RLP-STATSB · RLP Statistics B — Probability Models and Statistical Inference

Unit 4: Chi-Square, Regression Inference and the Capstone Study

Practice Workbook. The same problems, in the same order, as the Practice sections of the online lessons.

Name _____ Period _____

Lesson 4.1 — Chi-Square Goodness of Fit

FLUENCY

1. A 4-category variable is claimed to be uniform. Observed counts: 30, 22, 28, 10. Carry out a chi-square goodness-of-fit test.

2. A 3-category variable is claimed to be uniform. Observed counts: 39, 26, 25. Carry out a chi-square goodness-of-fit test.

3. A 3-category variable is claimed to be uniform. Observed counts: 38, 35, 17. Carry out a chi-square goodness-of-fit test.

4. A 3-category variable is claimed to be uniform. Observed counts: 29, 35, 26. Carry out a chi-square goodness-of-fit test.

5. A 3-category variable is claimed to be uniform. Observed counts: 77, 74, 49. Carry out a chi-square goodness-of-fit test.

6. A goodness-of-fit test has 6 categories and no parameters estimated from the data. What are the degrees of freedom?

7. The same test, but one parameter was estimated from the data. What are the df?

8. Find the P-value for $\chi^2 = 11.07$ with df = 5.

9. What is the mean of a chi-square distribution with df = 8?

10. An expected count is 3.2. What should you do before running the test?

APPLICATION

11. Springfield, 2000-2024 (NOAA daily): rainy days by weekday, Monday to Sunday: 363, 368, 392, 404, 396, 401, 365 (total 2,689). Test whether rain is spread evenly across weekdays.

12. Springfield, 2000-2024: days with at least 0.25 inch of rain, by month from October to April: Oct 105, Nov 85, Dec 69, Jan 69, Feb 77, Mar 114, Apr 128. Test whether heavy-rain days are spread across these months in proportion to their number of days, and name the months that contribute most.

13. Kansas City July weeks (2000-2024, 100 weeks of the first 28 days of July): a Binomial(7, 0.443) model for the number of days at 90°F or above. Pooling 0-1 and 5-7 days, observed 25, 18, 17, 12, 28 against expected 10.9, 22.1, 29.3, 23.3, 14.4. Carry out the goodness-of-fit test (p was estimated from the data).

CHALLENGE

14. Why is the P-value of a chi-square test always an upper-tail area, even though the alternative is 'the distribution differs in any way'?

REVIEW

15. Test H_0: $p = 0.2$ against H_a: $p > 0.2$, given 109 successes in a random sample of 500. Find the test statistic and the P-value.

16. (Unit 3) State the conclusion for P = 0.65 at $\alpha = 0.05$ in the weekday test.

Lesson 4.2 — Chi-Square Tests for Two-Way Tables

FLUENCY

1. A table has 3 rows and 4 columns. What are the df for a chi-square test of independence?

2. Row total 688, column total 690, grand total 2,256. Find the expected count.

3. Independence or homogeneity? One random sample of students is classified by grade and by preferred study time.

4. Independence or homogeneity? Separate random samples from three high schools are asked the same yes/no question.

5. For a 2×2 table, how is χ^2 related to the two-proportion z statistic?

APPLICATION

6. Springfield winter persistence table (2000-2024): rain-rain 234, rain-dry 350, dry-rain 353, dry-dry 1,319. Carry out the chi-square test of independence, with the expected counts.

7. Springfield winter days by yesterday's rain (dry / light 0.01-0.24 in / heavy ≥ 0.25 in) and today's (rain / dry): dry 353, 1,319; light 126, 241; heavy 108, 109. Test for association and describe the pattern with conditional proportions.

8. Kansas City, July-August 2000-2024: 90°F or above yesterday & today 466; yesterday, not today 182; not yesterday, 90°F or above today 179; neither 723. Test whether hot days are independent of the previous day, and explain what this means for any binomial model of hot days.

9. Rainy days by season, 2000-2024 (Dec-Feb, Mar-May, Jun-Aug, Sep-Nov): St. Louis 690, 866, 691, 599; Springfield 587, 838, 674, 590. Test whether the two stations have the same seasonal distribution of rainy days.

CHALLENGE

10. A student converts a two-way table of counts to percentages before computing χ^2. What goes wrong, and why?

REVIEW

11. (Unit 3) Two-proportion z-test: 294/688 against 396/1,568. Find the pooled proportion.

12. (RLP Statistics A) From a two-way table, how do you check independence?

Lesson 4.3 — Inference for the Slope of a Regression Line

FLUENCY

1. A slope estimate b = 0.30 has SE 0.12 with n = 27. Find t and the df for H_0: $\beta = 0$.
2. Find the two-sided P-value for t = 2.5 with df = 25.
3. Find the 95% interval for β when b = 0.30, SE = 0.12 and df = 25.
4. A slope test has residuals whose lag-1 correlation is 0.4. Is the reported P-value too large or too small, and why?
5. What do the letters in LINER stand for?

6. r = 0.5 with n = 27. Find the t statistic for H_0: $\beta = 0$ from r.

APPLICATION

7. Joplin annual mean temperature against year, 1950-2024 (75 years): b = 0.0243°F per year, SE = 0.00727, r = 0.364. Test H_0: $\beta = 0$ and give a 95% interval per decade.

8. St. Louis, same analysis (75 years): b = 0.0524°F per year, SE = 0.00799, r = 0.608. Compute t both from b/SE and from r, and interpret.

9. St. Louis precipitation against year, 1950-2024: b = 0.1810 in per year, SE = 0.0399, P ≈ 0.00. A student writes, 'The test proves St. Louis's rainfall has not changed.' Correct the conclusion.

CHALLENGE

10. Missouri's age-adjusted cancer death rate against year, 1999-2017 (CDC): b = -2.353 per 100,000 per year, SE = 0.0824, n = 19. Compute the 95% interval for β, then explain why a 'population' regression model is a delicate idea for these data.

REVIEW

11. A 4-category variable is claimed to be uniform. Observed counts: 27, 41, 26, 56. Carry out a chi-square goodness-of-fit test.

12. (RLP Statistics A) Interpret a slope of 0.0524°F per year for St. Louis annual mean temperature.

Lesson 4.4 — Choosing the Right Procedure and Writing It Up

FLUENCY

1. Choose the procedure: estimate the proportion of St. Louis days with rain from a random sample of 200 days.

2. Choose the procedure: do Springfield and Joplin differ in annual mean temperature over the same 30 years?

3. Choose the procedure: is rainfall at a station spread evenly over the days of the week?

4. Choose the procedure: is the number of 90°F days per year at Kansas City trending upward?

5. Choose the procedure: in a randomized experiment, do two groups of 12 differ in mean reaction time?

6. Choose the procedure: is preferred study time associated with grade level in a random sample of students?

7. Choose the procedure: what fraction of all 9,132 days in the St. Louis file were rainy?

8. Choose the procedure: do two independent random samples of households differ in the proportion with solar panels?

APPLICATION

9. A student wants to know whether heat waves in Kansas City have become longer. Propose a variable, a procedure and one condition that will be hard to meet.

10. A report states: 'We surveyed 400 visitors to our website; 62% support the plan (95% CI 57% to 67%).' Critique the choice of procedure.

11. Write the reproducibility steps a careful reader would need to recompute your St. Louis weekday chi-square test.

CHALLENGE

12. Two analysts study the same question: 'Is Springfield warmer than it used to be?' One uses a two-sample t-test on 1950-1979 against 1995-2024; the other uses a slope test on all 75 years. Compare the two approaches.

REVIEW

13. (Unit 2) Find the 95% interval for p with 66 successes in 200.

14. (Unit 1) Find the SD of $\hat{p}$ for n = 100 when p = 0.3117.

Unit 4 checkpoint

1. A goodness-of-fit test with 7 categories and no estimated parameters has df = _____.
2. $\chi^2 = 4.18$ with df = 6 gives $P \approx 0.65$. The conclusion is…
   A. Rain favors some weekdays
   B. Not convincing evidence that rain favors any weekday
   C. Rain is uniform, proven
   D. Reject H_0
3. True or false: A binomial goodness-of-fit test in which p was estimated from the data loses one extra degree of freedom.
4. The expected count for a cell with row total 688, column total 690 and grand total 2,256 is about _____.

5. For a 2 × 2 table, $\chi^2 = 68.8$. The two-proportion z statistic is about…
   A. 8.3
   B. 68.8
   C. 34.4
   D. 1.96
6. True or false: A chi-square statistic can be computed from percentages instead of counts.
7. A slope of 0.0303 with SE 0.00713 gives t ≈ _____.
8. The Springfield temperature slope test has $P < 0.001$ and $r^2 = 0.198$. Which is true?
   A. The trend explains most of the variation
   B. There is strong evidence of a trend, but it explains only about 20% of the variation
   C. The P-value is wrong
   D. r^2 should be 0.0007
9. True or false: Positively autocorrelated residuals make the slope test's P-value too small.
10. Which question needs no inference?
    A. Mean daily high over 2000-2024 from all 9,132 recorded days
    B. Mean daily high from a random sample of 30 days
    C. Whether two random samples differ
    D. Whether a slope is nonzero
11. Same 30 years at two stations: which procedure?
    A. Two-sample t
    B. Paired t
    C. Chi-square
    D. One-proportion z
12. True or false: A reproducibility appendix lets a careful reader obtain your numbers from the same data by following your steps.

Sejal Learning Systems · Royal Legacy Preparatory School · Generated from the course's practice files; generated problems are checked by computer, authored problems by their author.