STA 138 Project

Dandi Peng Yuhan Ning

2/1/2019

## 1. Introduction

With a belief that newborns are uniformly distributed among 12 months, here are the data we own to verify this hypothesis.

## 2. Summary

Our data has 1023 subjects, where the numbers of newborns in Spring, Summer, Fall and Winter are 251, 251, 325 and 196.

From the above barplot, comparing the heights for number of newborns in four seasons, it is apparently shown that there are the highest number of newborns in Fall and the lowest number of newborns in Winter with a visually comparable difference.

## 3. Analysis

With the null hypothesis that a subject is equally likely to be born in any season, we are supposed to compare the proportions of newborns in four seasons and make comparisions.

Since it brings us to the comparison of proportions, the Pearson’s Chi-Square Test and Wilson-Adjusted Bonferroni Corrected Confidence Intervals can be applied here.

**Pearson’s Chi-Square Test**

## 4. Interpretation: Interpret your tests and/or confidence intervals in terms of the problem.

1. Conclusion: Describe and interpret your findings. What conclusions, if any, can you come to?

### R Appendix

knitr::opts\_chunk$set(echo = TRUE)  
library(ggplot2)  
born <- read.csv('born.csv', header = T)  
nrow(born)  
born\_df <- as.data.frame(table(born))  
born\_df  
  
ggplot(born,aes(x=Season,fill=Season))+geom\_bar(stat='count',width = 0.4)+  
 theme\_minimal() + scale\_fill\_brewer(palette = 'Pastel1')  
  
  
## Wilson-Adjusted Bonferroni Corrected Confidence Intervals  
alpha <- .05 ## at the confidence level of 95%  
g <- 4  
W\_A\_CI <- function(a,b,alpha){  
 pi\_1 <- (a[1]+1)/(sum(a)+2)  
 pi\_2 <- (b[1]+1)/(sum(b)+2)  
 z\_b <- qnorm(1-alpha)  
 sigma\_b <- sqrt(pi\_1\*(1-pi\_1)/(sum(a)+2)+pi\_2\*(1-pi\_2)/(sum(b)+2))  
 wa\_ci <- c(pi\_1-pi\_2-z\_b\*sigma\_b, pi\_1-pi\_2+z\_b\*sigma\_b)  
 return(wa\_ci)  
}