R Notebook

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

# 6.118

## Planning another test to compare consumption.

### A) Solution:

#Here, level of significance  
a = .05  
  
# As it is a one-tailed test  
qnorm(a)

## [1] -1.644854

*So, if Z < -1.644854 then we can reject the null hypothesis*

### B) Solution:

#Here  
n = 100 # sample size  
m = 286 # mean  
S = 155 # standard deviation of the population  
s = S/sqrt(n) # standard deviation of the sample  
a = .05 # level of significance  
z = qnorm(a)  
# As Z = (x -mean)/standard deviation  
# So, x = Z\* standard deviation + mean  
  
x\_b = z\*s+m  
x\_b

## [1] 260.5048

*So, if mean of sample is less than 260.5048 then we can reject the null hypothesis*

### C) Solution:

#Here  
n = 100 # sample size  
m = 271 # alternate mean  
S = 155 # standard deviation of the population  
s = S/sqrt(n) # standard deviation of the sample  
x\_c = x\_b  
# As Z = (x bar-mean)/standard deviation  
z = (x\_c-m)/s  
# As it is a one-tailed test  
pnorm(z)

## [1] 0.2491675

*the probability that x¯ will fall in the region defined in part (b) is 0.2491675*

### D) Solution:

# Here  
m0 = 286 #old mean  
m1 = 271 # new mean  
a = .05 # level of significance   
S = 155 # standard deviation of the population  
n = 100 # sample size  
s = S/sqrt(n) # standard deviation of the sample  
z1 = qnorm(1-a)  
# We know that Power = P[Z > Z1 − |(m1 − m0)| / (S / √n)]  
p = 1- pnorm(z1-(abs(m1-m0)/s))  
p

## [1] 0.2491675

*So the power for n=100 is pretty low (.249) and not adequate for which we need bigger sample size.*

### E) Solution:

#Here  
m0 = 286 #old mean  
m1 = 271 # new mean  
S = 155 # standard deviation of the population  
a = .05 # level of significance   
z1 = qnorm(1-a)  
p = .8 # desired power  
z2 = qnorm(p)  
z1

## [1] 1.644854

z2

## [1] 0.8416212

# we know sample size, n = (z1+z2)^2\*S^2/(m0-m1)^2  
  
n = (z1+z2)^2\*S^2/(m1-m0)^2  
n

## [1] 660.1597

*So we would need a sample size of 660 (or 661).* *The results were validated by Statistical applet* [*https://digitalfirst.bfwpub.com/stats\_applet/stats\_applet\_9\_power.html*](https://digitalfirst.bfwpub.com/stats_applet/stats_applet_9_power.html)

# 6.124

## Stress by occupation.

### A) Solution:

df = data.frame("Occupation" = c("Professional", "Managerial", "Administrative","Sales", 'Mechanical', 'Service', 'Operator', 'Farm'),   
 "p" =c(.23,.22,.17,.15,.12,.13,.12,.08),"n"=c(2447,2552,2309,1811,1979,2592,2782,571))  
df

## Occupation p n  
## 1 Professional 0.23 2447  
## 2 Managerial 0.22 2552  
## 3 Administrative 0.17 2309  
## 4 Sales 0.15 1811  
## 5 Mechanical 0.12 1979  
## 6 Service 0.13 2592  
## 7 Operator 0.12 2782  
## 8 Farm 0.08 571

# with 95% confidence level for a two tailed test   
a = .05  
# So confidence level c = 1 - (a/2)  
c = 1 - (a/2)  
c

## [1] 0.975

# We know that CI (Proportion) = (P^ - Zc\*sqrt(p^(1-p^)/n), P^ + Zc\*sqrt(p^(1-p^)/n))  
df$CI\_left = df$p-qnorm(c)\*sqrt(df$p\*(1-df$p)/df$n)   
df$CI\_right = df$p+qnorm(c)\*sqrt(df$p\*(1-df$p)/df$n)  
df

## Occupation p n CI\_left CI\_right  
## 1 Professional 0.23 2447 0.21332598 0.2466740  
## 2 Managerial 0.22 2552 0.20392813 0.2360719  
## 3 Administrative 0.17 2309 0.15467856 0.1853214  
## 4 Sales 0.15 1811 0.13355462 0.1664454  
## 5 Mechanical 0.12 1979 0.10568283 0.1343172  
## 6 Service 0.13 2592 0.11705322 0.1429468  
## 7 Operator 0.12 2782 0.10792460 0.1320754  
## 8 Farm 0.08 571 0.05774801 0.1022520

# So the confidence intervals are  
print(paste('(',df$CI\_left,',',df$CI\_right,')'))

## [1] "( 0.213325977356449 , 0.246674022643551 )"   
## [2] "( 0.203928130486239 , 0.236071869513761 )"   
## [3] "( 0.154678558354342 , 0.185321441645658 )"   
## [4] "( 0.13355461828991 , 0.16644538171009 )"   
## [5] "( 0.105682830946921 , 0.134317169053079 )"   
## [6] "( 0.117053216225026 , 0.142946783774974 )"   
## [7] "( 0.107924601069392 , 0.132075398930608 )"   
## [8] "( 0.0577480054927639 , 0.102251994507236 )"

### B) Solution:

# five-number summary   
summary(df)

## Occupation p n CI\_left   
## Length:8 Min. :0.0800 Min. : 571 Min. :0.05775   
## Class :character 1st Qu.:0.1200 1st Qu.:1937 1st Qu.:0.10736   
## Mode :character Median :0.1400 Median :2378 Median :0.12530   
## Mean :0.1525 Mean :2130 Mean :0.13674   
## 3rd Qu.:0.1825 3rd Qu.:2562 3rd Qu.:0.16699   
## Max. :0.2300 Max. :2782 Max. :0.21333   
## CI\_right   
## Min. :0.1023   
## 1st Qu.:0.1338   
## Median :0.1547   
## Mean :0.1683   
## 3rd Qu.:0.1980   
## Max. :0.2467

*Mechanical and operator appear to have similar stress levels. Moreover, managerial and professionals also seem to have pretty similar stress levels.*

### C) Solution:

*My friend might be concerned with whether np>= 10 and n(1-p)>= 10 is satisfied or not.*

# 6.128

## Effect of sample size on significance.

### Solution:

# Here  
m0 = 0 # null hypothesis mean  
xbar = 4 # new mean  
S = 14 # standard deviation  
a = .05 # level of significance

# for n = 10  
n = 10  
s = S/sqrt(n)  
print(paste("Z-value =",(xbar-m0)/s))

## [1] "Z-value = 0.903507902905251"

print(paste("P-value =",pnorm((xbar-m0)/s, lower.tail = FALSE)))

## [1] "P-value = 0.183128197912392"

# for n = 20  
n = 20  
s = S/sqrt(n)  
print(paste("Z-value =",(xbar-m0)/s))

## [1] "Z-value = 1.27775312999988"

print(paste("P-value =",pnorm((xbar-m0)/s, lower.tail = FALSE)))

## [1] "P-value = 0.1006682426437"

# for n = 30  
n = 30  
s = S/sqrt(n)  
print(paste("Z-value =",(xbar-m0)/s))

## [1] "Z-value = 1.5649215928719"

print(paste("P-value =",pnorm((xbar-m0)/s, lower.tail = FALSE)))

## [1] "P-value = 0.0588006474973132"

# for n = 40  
n = 40  
s = S/sqrt(n)  
print(paste("Z-value =",(xbar-m0)/s))

## [1] "Z-value = 1.8070158058105"

print(paste("P-value =",pnorm((xbar-m0)/s, lower.tail = FALSE)))

## [1] "P-value = 0.0353799074015287"

# for n = 50  
n = 50  
s = S/sqrt(n)  
print(paste("Z-value =",(xbar-m0)/s))

## [1] "Z-value = 2.02030508910442"

print(paste("P-value =",pnorm((xbar-m0)/s, lower.tail = FALSE)))

## [1] "P-value = 0.0216758756304313"

*With increasing sample size, p-values decrease if we keep the sample mean the same in all cases. This shows that the increase in sample size gives us more and more evidence against null hypothesis and hence increases the probability of rejecting null hypothesis. For example, in the case above, for the sample size 10,20 and 30 , we will not reject null because p-value > 0.05 but for the case of 40 and 50, we will reject null.*

# 6.129

## Blood phosphorus level in dialysis patients.

### A) Solution:

arr = c(5.4,5.2,4.5,4.9,5.7,6.3)  
xbar = mean(arr)  
xbar

## [1] 5.333333

# Here  
S = .9 # standard deviation of the population  
n = length(arr) # sample size  
n

## [1] 6

# We know that CI (Proportion) = (P^ - Zc\*sqrt(p^(1-p^)/n), P^ + Zc\*sqrt(p^(1-p^)/n))  
# With 95% confidence level for a two-tailed test   
a = .05  
# and confidence level c = 1 - a/2  
c = 1 - a/2  
c

## [1] 0.975

# So, Confidence interval for the mean blood phosphorus level  
print(paste('(',xbar-qnorm(c)\*S/sqrt(n),',',xbar+qnorm(c)\*S/sqrt(n),')'))

## [1] "( 4.61319658188004 , 6.05347008478663 )"

### B) Solution:

*Since the confidence interval from part (a) includes 4.8 and the minimum value of possible mean is less than 4.8, there is not sufficient evidence that this patient has a mean phosphorus level that exceeds 4.8.*

# 6.130

## Cellulose content in alfalfa hay.

### A) Solution:

# We know that CI (Proportion) = (P^ - Zc\*sqrt(p^(1-p^)/n), P^ + Zc\*sqrt(p^(1-p^)/n))  
# With 90% confidence level for a two-tailed test   
a = .1  
# and confidence level c = 1 - a/2  
c = 1 - a/2  
c

## [1] 0.95

# Here  
S = 8 # standard deviation of the population  
xbar = 145 # mean  
n = 15 # sample size

# So, Confidence interval for the mean cellulose content in the population  
print(paste('(',xbar-qnorm(c)\*S/sqrt(n),',',xbar+qnorm(c)\*S/sqrt(n),')'))

## [1] "( 141.602404957796 , 148.397595042204 )"

### B) Solution:

*Here H0: m = 140* *H1: m > 140*

# So  
m = 140 # mean of the null hypothesis  
# We know that test statics Z = (xbar - m)/(S/sqrt(n))  
Z = (xbar - m)/(S/sqrt(n))  
Z

## [1] 2.420615

# So P- value of P(Z > 2.420615)  
  
pnorm(Z, lower.tail = FALSE)

## [1] 0.007747147

*As the p-value is less than level of significance we can reject the null hypothesis.*

### C) Solution:

*The main assumption we are taking is that the underlying population is normal and standard deviation is 8 mg/g.* *The second assumption we need to make that all the 15 sample are taken randomly, so that they can be independent.*

# 6.133

## CEO pay.

### A) Solution:

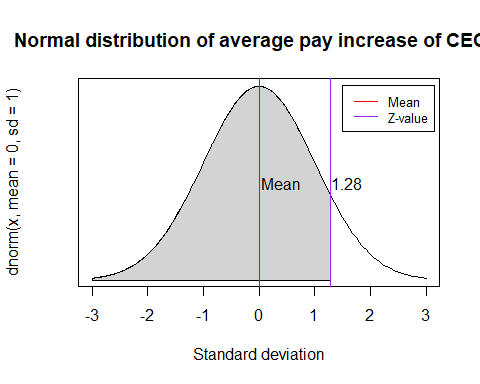
# Here  
xbar = 6.9 # mean of the sample  
S = 55 # Standard deviation  
n = 104 # Sample size  
m = 0 # mean of null hypothesis

Z = (xbar-m)/(S/sqrt(n))  
Z

## [1] 1.27939

# function for shading area in a graph  
colorArea <- function(from, to, density, ..., col="lightgray", dens=NULL){  
 y\_seq <- seq(from, to, length.out=500)  
 d <- c(0, density(y\_seq, ...), 0)  
 polygon(c(from, y\_seq, to), d, col=col, density=dens)  
}

curve(dnorm(x, mean = 0, sd = 1), from = -3, to = 3, yaxt = 'n', xlab = "Standard deviation", main= "Normal distribution of average pay increase of CEOs ")  
colorArea(from=-3, to=Z, dnorm)  
text(Z+.3,0.2, round(Z,2))  
text(0.4,0.2, "Mean")  
abline(v=0, col="red")  
abline(v=Z, col="purple")  
legend(1.5, .4, legend=c("Mean", "Z-value"),  
 col=c("red", "purple"), lty=1:1, cex=0.8)



#axis(1, at = c(-3,-2,-1,0,1,2,3), labels = c('-3S', '-2S', '-S','mean','S','2S','3S'))

### B) Solution:

# So the P-value is  
P = pnorm(Z, lower.tail = FALSE)  
P

## [1] 0.1003798

### C) Solution:

*As the P-value is greater than a = 0.05, we cannot reject the null hypothesis. Moreover, the study does not give strong evidence that the mean compensation of all CEOs went up.*