"Extra" Practice for Exam 2

Question #1: 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.
B. Based on the CI from A, is there a difference in attitude between the boys and girls? Provide justification for your response.
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.

#1 Solutions

A. 90% CI: (0.041, 0.159)
$\pi M - \pi F \pm z\alpha/2 * \text{sqrt}(\pi M(1-\pi M)/nM + \pi F(1-\pi F)/nF)$
$= 0.40-0.30 \pm 1.645 * \text{sqrt}(0.40(0.60)/400+0.30(0.70)/300)$
(0.0405, 0.1595)

B. We can conclude that there is a difference between the proportions because the confidence interval does not include 0.

C. H0: $\pi M - \pi F = 0$ vs HA: $\pi M - \pi F \neq 0$
$\pi = (yM+yF)/(nM+nF) = (160+90)/(400+300) = 0.357$
$Z = (\pi M - \pi F) / \text{sqrt}(\pi (1-\pi)(1/nM+1/nF)) = 0.10/\text{sqrt}(0.357 (1-0.357)(1/400+1/300))$
$= 2.73252$
p-value=2*P(Z>2.733)= 0.0063
Reject H0; conclude that boys are more likely to select Superman as their favorite character.

#1 R code

```
> prop.test(c(160,90), c(400,300), correct = FALSE, conf.le
vel = 0.90)

      2-sample test for equality of proportions without con
tinuity 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 #2: 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 $\alpha = 0.05$ level. State your hypotheses, test statistic, p-value and conclusion.
B. Calculate the Pearson residuals and state any conjectures that arise from these residuals.

#2 Solutions

A. H0: $\pi_1 = \pi_2 = \pi_3 = \pi_4$ vs HA: Not all the proportions are the same.
χ^2=36.133
p-value < 0.0001
Reject H0; the homicides are not equally spread among the seasons.

B. Pearson Residuals = $(n_i - E_i)/\sqrt{n\pi_i(1-\pi_i)} = (n_i - 375)/16.77$
Winter = -2.80
Spring = -0.17
Summer = 5.72
Fall = -2.74
Summer has more homicides than expected under the null hypothesis.

#2 R code

```
> chisq.test(c(328,372,471,329), p = c(0.25,0.25,0.25,0.25)
, 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

> #Calculating Pearson Residuals
> Counts <- c(328, 372, 471, 329)
> Props <- c(0.25, 0.25, 0.25, 0.25)
> Total <- sum(Counts)
> Exp <- Props*Total
> Resid <- Counts-Exp
> SEResid <- sqrt(Total*Props*(1-Props))
> PearsonResids <- Resid/SEResid
> PearsonResids
[1] -2.8025385 -0.1788854  5.7243340 -2.7429101
```