## (2 points) a) What is an estimate of the probability of a randomly-selected airline accident including a fatality?

Probability = (Number of accidents with one or more fatalities) / (Total number of accidents)

Number of accidents with one or more fatalities = 3381 Total number of accidents = 14944

Probability = 3381 / 14944 ≈ 0.2264

**Ans = 0.2264**

## (3 points) b) Some car insurance companies report that the likelihood of you (anyone) experiencing a car accident is ~70%. Test whether your estimate in part a) is different from 70% (in RStudio)? Paste your output here.

# Observed proportion (estimate from part a))  
p\_hat <- 3381 / 14944  
  
# Null hypothesis: The true proportion is 0.70  
# Alternative hypothesis: The true proportion is not equal to 0.70  
p\_null <- 0.70  
  
# Number of trials (total number of accidents)  
n <- 14944  
  
# Perform the one-sample proportion test  
prop.test(x = p\_hat \* n, n = n, p = p\_null)

##   
## 1-sample proportions test with continuity correction  
##   
## data: p\_hat \* n out of n, null probability p\_null  
## X-squared = 15970, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.7  
## 95 percent confidence interval:  
## 0.2195742 0.2330565  
## sample estimates:  
## p   
## 0.2262446

Ans = *Based on the extremely low p-value, we reject the null hypothesis and conclude that the estimate from part a) is significantly different from 70%.*

# (3 points) c) What is the confidence interval for your estimate in part a? (You can report the one that R gave you) How do you interpret this confidence interval in light of what you got in part a?

The confidence interval for the estimate obtained in part a) is provided in the output from the one-sample proportions test in RStudio:

95 percent confidence interval: 0.2195742 0.2330565

This means that we are 95% confident that the true proportion of airline accidents including a fatality falls within the interval [0.2196, 0.2331].

In light of the estimate obtained in part a), which was approximately 0.2262, we can see that this falls within the confidence interval. This indicates that our estimate is consistent with the range of values we would expect the true proportion to lie in 95% of the time. Therefore, our estimate aligns with the confidence interval, further supporting the conclusion drawn from the hypothesis test that the estimate is significantly different from 70%.

# (2 points) d) Why would someone use the Agresti-Coull method over the typical method found in most statistics books and programs? (be specific)

The Agresti-Coull method is often preferred over the typical method found in most statistics books and programs because it provides improved coverage probability for confidence intervals, especially when the sample size is small or the true proportion is close to 0 or 1.