r-worksheet - Composite Hypotheses

Younger Americans are better than their elders at separating factual from opinion statements in the news, according to a new analysis from Pew Research Center. Included in their study were 755 adults between the ages of 18 and 29 and 1108 adults between the ages of 50 and 64.

a. For the younger age group, 34% were able to correctly classify 5 factual and 5 opinion statements. For the older group, 22% made the classification correctly. Create 95% confidence intervals for the proportion of Americans of the appropriate age who would make the correct classification.

Confidence interval $= \hat{p} \pm z^*(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$, where \$z^* = \$ qnorm(1-0.025) since we want a 5% confidence interval

```
cat("The 95% confidence interval for the young group is: (", (0.34 -
qnorm(0.975)*sqrt(0.34*0.66/755)), ", ", (0.34 + qnorm(0.975)*sqrt(0.34*0.66/755)), ")\n")
```

```
## The 95% confidence interval for the young group is: ( 0.3062102 ,  0.3737898 )
```

```
cat("The 95% confidence interval for the old group is: (", (0.22 -
qnorm(0.975)*sqrt(0.22*0.78/1108)), ", ", (0.22 + qnorm(0.975)*sqrt(0.22*0.78/1108)), ")")
```

```
## The 95% confidence interval for the old group is: ( 0.1956086 ,  0.2443914 )
```

b. An educational program aimed at the youth is designed to help improve the ability to classify a fact from an opinion. With a hypothesis

$$H_0 : p \leq p_0 \quad \text{versus} \quad H_1 : p > p_0,$$

and for $p_0 = 0.34$, give the value of the power function $\pi(p)$ for $p = 0.34, 0.35, 0.36, 0.37.0.38$ and 0.39 with the choice of $\alpha = 0.02$ and a sample of size 750.

```
# Given values:
alpha<-0.02
N<-755 # I am using 755 since that was what was given in the first part of the worksheet
p0<-0.34
p<-c(0.34, 0.35,0.36,0.37,0.38,0.39)

# We will first find the critical value
pk<-p0 + qnorm(1-alpha)*sqrt(p0*(1-p0)/N)

# Now the power is the area under the distribution for the alternative hypothesis to the
# right of the critical value

# We need to find this for each of the curves with the different p-values
(1 - pnorm(pk, mean = p, sd = sqrt(p*(1-p)/N)))
```

```
## [1] 0.02000000 0.07164776 0.18890327 0.37915366 0.60257700 0.79449297
```

c. What qualititative change would you see in the power curve change if α is reduced to 0.01? Support your answer either with a clear explanation or a calculation.

If we decrease the value of alpha, we would expect the power to decrease. This is because when we decrease the value of alpha, the critical value of the distribution will increase, meaning the probability for a Type II error also increases. Therefore, since it is more likely to get a Type II error, the power decreases. Another way to think about this would be: since the value of pk is increasing due to the decrease in alpha, the area under the distribution for the alternative hypothesis to the right of pk decreases. Since this area is the measure of power and is decreasing, the power ultimately decreases. This can be seen in the calculations below.

```r
# Let's perform the same calculations for alpha = 0.01

# Given values:
alpha<-0.01
N<-755 # I am using 755 since that was what was given in the first part of the worksheet
p0<-0.34
p<-c(0.34, 0.35,0.36,0.37,0.38,0.39)

# We will first find the critical value
pk<-p0 + qnorm(1-alpha)*sqrt(p0*(1-p0)/N)

# Now the power is the area under the distribution for the alternative hypothesis to the
# right of the critical value

# We need to find this for each of the curves with the different p-values
(1 - pnorm(pk, mean = p, sd = sqrt(p*(1-p)/N)))
```

```
## [1] 0.01000000 0.04142679 0.12487206 0.28258891 0.49759934 0.71135888
```