

6.6 2010 Healthcare Law.

Legalization of marijuana, Part I.

Legalize Marijuana part II.

Sleep deprivation

Barking deer

Coffee and Depression.

DS606-HW6

[Code ▼](#)

George Cruz

2020-10-21

[Hide](#)

```
library(tidyverse)
library(openintro)
library(scales)
```

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% confidence 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. (a) *We are 95% confident 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 that 46% of Americans in this sample support the decision.

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

True, because at 95% confidence, this sample has a 3% margin of error.

(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%.* True. (d) *The margin of error at a 90% confidence level would be higher than 3%.* False. If we lower the confidence level the margin of error also lowers.

Legalization of marijuana, Part I.

(6.10 p.216) 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. (a) *Is 48% a sample statistic or a population parameter? Explain.* A Sample statistic. We are not inferring proportion but calculating a percentage of the answers.

(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

[Hide](#)

```
z = 1.96
n <- 1259
p <- .48
se <- sqrt((p*(1-p))/n)
ci_low <- p - (z * se)
ci_high <- p + (z * se)
ci_95 <- c(ci_low, ci_high)
ci_95
```

```
## [1] 0.4524028 0.5075972
```

(c) The central limit theorem states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough. This sample size is large enough (1259 observations and the variable is independent)

(d) It could be, if we hit our high limit in the confidence interval. A more accurate one would be half of Americans approve legalization.

Legalize Marijuana part II.

[Hide](#)

```
m = 0.02
p = .48
z = 1.96
se = m/z
n = (p*(1-p))/se**2
round(n,0)+1
```

```
## [1] 2398
```

We would need to question around 2398 US residents.

Sleep deprivation

[Hide](#)

```

p_ca <- 0.08
n_ca <- 11545
p_or <- 0.088
n_or <- 4691

p <- p_or - p_ca

se <- ( ((p_ca * (1 - p_ca)) / n_ca) + ((p_or * (1 - p_or)) / n_or)) ** 0.5
z <- 1.96

me <- z * se

ci_95 <- c(p - me, p + me )

ci_95

```

```
## [1] -0.001498128  0.017498128
```

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.

Woods | Cultivated grassplot | Deciduous forests | Other | Total 4 16 61 345 426

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

- H0 - Deer do not prefer to forage in certain habitats over others
- H1- Deer have a preference in habitat

(b) What type of test can we use to answer this research question? We can use a chi-squared test to answer this question.

(c) Check if the assumptions and conditions required for this test are satisfied. The chi-square goodness of fit test is appropriate when the following conditions are met: The sampling method is simple random sampling. The variable under study is categorical. The expected value of the number of sample observations in each level of the variable is at least 5.

Hide

```

obs <- c(woods=0.048 * 426, grassplot=0.147 * 426, forests = 0.396 * 426, other = (1-
(.048 + .147 + .396)) * 426)

obs

```

```
##      woods grassplot  forests   other
##    20.448    62.622   168.696   174.234
```

It appears all the conditions for the test are met.

(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.

Hide

```
chisq.test(x = c(4, 16, 61, 345), p = c(0.048, 0.147, 0.396, 1-(.048+.147+.396)))
```

```
##
## Chi-squared test for given probabilities
##
## data:  c(4, 16, 61, 345)
## X-squared = 284.06, df = 3, p-value < 2.2e-16
```

Hide

```
print(2.2e-16 < 0.5)
```

```
## [1] TRUE
```

The data provides convincing evidence that the null hypothesis is wrong. Because the value of p is less than 0.05, we reject H_0 and accept H_1 : Deer have a preference in habitat.

Coffee and Depression.

(6.50, p. 248) 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.

depression	less_1_cup_week	2_to_6_cups_week	1_cup_day	2_to_3_cups_day	4_cups_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?

The chi square test will be appropriate to test for association between these two variables.

(b) Write the hypotheses for the test you identified in part (a). H_0 : There is no connection between coffee intake and depression. H_1 : There is a connection between those two variables.

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

Hide

```
total <- 50739
total_no <- 48132
total_yes <- 2607

prop_yes <- total_yes/total
prop_no <- total_no/total

c(yes=percent(prop_yes, accuracy = 0.01), no=percent(prop_no, accuracy = 0.01))
```

```
##      yes      no
## "5.14%" "94.86%"
```

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

[Hide](#)

```
expected <- prop_yes * 6617

contri <- ((373 - expected) ** 2)/expected

c(expected= expected, contrib = contri)
```

```
##      expected      contrib
## 339.985396    3.205914
```

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

[Hide](#)

```
p_val <- pchisq(20.93, df=4, lower.tail=FALSE)
p_val
```

```
## [1] 0.0003269507
```

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

Because the value of p is less than 0.05, we reject H0 and accept H1

(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. Do you agree with this statement? Explain your reasoning.

This probably means that, even though there is some connection between the amount of coffee intake and depression, a correlation does not mean causality and other factors might have to be taken into consideration before making a clinical recommendation on the benefits of coffee to prevent depression.

...