Chapter 6

Andrew Goldberg

IS606

6.6 2010 Healthcare Law. On June 28, 2012 the U.S. Supreme Court upheld the much debated 2010 healthcare law, declaring it constitutional. A Gallup poll released the day after this decision indicates that 46% of 1,012 Americans agree with this decision. At a 95% con\_dence level, this sample has a 3% margin of error. Based on this information, determine if the following statements are true or false, and explain your reasoning.39

(a) We are 95% con\_dent that between 43% and 49% of Americans in this sample support the decision of the U.S. Supreme Court on the 2010 healthcare law.

**False, we know exactly how many in the sample support the healthcare law**

(b) We are 95% con\_dent that between 43% and 49% of Americans support the decision of the U.S. Supreme Court on the 2010 healthcare law.

**True, that is the percent of americans taking into account of the MoE that support the law**

(c) If we considered many random samples of 1,012 Americans, and we calculated the sample proportions of those who support the decision of the U.S. Supreme Court, 95% of those sample proportions will be between 43% and 49%.

(d) The margin of error at a 90% con\_dence level would be higher than 3%.

**False, it would be smaller, because we are asking for less precision**

6.12 Legalization of marijuana, Part I. The 2010 General Social Survey asked 1,259 US residents: \Do you think the use of marijuana should be made legal, or not?" 48% of the respondents said it should be made legal.44

(a) Is 48% a sample statistic or a population parameter? Explain.

**Sample statistic, it is based on a sample.**

(b) Construct a 95% confidence interval for the proportion of US residents who think marijuana should be made legal, and interpret it in the context of the data.

Se = sqrt((.48\*.52)/1259) = 1.4%

.48±1.96\*.014 = **(.45, .51)**

(c) A critic points out that this 95% confidence interval is only accurate if the statistic follows a normal distribution, or if the normal model is a good approximation. Is this true for these data? Explain.

**Yes, There’s more than 10 of both legal/not legal responses, and the data, presumably is based on a simple random sample of less than 10% of the population**

(d) A news piece on this survey's findings states, \Majority of Americans think marijuana should be legalized." Based on your confidence interval, is this news piece's statement justified?

**48% is below a majority, although it is possible, with the upper end of the confidence interval at 51%. But no, the statement is not justified.**

6.20 Legalize Marijuana, Part II. As discussed in Exercise ??, the 2010 General Social Survey reported a sample where about 48% of US residents thought marijuana should be made legal. If we wanted to limit the margin of error of a 95% confidence interval to 2%, about how many Americans would we need to survey?

1.96 \* (sqrt((.48\*.52)/n)) < .02

**2397.16 < n, or 2398**

6.28 Sleep deprivation, CA vs. OR, Part I. According to a report on sleep deprivation by the Centers for Disease Control and Prevention, the proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 8.0%, while this proportion is 8.8% for Oregon residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents. Calculate a 95% confidence interval for the difference between the proportions of Californians and Oregonians who are sleep deprived and interpret it in context of the data.53

SE ≈ √((.08\*.92)/11545)+((.088 \* .912)/4691)

SE ≈ √.002348 = .0048

.08 ± 1.96 \* .0048 = (.071, .089)

6.44 Barking deer. Microhabitat factors associated with forage and bed sites of barking deer in Hainan Island, China were examined from 2001 to 2002. In this region woods make up 4.8% of the land, cultivated grass plot makes up 14.7%, and deciduous forests makes up 39.6%. Of the 426 sites where the deer forage, 4 were categorized as woods, 16 as cultivated grassplot, and 61 as deciduous forests. The table below summarizes these data.62

Woods: 4

Cultivated grassplot: 16

Deciduous forests: 67

Other: 345

Total: 426

(a) Write the hypotheses for testing if barking deer prefer to forage in certain habitats over others.

**Ho: Barking deer show no preference for foraging habitats**

**Ha: Barking deer have habitat preferences for foraging**

(b) What type of test can we use to answer this research question?

**Chi-squared goodness of fit test**

(c) Check if the assumptions and conditions required for this test are satisfied.

Expected counts:

Cases are independent

All have a sample size above 5, so good to go

(d) Do these data provide convincing evidence that barking deer prefer to forage in certain habitats over others? Conduct an appropriate hypothesis test to answer this research question.

Woods: (4-20.5)^2/20.5 = 13.2

Cultivated grass: (16-62.6)^2 /62.6 = 34.7

Deciduous forests: (67-168.7)^2/168.7 = 61.3

Woods: .048 \* 426 = 20.5

Cultivated grass: 62.6

Deciduous forests: 168.7

X^2 = 109.25

With df = 2, 109 is way off the chart, so we reject the null in favor of the alternative: deer prefer to forage in certain habitats

6.48 Coffee and Depression. Researchers conducted a study investigating the relationship between caffeinated coffee consumption and risk of depression in women. They collected data on 50,739 women free of depression symptoms at the start of the study in the year 1996, and these women were followed through 2006. The researchers used questionnaires to collect data on caffeinated coffee consumption, asked each individual about physician-diagnosed depression, and also asked about the use of antidepressants. The table below shows the distribution of incidences of depression by amount of caffeinated coffee consumption.63

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | <1 week | 2-6 week | 1 day | 2-3 day | >=4 day | Total |
| Yes | 670 | 373 | 905 | 564 | 95 | 2607 |
| No | 11545 | 6244 | 16329 | 11726 | 2288 | 48132 |
| Total | 12215 | 6617 | 17234 | 12290 | 2383 | 50739 |

(a) What type of test is appropriate for evaluating if there is an association between coffee intake

and depression?

**Chi-squared goodness of fit test**

(b) Write the hypotheses for the test you identified in part (a).

**Ho: there is no association between coffee intake and depression**

**Ha: there is an association between coffee intake and depression**

(c) Calculate the overall proportion of women who do and do not suffer from depression.

48132/50739 = **.949**

(d) Identify the expected count for the highlighted cell, and calculate the contribution of this cell to the test statistic, i.e. (Observed 􀀀 Expected)2=Expected.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| expected | <1 | 2-6 | 1 | 2-3 | >=4 | Total |
| Yes | 627.6139656 | **339.9854** | 885.4932 | 631.4675 | 122.44 | 2607 |
| No | 11587.38603 | 6277.015 | 16348.51 | 11658.53 | 2260.56 | 48132 |

=(373-340)^2/340 = **3.2**

(e) The test statistic is x^2 = 20.93. What is the p-value?

Df = (5-1)(2-1) = 4

**p-value is below .001**

(f) What is the conclusion of the hypothesis test?

**We reject the null in favor of the alternative**

(g) One of the authors of this study was quoted on the NYTimes as saying it was \too early to recommend that women load up on extra coffee" based on just this study.64 Do you agree with

this statement? Explain your reasoning.

**Yes, it is an observational study, not an experimental study, so there’s no causal link between coffee and depression, just an association.**