Assignment1024

Suz

2019 10 24

1. Children’s IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. What proportion of children are expected to have an IQ between 80 and 120?

mean=100 std=15 lowerbound=80 upperbound=120

#probability from a normal distribution  
mean=100  
std=15  
lowerbound=80  
upperbound=120  
  
pnorm(upperbound, mean, std, lower.tail=TRUE)- pnorm(lowerbound, mean, std, lower.tail=TRUE)

## [1] 0.8175776

1. Generate data for hypothesis testing (student’s T-test) and show me one exmaple each of a significant result and a non-significant result. Explain your thinking.

Data from: <http://fcon_1000.projects.nitrc.org/indi/adhd200/>

raw\_adhd<-read.csv(file="adhd200.csv")  
df\_adhd<-data.frame(raw\_adhd)  
an\_children<-df\_adhd[df\_adhd$DX != "pending",] #disregard pending diagnosis  
girls<-an\_children[an\_children$Gender==0,]  
girls\_not<-girls[girls$DX==0,]  
girls\_diagnosed <- girls[girls$DX!=0,]  
boys<-an\_children[an\_children$Gender==1,]  
boys\_not<-boys[boys$DX==0,]  
boys\_diagnosed <- boys[boys$DX!=0,]  
  
  
#girls\_diagnosed  
#boys\_diagnosed  
  
boys\_age <-boys\_diagnosed$Age  
girls\_age<-girls\_diagnosed$Age

T test for the age of girls and boys who were diagnosed:

res1<-t.test(boys\_age, girls\_age, var.equal = TRUE)  
res1

##   
## Two Sample t-test  
##   
## data: boys\_age and girls\_age  
## t = 2.0376, df = 75, p-value = 0.04511  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.03977795 3.51951617  
## sample estimates:  
## mean of x mean of y   
## 12.09200 10.31235

Significance of comparing the age of girls and boys who were diagnosed:

res1$p.value < 0.05

## [1] TRUE

This implies that the difference between the mean of the two dataset can be significant. However, shapiro test showed that the boys data was not normally distributed, which could show a wrong result.

T test for the age boys who were diagnosed with adhd or boys who were not:

res2<-t.test(boys\_not$Age, boys\_diagnosed$Age, var.equal=TRUE)  
res2

##   
## Two Sample t-test  
##   
## data: boys\_not$Age and boys\_diagnosed$Age  
## t = -1.5955, df = 104, p-value = 0.1136  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -2.2654290 0.2453421  
## sample estimates:  
## mean of x mean of y   
## 11.08196 12.09200

Significance of comparing the age of boys diagnosed with adhd, and typical for age:

res2$p.value <0.05

## [1] FALSE

This implies that the difference between the ages of boys who were diagnosed and who were not is not siginificant, which is most likely the result of using a dataset that was focusing on a certain age interval of children.