these (or more) is about 0.017.

$p\text{-value} \approx 0.0166$

Scoring: 1 point each part.

Question 7 Observational or Experimental (3 points)

(a) Study 1: convincing evidence of an association between tracker use and meeting the goal for adults in the population; no causal conclusion. Study 2: convincing evidence that using a tracker affects whether volunteers meet the goal (a causal conclusion), though only for volunteers like those in the study. (b) Study 2 used random assignment, which balances other variables; Study 1 is observational, so confounding variables are possible. Study 1, with a random sample, can be generalized to the population.

Study 1: association, generalizable; Study 2: cause and effect, for the volunteers

Scoring: 1 point for each study conclusion, 1 point for the explanation.

Question 8 Contributions to Chi-Square (3 points)

(a) The two largest contributions are service A for 30 or older (3.65) and service A for under 30 (3.62). (b) Under-30 customers prefer A more than expected (58 observed vs 45.2 expected), while customers 30 or older prefer A less than expected (32 vs 44.8). (c) Service B contributes very little (0.29 and 0.29), so it shows the least association with age.

(a) A for both age groups; (b) under 30 over-represented for A; (c) service B

Scoring: 1 point each part.