

MSCR 500 Homework Assignment 6

Anish Shah

11/12/19

Do the problems by hand first:

- all parts of hypothesis test
- show details of test statistic calculation and p-value determination
- use the stats tables from Weiss
- use 5% level for significance / alpha

Verify all problems by analyzing data in SAS

- write the test statistic and p-value in appropriate spaces on pages 2-4
- code included at end of each

Question 1

Question description

- 1000 people chosen at random from population
- 350 exposed and had disease
- 100 exposed but did no disease
- 110 not exposed and had disease

Is the occurrence of disease related to exposure of the agent?

Answers to Q1

Hypothesis:

H0: There is no association between exposure and disease.

H1: There is an association between exposure and disease.

$\alpha = 0.05$

Hypothesis test: chi-square test of independence

This is the correct test to be used because we have two independent samples from each population, with theoretically independent frequencies. This follows the chi-square distribution with 1 degree of freedom. Also follows a 2x2 contingency table pattern.

$N(D|E) = 350$

$N(D|E') = 110$

$N(D'|E) = 100$

$N(D'|E') = 440$

$N = 1000$

Total exposed = 450

Total unexposed = 550

Total disease = 460

Total no disease = 540

If H0 is true, then $p_1 = p_2 = p_p$ (pooled estimate)

$p_p = \text{total dz} / \text{total} = 460 / 1000 = .46$

$D|E = 350$, expected 207

$D|E' = 110$, expected 243

$D'|E = 100$, expected 253

$D'|E' = 440$, expected 297

$$\begin{aligned}\chi^2 &= \text{sum}[(\text{observed} - \text{expected})^2 / \text{expected}] \\ &= (350-207)^2/207 + (110-243)^2/243 + (100-253)^2/253 + (440-297)^2/297 \\ &= 98.787 + 72.794 + 92.525 + 68.851 \\ &= 332.957\end{aligned}$$

$$P(X^2 = 332.957) < 0.005$$

SAS values: $X^2 = 332.6176$, $p < 0.0001$

Reject H0 at alpha = 0.05

Based on this chi square test of independence, we can reject the null hypothesis that proportion of disease between exposure and unexposed is the same (at an alpha of 0.05). We can thus accept the H1 such that there is likely a difference in disease probability in the exposed versus unexposed groups.

Question 2

Question description

Condition = died, lived, total

Seat belt = 15, 335, 350

Harness = 10, 240, 250

None = 25, 375, 400

Total dead = 50

Total lived = 950

Total = 1000

Answer to Q2

H0: Safety device and survival are not associated.

H1: Safety device type and survival are associated.

alpha = 0.05

Test = chi-square test of independence using an RxC table

The X^2 test is the appropriate test since we have independent random sample in the form of a contingency / frequency table. We can tabulate the proportions in each group. The degrees of freedom will be RxC (2x3), leading to a DF of 2. This will also be the right test as >80% of the expected values are >5.

$$pp = 50/1000 = 0.05$$

Expected by seat belt = 17.5, 332.5

Expected by harness = 12.5, 237.5

Expected by no device = 20, 380

$$\begin{aligned}X^2 &= \text{sum}[(\text{observed} - \text{expected})^2 / \text{expected}] \\ &= (15-17.5)^2/17.5 + (335-332.5)^2/332.5 + (10-12.5)^2/12.5 + (240-237.5)^2/237.5 + (25-20)^2/20 + (375-380)^2/380 \\ &= 2.2180\end{aligned}$$

$$P(X^2 = 2.2180, df = 2) > 0.10$$

SAS values: $X^2 = 2.2180$, $p = 0.3299$

Fail to reject H0 at alpha = 0.05

The data provided by the chi square test statistic suggests that our p-value is > 0.10, and thus based on our alpha of 0.05, we do not have sufficient evidence to reject the null hypothesis. It looks likely safety device does not have an association with car crashes.

Question 3

Question description

Total in 1983 = 2637
Drinking in 1983 = 1250
Total in 1987 = 2657
Drinking in 1987 = 991

Answer to Q3

H0: $p_1 - p_2 = 0$
H1: $p_1 - p_2 \neq 0$

p_1 = proportion of drunk drivers in 1983
 p_2 = proportion of drunk drivers in 1987

$\alpha = 0.05$

Chi-square test of homogeneity

This is the correct test to use because we have two populations that can be pooled together, or separated, by a factor (change in legal age of drinking / year). The contingency table describes frequencies, and the sample is independent/random.

$p_1 = 0.4740$ (1983)
 $p_2 = 0.3730$ (1987)

$pp = 0.4233$

Drunk in 1983 = 1250, expected 1116.2421
Sober in 1983 = 1387, expected 1520.7579
Drunk in 1987 = 991, expected 1124.7081
Sober in 1987 = 1666, expected 1532.2919

$$X^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(1250-1116.2421)^2}{1116.2421} + \frac{(1387-1520.7579)^2}{1520.7579} + \frac{(991-1124.7081)^2}{1124.7081} + \frac{(1666-1532.2919)^2}{1532.2919}$$
$$= 55.3556$$

$P(X^2 = 55.3556, df = 1) < 0.005$

SAS values: $X^2 = 55.3553$, $P < 0.0001$

Reject H0 at $\alpha = 0.05$

We can safely reject the null hypothesis at α of 0.05, based on our test statistic, and accept the H1 that there is a difference in proportion of drunk drivers after the legal age was changed.