

Name:

Solution

1. An economist believes that women are more successful business people than men. He collects data to assess whether the proportion of women-owned businesses that fail is smaller than the proportion of men-owned businesses that fail. He observes a random sample of 148 small businesses for a three-year period. During that time, 106 of the businesses headed by men and 7 of the 42 businesses headed by women fail. The attached SAS output can be used to answer some of the following questions. In the dataset used in SAS, the codings are:

gender: M = men
 F = women
 failed: F = failed
 S = succeeded (did not fail)

(a) Should the economist carry out a one-sided or two-sided hypothesis test? Briefly state why. *One-sided. He believes that the proportion of failures in women-owned businesses is smaller than in men-owned.*

2. Write the null and alternative hypotheses that the economist wishes to test. Use conventional statistical symbols.

3. $H_0: P_W = P_M$
 $H_A: P_W < P_M$

(c) Calculate the expected count for the first cell of the contingency table (layout as in the SAS output). Numeric answer; show your work.

3. $\frac{22}{148}(42) = 6.24$

(d) Should the economist reject the null hypothesis? Why or why not?

2. *No. The data goes in the direction of supporting H_0 . $P_W = 16.67 > P_M = 14.15$.*

(e) What does your answer mean in terms of business failures for women versus men?

3. *The data does not give evidence that the proportion of women-owned businesses that fail is smaller than the proportion of men-owned businesses that fail.*

(f) Why was the Chi-square test appropriate for this problem, instead of the two-sample t test?

3. *We are comparing population proportions. F-test is for means.*

2. You are interested in estimating the proportion of failures in the population of women-owned businesses. You will use the economist's data on women-owned businesses from the previous problem (7 failures in 42 businesses).

(a) Are the rules of thumb met for the use of the large-sample normal approximation? List the rules of thumb and assess each one for these data.

3. *Population at least 10 times as large as sample. Yes, there are at least 420 women-owned businesses. At least 10 successes and 10 failures in sample. No.*

(b) Use the plus-4 method to calculate a 95% confidence interval for the population proportion of failures in women-owned businesses. (Numeric answer; show your work.)

4. $\hat{p} = \frac{7+2}{42+4} = .1957$
 $\hat{p} \pm 1.96 \sqrt{\hat{p}(1-\hat{p})} = (.0810, .3104)$

(c) What was the margin of error in your confidence interval? (Numeric answer).

2. $1.96 \sqrt{\hat{p}(1-\hat{p})} = .118$

(12)

3. This problem is based on data on per capita gross domestic product and life expectancy from the Gap Minder project (gapminder.org). Data from 143 countries in 2007 are used.

(a) I was interested in using per capita gross domestic product (gdppercap) to predict life expectancy (lifeexp). Refer to the SAS output to explain why I chose to use the log of per capita gross domestic product (loggdppc) instead of the untransformed version. The relationship between

2 The relationship between $\log(\text{gdppc})$ and lifeexp is more linear.

(b) I wish to test the null hypothesis that there is no linear relationship between $\log(\text{gdppc})$ and lifeexp in the population of all countries. Write the null and alternative hypotheses using the statistical symbols we used in class.

3 $H_0: \beta = 0$
 $H_A: \beta \neq 0$

(c) Cite two different parts of the SAS output to explain why I should or should not reject H_0 at significance level $\alpha = 0.05$.

3 95% CI for β is (6.328, 8.017) which does not contain 0.

provide for t-test for $\beta = 0$ is $t = 2.203$ which is < 2.000

(d) Country A has 2 log units higher $\log(\text{gdppc})$ than Country B. What is the expected difference between lifeexp in Country A and Country B? (Numeric answer)

2 Give a 95% prediction interval for life expectancy in an individual new country with $\log(\text{gdppc})$ equal to 10.45.

2 (65.98, 94.42)

general

(f) Would the 95% confidence interval for the mean life expectancy in all countries with $\log(\text{gdppc})$ equal to 10.45 be wider or narrower than the 95% prediction interval that you gave above? Briefly justify your answer.

2 narrower. The prediction interval is wider than the confidence interval because it includes individual variability.

(14)