SHAIKH SHAWON AREFIN SHIMON skshimon@cs.colostate.edu
STAT511 HW9 830201247

1) **Question:**
An investigator is interested in estimating the proportion of cats (over age 7) suffering from diabetes. The investigator would like to have a 95% ME of 10% or less. Answers should be based on the large sample normal approximation.
A. Using a conjectured proportion of 0.20, what sample size is required?

Answer:

Confidence Interval width, **2E** = 10% = 0.10

Margin of Error, **E, half width of CI = 0.05**

Conjectured proportion, $\hat{\pi} = \mathbf{0.20}$

For 95% confidence interval, $Z_{\alpha/2} = Z_{0.025} = \mathbf{1.960}$

Sample size n = $\frac{Z_{\alpha/2}^2(\hat{\pi}(1-\hat{\pi}))}{E^2} = \frac{1.96^2*(0.2*0.8)}{0.05^2} = 245.8534 \approx 246$

B. Without using the conjectured proportion from above, what (maximum) sample size is required?

Answer:

The choice of Conjectured proportion, $\hat{\pi} = \mathbf{0.5}$ will give the maximum sample size.

Maximum sample size n = $\frac{Z_{\alpha/2}^2(\hat{\pi}(1-\hat{\pi}))}{E^2} = \frac{1.96^2*(0.5*0.5)}{0.05^2} = 384.1459 \approx 385$

2) **Question:**
The Cartoon Network conducted a nation-wide survey to assess viewer attitudes toward Superman. Using a simple random sample, they selected 400 boys and 300 girls. Forty percent of the boys stated that Superman is their favorite cartoon character, compared to thirty percent of the girls.
A. Calculate the 90% confidence interval for the true percent difference in viewer attitude between the boys and the girls using the normal approximation.

Answer:

	Superman Yes	Superman No	
Boys	160	240	400
Girls	90	210	300
	250	450	700

Using prop.test,

```
2-sample test for equality of proportions without continuity
        correction

data:  c(160, 90) out of c(400, 300)
X-squared = 7.4667, df = 1, p-value = 0.006285
alternative hypothesis: two.sided
90 percent confidence interval:
 0.04069396 0.15930604
sample estimates:
prop 1 prop 2
   0.4    0.3
```

$y_1 = 400 * 0.4$ = 160, $n_1 = 400$

$y_2 = 300 * 0.3$ = 90, $n_2 = 300$

90% CI for $\pi_1 - \pi_2 = (0.04, 0.16)$

B. Based on the CI from A, is there a difference in attitude between the boys and girls? Provide justification for your response.

The 90% CI does not include 0, which means that there is a 90% certainty that there is a difference in attitude between boys and girls.

C. Using α=0.10, run a **two-sided** test comparing the proportion of boys vs girls that select Superman as their favorite character. Give your test statistic, p-value and conclusion.

H_0 : $\boldsymbol{\pi_{boys} - \pi_{girls} = 0 : \pi_1 - \pi_2 = 0}$
H_A : $\boldsymbol{\pi_1 - \pi_2 \neq 0}$

Using two sided prop.test, we get,

```
2-sample test for equality of proportions without continuity
        correction

data:  c(160, 90) out of c(400, 300)
X-squared = 7.4667, df = 1, p-value = 0.006285
alternative hypothesis: two.sided
90 percent confidence interval:
 0.04069396 0.15930604
sample estimates:
prop 1 prop 2
   0.4    0.3
```

Test statistics, z = 2.732
P-value = 0.006285 <0.10, reject H_0.

Conclusion:
There is a difference in attitude between boys and girls in the population from which the sample was taken.

3) **Question:**
This is problem 10.31 in the 6th edition of O&L. Does weather affect the occurrence of violent crimes? Sociologists have long debated whether certain atmospheric conditions are associated with increases in the homicide rate. A researcher classified 1500 homicides in the southwest US according to the season in which the homicide occurred.

	Winter	Spring	Summer	Fall
# of Homicides	328	372	471	329

A. Test the hypothesis that the homicide rates are equal among the four seasons using α = 0.05 level. State your hypotheses, test statistic, p-value and conclusion.

If the homicide rates are equal among the 4 seasons ,

H_0: $\pi_{Winter} = 1/4$
$\pi_{Spring} = 1/4$
$\pi_{Summer} = 1/4$
$\pi_{Fall} = 1/4$

H_A: At least one of the cell probabilities are different from the hypothesized value.

df = k-1 = 3

```
        Chi-squared test for given probabilities

data:  c(328, 372, 471, 329)
X-squared = 36.133, df = 3, p-value = 7.018e-08
```

Test statistics, $X^2 = 36.133$
p-value < 0.05. So we can reject the null hypothesis that homicide rates are equal among the four seasons with 95% confidence.

B. Calculate the Pearson residuals and state any conjectures that arise from these residuals.

Pearson's Residuals:

$r_{Winter} = -2.8025385$
$r_{Spring} = -0.1788854$
$r_{Summer} = 5.7243340$
$r_{Fall} = -2.7429101$

If the null hypothesis is true, there is only a 5% chance of a Pearson's residual to take a value outside of the interval between -1.96 and +1.96.

SHAIKH SHAWON AREFIN SHIMON skshimon@cs.colostate.edu
STAT511 HW9 830201247

Conjecture:
Data inconsistent with the assumption of proportion = 0.25 except for spring.

4) **Question:**
An experiment involving subjects with schizophrenia compared "personal therapy" to "family therapy". Only 2 out of 23 subjects assigned to the personal therapy group suffered psychotic relapses in the first year of the study, compared to 8 of the 24 subjects assigned to the family therapy group. The investigators were interested in testing the null hypothesis that the relapse rate is the same for personal and family therapies.

A) Report the test statistic and p-value from the chi-squared test.

H_0: $\pi_{personal\ therapy} = \pi_{Family\ Therapy}$
H_A: $\pi_{personal\ therapy}=\pi_{Family\ Therapy}$

	Relapse	No Relapse	
Personal Therapy	2	21	23
Family Therapy	8	16	24
	10	37	47

```
	Pearson's Chi-squared test

data:  schizophrenia_data
X-squared = 4.2563, df = 1, p-value = 0.0391
```

p-value = 0.0391

B)

```
	Fisher's Exact Test for Count Data

data:  schizophrenia_data
p-value = 0.07226
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.0180406 1.1769690
sample estimates:
odds ratio
  0.197105
```

p-value = 0.07226

C) Sample sizes are small. Fisher's exact test is appropriate.

APPENDIX:

```
> #HW9
>
> #Question 1
> z_alpha_by2 <- qnorm(1-(0.05/2))
> z_alpha_by2
[1] 1.959964
> conjectured_pie_hat <- 0.20
> E <- 0.05
> sample_size_n <- (z_alpha_by2*z_alpha_by2)*(conjectured_pie_hat*(1-conjectured_pie_hat))
> sample_size_n
[1] 245.8534
>
> #Maximum sample size
> conjectured_pie_hat <- 0.50
> sample_size_n <- (z_alpha_by2*z_alpha_by2)*(conjectured_pie_hat*(1-conjectured_pie_hat))
> sample_size_n
[1] 384.1459
>
> boys_superman <- 400*40/100
> boys_superman
[1] 160
> girls_superman <- 300*30/100
> girls_superman
[1] 90
>
> #Question 2
> boys_girls <- matrix(c(160,240,90,210), nrow = 2, byrow = TRUE)
> colnames(boys_girls) <- c("superman yes","superman no")
> rownames(boys_girls) <- c("boys", "girls")
> boys_girls
      superman yes superman no
boys           160         240
girls           90         210
> prop.table(boys_girls,1)
      superman yes superman no
boys           0.4         0.6
girls          0.3         0.7
>
> prop.test(c(160,90),c(400,300), alternative = "two.sided", conf.level = 0.90, correct =

        2-sample test for equality of proportions without continuity correction

data:  c(160, 90) out of c(400, 300)
X-squared = 7.4667, df = 1, p-value = 0.006285
alternative hypothesis: two.sided
90 percent confidence interval:
 0.04069396 0.15930604
sample estimates:
prop 1 prop 2
   0.4    0.3


>
> #Question 3
>
> chisq.test(c(328,372,471,329),p = c(1/4,1/4,1/4,1/4), correct = FALSE)

        Chi-squared test for given probabilities

data:  c(328, 372, 471, 329)
X-squared = 36.133, df = 3, p-value = 7.018e-08

>
```

```
> Counts <- c(328,372,471,329)
> probs <- c(1/4,1/4,1/4,1/4)
>
> total <- sum(Counts)
> total
[1] 1500
> Exp <- probs*total
> Exp
[1] 375 375 375 375
> Resid <- Counts - Exp
> SEResid <- sqrt(total*probs*(1-probs))
> PearsonResids <- Resid/SEResid
> PearsonResids
[1] -2.8025385 -0.1788854  5.7243340 -2.7429101
>
> critval <- qchisq(0.95,df=3)
> critval
[1] 7.814728
>
> #Question 4
>
> schizophrenia_data <- matrix(c(2,21,8,16),byrow = TRUE, nrow =2)
> colnames(schizophrenia_data) <- c("Relapse","No Relapse")
> rownames(schizophrenia_data) <- c("Personal Therapy","Family Therapy")
> schizophrenia_data
                 Relapse No Relapse
Personal Therapy       2         21
Family Therapy         8         16
> chisq.test(schizophrenia_data,correct = FALSE)

        Pearson's Chi-squared test

data:  schizophrenia_data
X-squared = 4.2563, df = 1, p-value = 0.0391

Warning message:
In chisq.test(schizophrenia_data, correct = FALSE) :
  Chi-squared approximation may be incorrect
> fisher.test(schizophrenia_data)

        Fisher's Exact Test for Count Data

data:  schizophrenia_data
p-value = 0.07226
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.0180406 1.1769690
sample estimates:
odds ratio
  0.197105


>
```