Week 26

AP Statistics

Homework bundle for this week

本周作业合集

exercises.lol

Contents 目录

8. Inference for Categorical Data: Chi-Square

Starter Quiz · 课前小测

AP Statistics Name: ____________ Score: ______

1. Chi-square tests are designed for which kind of data?
   - ☐ categorical counts
   - ☐ measured means
   - ☐ a single proportion only
   - ☐ continuous data

2. A fair die is rolled 60 times. What is the expected count for each face ($60 \times 1/6$)?

   ____

3. One category has observed 8, expected 10. Compute $(\text{observed} - \text{expected})^2/\text{expected}$.

   ____

4. Row total 60, column total 50, grand total 100. Find the expected count ($\text{row} \cdot \text{col}/\text{grand}$).

   ____

5. You sample 100 boys and 100 girls SEPARATELY and compare favorite subject. This is a test of...
   - ☐ homogeneity
   - ☐ independence
   - ☐ goodness-of-fit
   - ☐ a mean

6. With $\chi^2 = 6.0$ at df $= 1$, the p-value ≈ 0.014. At $\alpha = 0.05$, the decision is...
   - ☐ reject H0
   - ☐ fail to reject H0
   - ☐ accept H0
   - ☐ double the p-value

7. One categorical variable checked against a claimed distribution (one sample) calls for a...
   - ☐ goodness-of-fit test
   - ☐ test of independence
   - ☐ two-sample t-test
   - ☐ test of homogeneity

8. A goodness-of-fit test involves how many c ical variables?

- ☐ one
- ☐ two
- ☐ three
- ☐ none

9. For a goodness-of-fit test with 6 categories, what are the degrees of freedom?

    ____

10. For a 2×3 two-way table, what are the degrees of freedom $(r-1)(c-1)$?

    ____

AP Statistics

1. **categorical counts**

   Chi-square compares counts in categories.

2. **10**

   $60 \times 1/6 = 10$.

3. **0.4**

   $(8-10)^2/10 = 4/10 = 0.4$.

4. **30**

   $60 \times 50 \ / \ 100 = 30$.

5. **homogeneity**

   Several separate samples compared on one variable → homogeneity.

6. **reject H0**

   $0.014 \leq 0.05 \rightarrow$ reject H0.

7. **goodness-of-fit test**

   One variable vs. a distribution → goodness-of-fit.

8. **one**

   GOF checks one variable against a claimed distribution.

9. **5**

   df = categories − 1 = 6 − 1 = 5.

10. **2**

    $(2-1)(3-1) = 1\times2 = 2$.

8.1 Introducing Statistics: Are My Results Unexpected?

Name: ____________ Class: ________ Date: ________ **Total: 8 marks**

KEY WORDS chi-square 卡方 • expected counts 期望计数 • observed counts 观测计数

Objective

Build the skills to answer exam questions on **whether results are unexpected**.

You must be able to:

- explain that a **chi-square** 卡方 test compares **observed** counts to **expected** counts
- describe **observed counts** 观测计数 and **expected counts** 期望计数
- recognise that chi-square tests are for **categorical** data

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Chi-square tests

A **chi-square** test judges whether **observed** counts differ from what we would **expect** if a claim were true.

- **Observed counts** — what the data actually show in each category.
- **Expected counts** — what we would predict under the null hypothesis.

Large gaps between observed and expected point to a real effect. Chi-square tests are used for **categorical** data (counts).

■ A worked comparison

A die is rolled 60 times. If it is fair, each face is **expected** 10 times. Suppose a six is **observed** 18 times: the gap $(18 - 10)$ is large, hinting the die is not fair. Chi-square adds up these gaps across all six faces to give one number.

2 Practice

2.1 State what a chi-square test compares. [1]

2.2 State what “observed counts” are. [1]

2.3 State what “expected counts” are. [1]

3 Exam-style questions

3.1 A chi-square test compares observed counts to [1]

- **A** means
- **B** expected counts
- **C** slopes
- **D** proportions directly

3.2 Chi-square tests are used for _________ data. [1]

- **A** quantitative
- **B** categorical
- **C** continuous
- **D** paired

3.3 A die is rolled 60 times to test whether it is fair.

(a) Name the kind of data collected. [1]

(b) State the expected count for each face. [1]

(c) Name the test comparing observed to expected counts. [1]

Solutions

2.1 observed counts with the counts expected under the null hypothesis.

2.2 the counts that actually occur in each category of the data.

2.3 the counts predicted in each category if the null hypothesis were true.

3.1 **B**.

3.2 **B**.

3.3 (a) categorical (counts of each face). (b) $60 \div 6 = 10$. (c) a chi-square (goodness-of-fit) test.

AP STATISTICS • EXERCISE SHEET

8.2 Setting Up a Chi-Square Goodness of Fit Test

Name: ______________ Class: __________ Date: ____________ **Total: 9 marks**

KEY WORDS chi-square 卡方 • degrees of freedom 自由度

Objective

Build the skills to answer exam questions on **setting up a chi-square goodness-of-fit test**.

You must be able to:

- state the hypotheses for a goodness-of-fit test
- calculate **expected counts** as $n \cdot p_i$
- write the statistic $\chi^2 = \sum \dfrac{(O-E)^2}{E}$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Hypotheses

- **Null** H_0: the distribution **matches** the claimed proportions.
- **Alternative** H_a: at least one proportion is different.

■ Expected counts and the statistic

The expected count in a category is $E = n \cdot p_i$. The test statistic is

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

with degrees of freedom = (number of categories) − 1.

2 Practice

2.1 State the null hypothesis for a goodness-of-fit test. [1]

2.2 Write the formula for an expected count. [1]

2.3 For $n = 200$ and a claimed proportion of 0.25, find the expected count. [2]

3 Exam-style questions

3.1 The expected count for a category is [1]

- **A** n
- **B** $n \cdot p_i$
- **C** $O - E$
- **D** $\sqrt{n}$

3.2 The chi-square statistic is [1]

- **A** $\sum \frac{(O-E)^2}{E}$
- **B** $\sum (O - E)$
- **C** $\sum \frac{O}{E}$
- **D** $\sqrt{O - E}$

3.3 A fair die is rolled 120 times.

(a) State the expected count for each face. [2]

(b) Write the chi-square statistic formula. [1]

Solutions

2.1 the distribution matches the claimed proportions.

2.2 $E = n \cdot p_i$.

2.3 $E = 200 \times 0.25 = 50$.

3.1 **B**.

3.2 **A**.

3.3 (a) $120 \times \dfrac{1}{6} = 20$. (b) $\chi^2 = \sum \dfrac{(O-E)^2}{E}$.

8.3 Carrying Out a Chi-Square Test for Goodness of Fit

Name: ______________ Class: __________ Date: ____________ **Total: 8 marks**

KEY WORDS degrees of freedom 自由度 • chi-square 卡方

Objective

Build the skills to answer exam questions on **carrying out a chi-square goodness-of-fit test**.

You must be able to:

- compute $\chi^2 = \sum \frac{(O-E)^2}{E}$ with degrees of freedom $= k - 1$
- find the **p-value** and reach a decision
- state a conclusion in context

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ The procedure

1. State the hypotheses and check conditions (all expected counts ≥ 5).
2. Compute $\chi^2 = \sum \frac{(O-E)^2}{E}$, with df $= k - 1$ (categories minus 1).
3. Find the **p-value** from the chi-square distribution.
4. Compare to α and conclude **in context**.

A **large** χ^2 gives a **small** p-value — stronger evidence against the claimed distribution.

■ A worked statistic

For observed $(18, 12, 30)$ and expected $(20, 20, 20)$,

$$\chi^2 = \frac{(18-20)^2}{20} + \frac{(12-20)^2}{20} + \frac{(30-20)^2}{20} = 0.2 + 3.2 + 5.0 = 8.4,$$

with df $= 3 - 1 = 2$. This gives a p-value $\approx 0.015 < 0.05$ – reject the claimed distribution.

2 Practice

2.1 State the degrees of freedom for a goodness-of-fit test with k categories. [1]

2.2 State what a large chi-square value suggests. [1]

2.3 A goodness-of-fit test gives a p-value of 0.03 at $\alpha = 0.05$. State the decision. [1]

3 Exam-style questions

3.1 The degrees of freedom for a goodness-of-fit test with 6 categories is [1]

- **A** 6
- **B** 5
- **C** 4
- **D** 1

3.2 A large chi-square statistic gives a __________ p-value. [1]

- **A** large
- **B** small
- **C** zero
- **D** negative

3.3 A goodness-of-fit test of a die (6 faces) gives $\chi^2 = 15$ with df $= 5$ and p-value 0.01.

(a) State the degrees of freedom. [1]

(b) State the decision at $\alpha = 0.05$. [1]

(c) State the conclusion in context. [1]

Solutions

2.1 $k - 1$.

2.2 a large difference between observed and expected counts (evidence against H_0).

2.3 reject H_0 $(0.03 < 0.05)$.

3.1 **B**.

3.2 **B**.

3.3 (a) 5. (b) reject H_0. (c) there is convincing evidence the die does not follow the claimed (fair) distribution.

AP STATISTICS • EXERCISE SHEET

8.4 Expected Counts in Two-Way Tables

Name: ____________ Class: ________ Date: __________ **Total: 9 marks**

Objective

Build the skills to answer exam questions on **expected counts in two-way tables**.

You must be able to:

- calculate an expected count as $E = \dfrac{\text{(row total)(column total)}}{\text{grand total}}$
- explain that expected counts assume **no association** between the variables

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Expected counts

For each cell of a two-way table, if the two variables were **not associated** (the null hypothesis):

$$E = \frac{\text{(row total)(column total)}}{\text{grand total}}.$$

Example

A cell with row total 40, column total 50, and grand total 200: $E = \dfrac{40 \times 50}{200} = 10$.

2 Practice

2.1 Write the formula for an expected count in a two-way table. [1]

2.2 A cell has row total 60, column total 80, grand total 240. Find the expected count. [2]

2.3 State what expected counts assume. [1]

3 Exam-style questions

3.1 The expected count in a two-way table cell is [1]

- **A** $\dfrac{(\text{row})(\text{column})}{\text{grand total}}$
- **B** row + column
- **C** the observed count
- **D** grand total $\div 2$

3.2 Expected counts are computed assuming [1]

- **A** an association
- **B** no association (the null hypothesis)
- **C** bias
- **D** unequal sample sizes

3.3 In a table, a cell has row total 50, column total 100, grand total 250.

(a) Find the expected count. [2]

(b) State the assumption behind it. [1]

Solutions

2.1 $E = \dfrac{\text{(row total)(column total)}}{\text{grand total}}$.

2.2 $E = \dfrac{60 \times 80}{240} = 20$.

2.3 that there is no association between the two variables.

3.1 **A**.

3.2 **B**.

3.3 (a) $E = \dfrac{50 \times 100}{250} = 20$. (b) no association between the variables.

8.5 Setting Up a Chi-Square Test for Homogeneity or Independence

Name: ______________ Class: __________ Date: ____________ **Total: 10 marks**

KEY WORDS independence 独立性 • homogeneity 同质性 • chi-square 卡方 • degrees of freedom 自由度

Objective

Build the skills to answer exam questions on **setting up a chi-square test for homogeneity or independence**.

You must be able to:

- distinguish a test of **homogeneity** 同质性 from a test of **independence** 独立性
- calculate the degrees of freedom $(r-1)(c-1)$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Homogeneity vs independence

- **Homogeneity** — compares the distribution of one categorical variable across **several groups** (several samples).
- **Independence** — tests whether **two** categorical variables are associated within **one** population (one sample).

Both use the same statistic $\chi^2 = \sum \frac{(O-E)^2}{E}$ and the same degrees of freedom:

$$\text{df} = (r-1)(c-1).$$

Which one, and its df

A 3×4 table (3 rows, 4 columns) has $\text{df} = (3-1)(4-1) = 6$. If the rows are **separate samples** (say, three schools), it is a test of **homogeneity**; if the table comes from **one sample** cross-classified by two variables, it is a test of **independence**.

2 Practice

2.1 State the difference between a test of homogeneity and a test of independence. [2]

2.2 Write the degrees of freedom for an $r \times c$ table. [1]

2.3 For a 3×4 table, find the degrees of freedom. [2]

3 Exam-style questions

3.1 The degrees of freedom for a two-way chi-square with r rows and c columns is [1]

- **A** rc
- **B** $(r-1)(c-1)$
- **C** $r+c$
- **D** $rc-1$

3.2 A test of whether two categorical variables are associated in one population is a test of [1]

- **A** homogeneity
- **B** independence
- **C** goodness of fit
- **D** means

3.3 A 2×3 table is analysed.

(a) Find the degrees of freedom. [2]

(b) Name the test that compares distributions across several groups. [1]

Solutions

2.1 homogeneity compares one variable's distribution across several groups; independence tests whether two variables are associated in one population.

2.2 $(r-1)(c-1)$.

2.3 $(3-1)(4-1) = 2 \times 3 = 6$.

3.1 **B**.

3.2 **B**.

3.3 (a) $(2-1)(3-1) = 2$. (b) a test of homogeneity.

5. A researcher is investigating whether there is an association among professional athletes between age-group (in years) and type of sport played (basketball, football, baseball). The age-group and type of sport played for all 4,193 professional athletes in these sports for a recent year are given in the two-way table.

Age-Group by Type of Sport Played

Age (years)	Basketball	Football	Baseball	Total
Age < 25	232	807	259	1,298
$25 \leq$ Age < 30	175	1,326	620	2,121
$30 \leq$ Age < 35	90	287	276	653
$35 \leq$ Age	19	41	61	121
Total	516	2,461	1,216	4,193

Part A

Consider the two-way table.

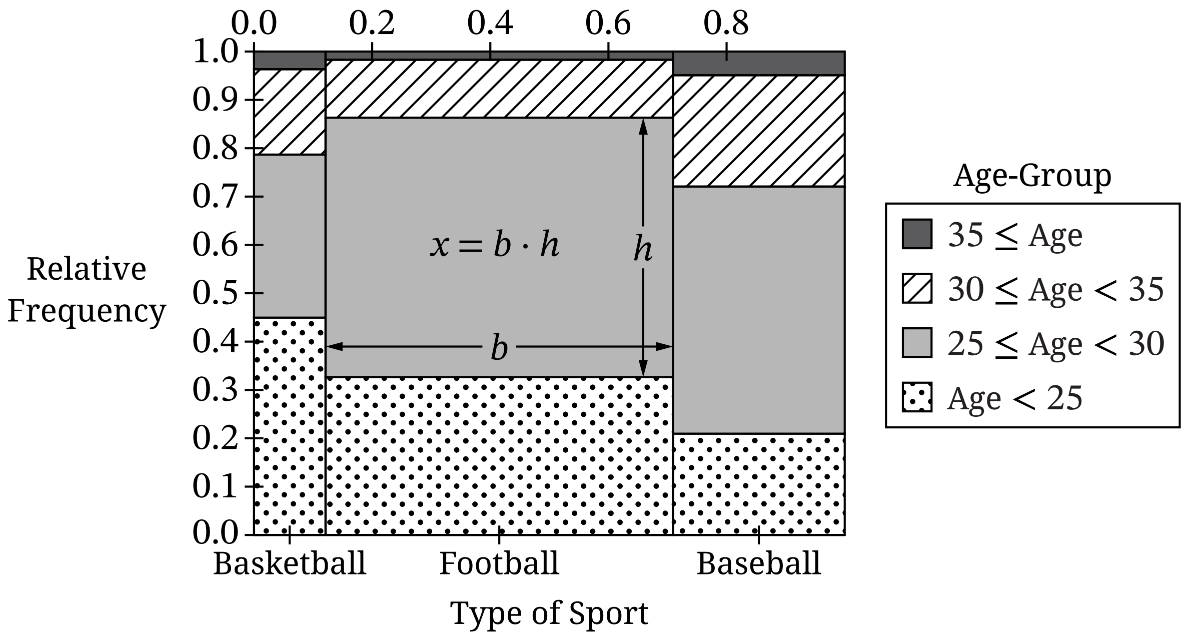
i. What is the probability a randomly selected professional athlete is a football player? Show your work.

ii. What is the probability a randomly selected professional athlete is in the age-group of "$25 \leq$ Age < 30" given they are a football player? Show your work.

Part B

A mosaic plot was constructed using the information in the two-way table. Use the mosaic plot to answer the following questions.

Relative Frequency of Age-Group by Type of Sport Played



i. The b and h displayed in the mosaic plot each represent a probability calculated in part A. Which probability does b correspond to: the probability calculated in part A (i) or the probability calculated in part A (ii)?

ii. What probability does the x displayed in the mosaic plot represent in context?

Part C

Use the information to answer the following questions.

i. Are the events "Baseball" and "$35 \leq$ Age" mutually exclusive events? Explain.

ii. Are the events "Baseball" and "$35 \leq$ Age" independent events? Show your work.

Part D

Consider the data of all professional athletes from the two-way table. Determine if it is appropriate to carry out a chi-square test for independence to investigate whether there is an association between age-group and sport played using these data. Explain your answer.

5. Baseball cards are trading cards that feature data on a player's performance in baseball games. Michelle is at a national baseball card collector's convention with approximately 20,000 attendees. She notices that some collectors have both regular cards, which are easily obtained, and rare cards, which are harder to obtain. Michelle believes that there is a relationship between the number of months a collector has been collecting baseball cards and whether the majority of the cards (cards appearing more often) in their collection are regular or rare. She obtains information from a random sample of 500 baseball card collectors at the convention and records how many full months they have been collecting baseball cards and whether the majority of the cards in their card collection are regular or rare. Her results are displayed in a two-way table.

Majority Type of Baseball Cards and Months of Collecting Baseball Cards

	Fewer Than 6 Months	6 - 10 Months	11 - 15 Months	16 - 20 Months	21 or More Months	**Total**
Has a Majority of Regular Baseball Cards	80	84	71	76	112	423
Has a Majority of Rare Baseball Cards	11	16	9	6	35	77
Total	91	100	80	82	147	500

(a) If one collector from the sample is selected at random, what is the probability that the collector has been collecting baseball cards for 11 or more months and has a majority of regular baseball cards? Show your work.

(b) Given that a randomly selected collector from the sample has been collecting baseball cards for fewer than 6 months, what is the probability the collector has a majority of regular baseball cards? Show your work.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 5** on this page.

(c) Michelle believes there is a relationship between the number of months spent collecting baseball cards and which type of card is the majority in the collection (regular or rare).

(i) Name the hypothesis test Michelle should use to investigate her belief. Do not perform the hypothesis test.

(ii) State the appropriate null and alternative hypotheses for the hypothesis test you identified in (c-i). Do not perform the hypothesis test.

(d) After completing the hypothesis test described in part (c), Michelle obtains a p-value of 0.0075. Assuming the conditions for inference are met, what conclusion should Michelle make about her belief? Justify your response.

GO ON TO THE NEXT PAGE.

5. A research center conducted a national survey about teenage behavior. Teens were asked whether they had consumed a soft drink in the past week. The following table shows the counts for three independent random samples from major cities.

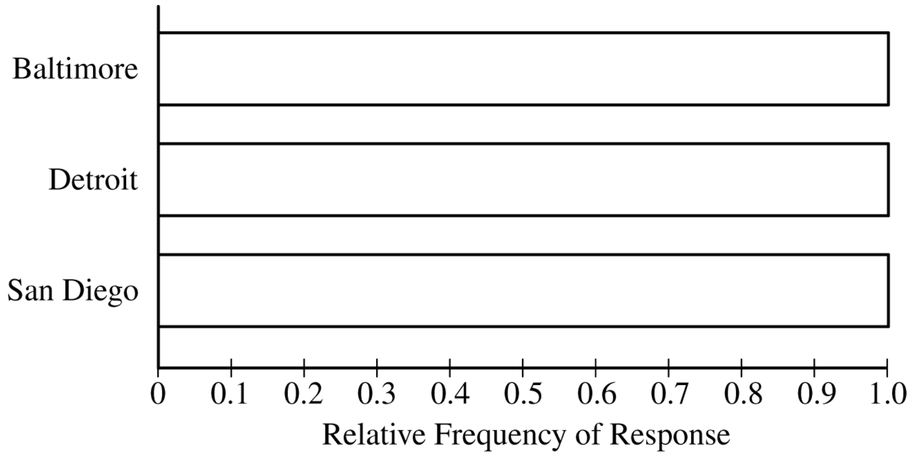
	Baltimore	Detroit	San Diego	Total
Yes	727	1,232	1,482	3,441
No	177	431	798	1,406
Total	904	1,663	2,280	4,847

(a) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the past week because Baltimore has the least number of teens who consumed a soft drink. Is the researcher's claim correct? Explain your answer.

GO ON TO THE NEXT PAGE.

(b) Consider the values in the table.

(i) Construct a segmented bar chart of relative frequencies based on the information in the table.



(ii) Which city had the smallest proportion of teens who consumed a soft drink in the previous week? Determine the value of the proportion.

(c) Consider the inference procedure that is appropriate for investigating whether there is a difference among the three cities in the proportion of all teens who consumed a soft drink in the past week.

(i) Identify the appropriate inference procedure.

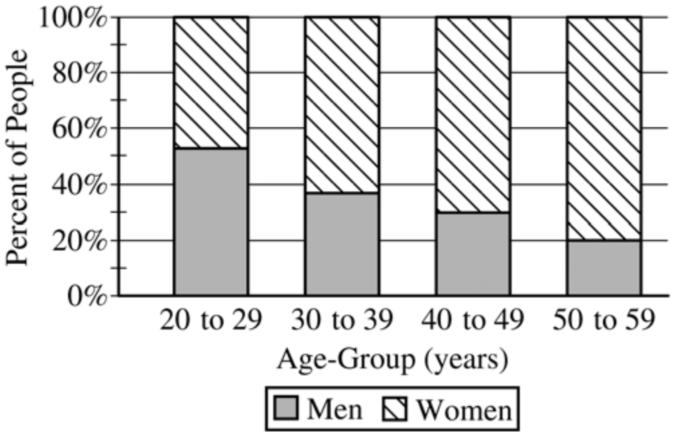
(ii) Identify the hypotheses of the test.

GO ON TO THE NEXT PAGE.

5. The table and the bar chart below summarize the age at diagnosis, in years, for a random sample of 207 men and women currently being treated for schizophrenia.

Age-Group (years)

	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
Men	53	23	9	3	88
Total	99	63	30	15	207



Do the data provide convincing statistical evidence of an association between age-group and gender in the diagnosis of schizophrenia?

AP Statistics: **2016**

2. Product advertisers studied the effects of television ads on children's choices for two new snacks. The advertisers used two 30-second television ads in an experiment. One ad was for a new sugary snack called Choco-Zuties, and the other ad was for a new healthy snack called Apple-Zuties.

For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad nor the Apple-Zuties ad.

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices.

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties
A	Choco-Zuties only	21	4
B	Apple-Zuties only	13	12
C	Neither	22	3

(a) Do the data provide convincing statistical evidence that there is an association between type of ad and children's choice of snack among all children similar to those who participated in the experiment?

(b) Write a few sentences describing the effect of each ad on children's choice of snack.

8.6 Carrying Out a Chi-Square Test for Homogeneity or Independence

Name: ______________ Class: __________ Date: ____________ **Total: 9 marks**

KEY WORDS chi-square 卡方 • test for homogeneity 同质性

Objective

Build the skills to answer exam questions on **carrying out a chi-square test for homogeneity or independence**.

You must be able to:

- compute χ^2 from a two-way table and find the **p-value**
- compare it to α and reach a decision
- state a conclusion about association in context

1 Worked examples

Study these first. Each one shows the method for a question type used later.

The procedure

1. State the hypotheses; check conditions (expected counts ≥ 5).
2. Compute the expected counts, then $\chi^2 = \sum \frac{(O-E)^2}{E}$, with df $= (r-1)(c-1)$.
3. Find the **p-value**.
4. Compare to α: a **small** p-value $\rightarrow$ **reject** H_0 $\rightarrow$ the variables are **associated**.

Finding an expected count

Each expected count is $\frac{\text{row total} \times \text{column total}}{\text{grand total}}$. For a cell whose row totals 40 and column totals 50, with grand total 200:

$$E = \frac{40 \times 50}{200} = 10.$$

Compute every E, then sum $\frac{(O-E)^2}{E}$.

2 Practice

2.1 State the steps to carry out a chi-square test for a two-way table. [2]

2.2 State what a small p-value indicates about the two variables. [1]

2.3 A test of independence gives a p-value of 0.02 at $\alpha = 0.05$. State the decision. [1]

3 Exam-style questions

3.1 A chi-square test of independence with a small p-value suggests the variables are [1]

- **A** independent
- **B** associated
- **C** exactly equal
- **D** unrelated

3.2 A chi-square test rejects H_0 when the p-value is [1]

- **A** above α
- **B** below α
- **C** exactly zero
- **D** exactly one

3.3 A test of independence gives $\chi^2 = 10$, df $= 2$, p-value 0.007 at $\alpha = 0.05$.

(a) State the decision. [1]

(b) State whether the variables are associated. [1]

(c) State the conclusion in context. [1]

Solutions

2.1 state hypotheses and check conditions; compute expected counts and χ^2; find the p-value; compare to α and conclude.

2.2 the two variables are associated (not independent).

2.3 reject H_0 $(0.02 < 0.05)$.

3.1 B.

3.2 B.

3.3 (a) reject H_0. (b) yes, they are associated. (c) there is convincing evidence of an association between the two variables.

5. Baseball cards are trading cards that feature data on a player's performance in baseball games. Michelle is at a national baseball card collector's convention with approximately 20,000 attendees. She notices that some collectors have both regular cards, which are easily obtained, and rare cards, which are harder to obtain. Michelle believes that there is a relationship between the number of months a collector has been collecting baseball cards and whether the majority of the cards (cards appearing more often) in their collection are regular or rare. She obtains information from a random sample of 500 baseball card collectors at the convention and records how many full months they have been collecting baseball cards and whether the majority of the cards in their card collection are regular or rare. Her results are displayed in a two-way table.

Majority Type of Baseball Cards and Months of Collecting Baseball Cards

	Fewer Than 6 Months	6 - 10 Months	11 - 15 Months	16 - 20 Months	21 or More Months	**Total**
Has a Majority of Regular Baseball Cards	80	84	71	76	112	423
Has a Majority of Rare Baseball Cards	11	16	9	6	35	77
Total	91	100	80	82	147	500

(a) If one collector from the sample is selected at random, what is the probability that the collector has been collecting baseball cards for 11 or more months and has a majority of regular baseball cards? Show your work.

(b) Given that a randomly selected collector from the sample has been collecting baseball cards for fewer than 6 months, what is the probability the collector has a majority of regular baseball cards? Show your work.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 5** on this page.

(c) Michelle believes there is a relationship between the number of months spent collecting baseball cards and which type of card is the majority in the collection (regular or rare).

(i) Name the hypothesis test Michelle should use to investigate her belief. Do not perform the hypothesis test.

(ii) State the appropriate null and alternative hypotheses for the hypothesis test you identified in (c-i). Do not perform the hypothesis test.

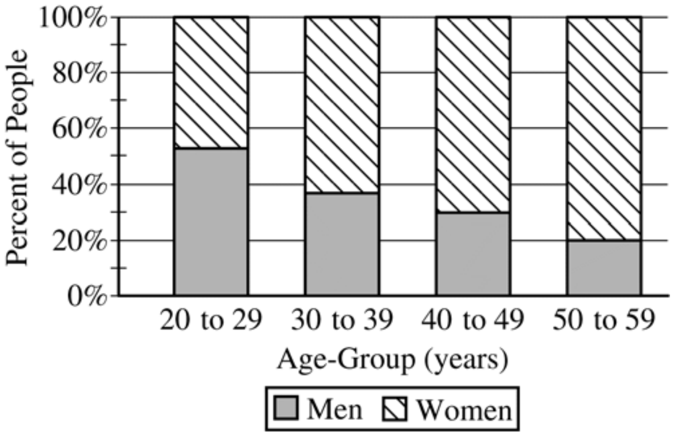
(d) After completing the hypothesis test described in part (c), Michelle obtains a p-value of 0.0075. Assuming the conditions for inference are met, what conclusion should Michelle make about her belief? Justify your response.

GO ON TO THE NEXT PAGE.

5. The table and the bar chart below summarize the age at diagnosis, in years, for a random sample of 207 men and women currently being treated for schizophrenia.

Age-Group (years)

	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
Men	53	23	9	3	88
Total	99	63	30	15	207



Do the data provide convincing statistical evidence of an association between age-group and gender in the diagnosis of schizophrenia?

AP Statistics: **2016**

2. Product advertisers studied the effects of television ads on children's choices for two new snacks. The advertisers used two 30-second television ads in an experiment. One ad was for a new sugary snack called Choco-Zuties, and the other ad was for a new healthy snack called Apple-Zuties.

For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad nor the Apple-Zuties ad.

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices.

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties
A	Choco-Zuties only	21	4
B	Apple-Zuties only	13	12
C	Neither	22	3

(a) Do the data provide convincing statistical evidence that there is an association between type of ad and children's choice of snack among all children similar to those who participated in the experiment?

(b) Write a few sentences describing the effect of each ad on children's choice of snack.

STATISTICS
SECTION II
Part A
Questions 1-5

Spend about 65 minutes on this part of the exam.

Percent of Section II score—75

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. An administrator at a large university is interested in determining whether the residential status of a student is associated with level of participation in extracurricular activities. Residential status is categorized as on campus for students living in university housing and off campus otherwise. A simple random sample of 100 students in the university was taken, and each student was asked the following two questions.

   - Are you an on campus student or an off campus student?
   - In how many extracurricular activities do you participate?

   The responses of the 100 students are summarized in the frequency table shown.

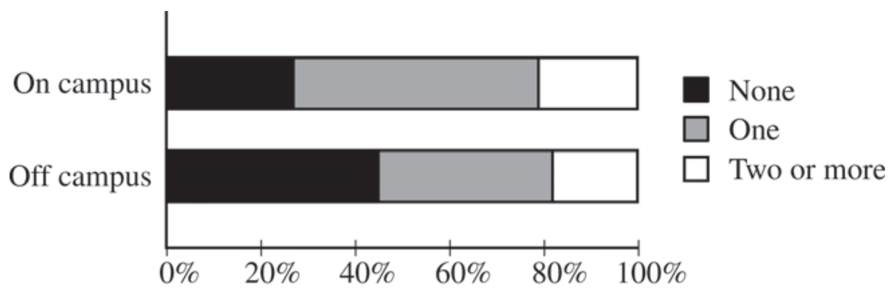
	Residential Status		
Level of Participation in Extracurricular Activities	On campus	Off campus	Total
No activities	9	30	39
One activity	17	25	42
Two or more activities	7	12	19
Total	33	67	100

(a) Calculate the proportion of on campus students in the sample who participate in at least one extracurricular activity and the proportion of off campus students in the sample who participate in at least one extracurricular activity.

On campus proportion:

Off campus proportion:

The responses of the 100 students are summarized in the segmented bar graph shown.



(b) Write a few sentences summarizing what the graph reveals about the association between residential status and level of participation in extracurricular activities among the 100 students in the sample.

(c) After verifying that the conditions for inference were satisfied, the administrator performed a chi-square test of the following hypotheses.

H_0 : There is no association between residential status and level of participation in extracurricular activities among the students at the university.

H_a : There is an association between residential status and level of participation in extracurricular activities among the students at the university.

The test resulted in a *p*-value of 0.23. Based on the *p*-value, what conclusion should the administrator make?

AP STATISTICS • EXERCISE SHEET

8.7 Skills Focus: Selecting an Appropriate Inference Procedure for Categorical Data

Name: ______________ Class: __________ Date: ____________ **Total: 8 marks**

KEY WORDS chi-square 卡方

Objective

Build the skills to answer exam questions on **selecting a categorical inference procedure**.

You must be able to:

- choose a **goodness-of-fit** test, a test of **homogeneity**, or a test of **independence**
- distinguish these chi-square procedures from a one- or two-proportion procedure

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Which categorical procedure?

- **Goodness of fit** — does **one** categorical variable match a claimed distribution?
- **Homogeneity** — is the distribution of one variable the **same across several groups**?
- **Independence** — are **two** categorical variables associated in **one** sample?
- **One/two proportion** — for a single proportion or the difference of two proportions (a two-category variable).

Worked choices

"Does this die match a uniform distribution?" – one variable versus a claim → **goodness of fit**. "Do three schools have the same grade distribution?" – one variable across groups → **homogeneity**. "Are gender and subject choice associated in one sample?" – two variables, one sample → **independence**.

2 Practice

2.1 State when to use a goodness-of-fit test. [1]

2.2 State when to use a test of independence. [1]

2.3 State the difference between a two-proportion test and a chi-square test of homogeneity. [1]

3 Exam-style questions

3.1 To test whether one categorical variable's distribution matches claimed proportions, use [1]

- **A** a t-test
- **B** a goodness-of-fit test
- **C** regression
- **D** a z-interval

3.2 To test the association between two categorical variables in one sample, use [1]

- **A** a goodness-of-fit test
- **B** a test of independence
- **C** a mean test
- **D** a slope test

3.3 A researcher compares the distribution of blood types across 3 hospitals.

(a) Name the type of data. [1]

(b) Name a suitable chi-square test. [1]

(c) State the degrees-of-freedom formula for the table. [1]

Solutions

2.1 when checking whether one categorical variable matches a claimed distribution.

2.2 when testing whether two categorical variables are associated in one sample.

2.3 a two-proportion test compares two groups on a two-category variable; homogeneity compares a variable with more than two categories across groups.

3.1 **B**.

3.2 **B**.

3.3 (a) categorical (counts). (b) a chi-square test of homogeneity. (c) $(r-1)(c-1)$.