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.

library(plotrix)

# Problem 1.169 How much vitamin C do you need?

The Food and Nutrition Board of the Institute of Medicine working in cooperation with scientists

from Canada have used scientific data to answer this question for a variety of vitamins and

minerals.47 Their methodology assumes that needs, or requirements, follow a distribution. They

have produced guidelines called dietary reference intakes for different gender-by-age combinations. For vitamin C, there are three dietary reference intakes: the estimated average requirement (EAR), which is the mean of the requirement distribution; the recommended dietary allowance (RDA), which is the intake that would be sufficient for 97% to 98% of the population; and the tolerable upper level (UL), the intake that is unlikely to pose health risks. For women aged 19 to 30 years, the EAR is 60 milligrams per day (mg/d), the RDA is 75 mg/d, and the UL is 2000 mg/d.

(a) The researchers assumed that the distribution of requirements for vitamin C is Normal. The EAR gives the mean. From the definition of the RDA, let’s assume that its value is the 97.72 percentile. Use this information to determine the standard deviation of the requirement distribution.

(b) Sketch the distribution of vitamin C requirements for 19- to 30-year-old women. Mark the EAR, the RDA, and the UL on your plot.

# A) Solution:

# EAR  
mean = 60  
# percent of RDA (p-value)  
pz = .9772  
# RDA  
A = 75  
  
# find z-value from p-value  
z= round(qnorm(pz, lower.tail = TRUE),2)  
print(paste("z score for RDA is", z))

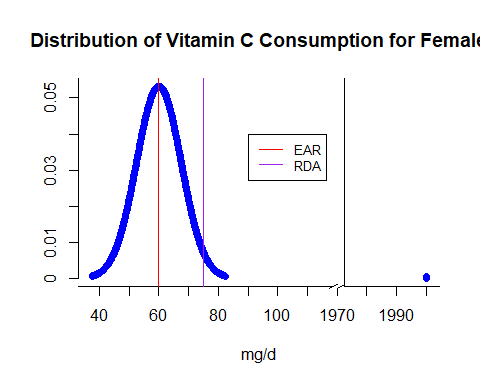
## [1] "z score for RDA is 2"

sd = (A-mean)/z  
print(paste("Standard deviation of the requirement distribution is",sd))

## [1] "Standard deviation of the requirement distribution is 7.5"

# B) Solution:

ul = 2000  
xmin = mean-sd\*3  
xmax = mean+sd\*3  
# Create a sequence of numbers between xmin and 2000 incrementing by 0.1  
x <- seq(xmin, ul, by = .1)  
xgap <- ifelse(x>xmax,2000, x)  
# Choose the mean as 60 and standard deviation as 7.5.  
y <- dnorm(x, mean, sd)  
par(bty="n") # deleting the box  
gap.plot(xgap,y, gap=c(120,1970), gap.axis="x", pch=16,  
 col="blue", ylim=range(y),  
 xtics=c(seq(30, 120, by = 10),seq(1970, 2010, by = 10)),   
 xticlab=c(seq(30, 120, by = 10),seq(1970, 2010, by = 10)), xlab = 'mg/d',   
 ylab ='', main="Distribution of Vitamin C Consumption for Female")  
abline(v=mean, col="red") #add EAR line  
abline(v=A, col="purple") # add RDA line  
abline(v=seq(120,1970,100), col="white") # hiding vertical lines  
axis.break(1,120,style="slash") # plotting slashes for breakpoints  
legend(90, .04, legend=c("EAR", "RDA"),  
 col=c("red", "purple"), lty=1:1, cex=0.8)



# Problem 1.170 How much vitamin C do men need?

Refer to the previous exercise. For men aged 19 to 30 years, the EAR is 75 milligrams per day

(mg/d), the RDA is 90 mg/d, and the UL is 2000 mg/d. Answer the questions in the previous exercise for this population.

# A) Solution:

#EAR  
mean = 75  
# percent of RDA (p-value)  
pz = .9772  
# RDA  
A = 90  
  
#find z-value from p-value  
z= round(qnorm(pz, lower.tail = TRUE),2)  
print(paste("z score for RDA is", z))

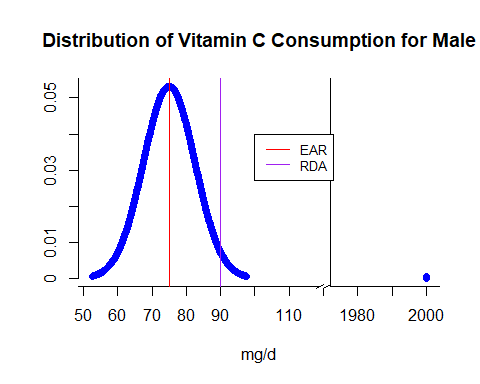
## [1] "z score for RDA is 2"

sd = (A-mean)/z  
print(paste("Standard deviation of the requirement distribution is",sd))

## [1] "Standard deviation of the requirement distribution is 7.5"

# B) Solution:

ul = 2000  
xmin = mean-sd\*3  
xmax = mean+sd\*3  
# Create a sequence of numbers between xmin and 2000 incrementing by 0.1  
x <- seq(xmin, ul, by = .1)  
xgap <- ifelse(x>xmax,2000, x)  
# Choose the mean as 60 and standard deviation as 7.5.  
y <- dnorm(x, mean, sd)  
par(bty="n") # deleting the box  
gap.plot(xgap,y, gap=c(120,1970), gap.axis="x", pch=16,  
 col="blue", ylim=range(y),  
 xtics=c(seq(30, 120, by = 10),seq(1970, 2010, by = 10)),   
 xticlab=c(seq(30, 120, by = 10),seq(1970, 2010, by = 10)), xlab = 'mg/d',   
 ylab ='', main="Distribution of Vitamin C Consumption for Male")  
abline(v=mean, col="red") #add EAR line  
abline(v=A, col="purple") # add RDA line  
abline(v=seq(120,1970,100), col="white") # hiding vertical lines  
axis.break(1,120,style="slash") # plotting slashes for breakpoints  
legend(100, .04, legend=c("EAR", "RDA"),  
 col=c("red", "purple"), lty=1:1, cex=0.8)



# Problem 1.171 How much vitamin C do women consume?

To evaluate whether or not the intake of a vitamin or mineral is adequate, comparisons are made between the intake distribution and the requirement distribution. Here is some information about the distribution of vitamin C intake, in milligrams per day, for women aged 19 to 30 years:

Table

Description automatically generated

(a) Use the 5th, the 50th, and the 95th percentiles of this distribution to estimate the mean and

standard deviation of this distribution assuming that the distribution is Normal. Explain your method

for doing this.

(b) Sketch your normal intake distribution on the same graph with a sketch of the requirement

distribution that you produced in part (b) of Exercise 1.69.

(c) Do you think that many women aged 19 to 30 years are getting the amount of vitamin C that they need? Explain your answer.

# A) Solution:

# percent of RDA (p-value)  
pz5 = 0.05  
  
  
# find z-value from p-value  
z5= round(qnorm(pz5),2)  
print(paste("z score for 5th percentile is", z5))

## [1] "z score for 5th percentile is -1.64"

# percent of RDA (p-value)  
pz50 = 0.50  
# RDA  
A50 = 79  
# find z-value from p-value  
z50= round(qnorm(pz50, lower.tail = TRUE),2)  
print(paste("z score for 50th percentile is", z50))

## [1] "z score for 50th percentile is 0"

# percent of RDA (p-value)  
pz95 = 0.95  
  
# find z-value from p-value  
z95= round(qnorm(pz95, lower.tail = TRUE),2)  
print(paste("z score for 95th percentile is", z95))

## [1] "z score for 95th percentile is 1.64"

We know z = (x-mean)/sd

So, mean = x-(sd\*z)

But, for 50th percentile, z is 0 as p(z<=0.00) = .50

So, mean = A50-(sd\*0)

Finally, mean = A50

mean = A50  
print(paste("The value of mean is",mean))

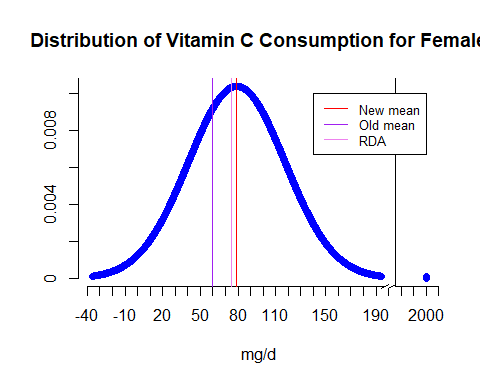
## [1] "The value of mean is 79"

A95 = 142  
sd = (A95-mean)/z95  
print(paste("The value of standard deviation is",sd))

## [1] "The value of standard deviation is 38.4146341463415"

# B) Solution:

ul = 2000  
xmin = mean-sd\*3  
xmax = mean+sd\*3  
# Create a sequence of numbers between xmin and 2000 incrementing by 0.1  
x <- seq(xmin, ul, by = .1)  
xgap <- ifelse(x>xmax,2000, x)  
# Choose the mean as 60 and standard deviation as 7.5.  
y <- dnorm(x, mean, sd)  
par(bty="n") # deleting the box  
gap.plot(xgap,y, gap=c(200,1970), gap.axis="x", pch=16,  
 col="blue", ylim=range(y),  
 xtics=c(seq(-40, 200, by = 10),seq(1970, 2010, by = 10)),   
 xticlab=c(seq(-40, 200, by = 10),seq(1970, 2010, by = 10)), xlab = 'mg/d',   
 ylab ='', main="Distribution of Vitamin C Consumption for Female")  
abline(v=mean, col="red") #add new mean line  
abline(v=60, col="purple") # add old mean line  
abline(v=75, col="violet") # add RDA line  
abline(v=seq(200,1970,100), col="white") # hiding vertical lines  
axis.break(1,200,style="slash") # plotting slashes for breakpoints  
legend(140, .01, legend=c("New mean", "Old mean", "RDA"),  
 col=c("red", "purple", "violet"), lty=1:1, cex=0.8)



# C) Solution:

*As per the new graph, we can see that the new mean (79) is more than the previously estimated mean (60) and the new mean (79) which is at 50th percentile is also more than the RDA (75). So we can confidently say that majority of the female population consumes more vitamin C than the RDA on a daily basis.*

# Problem 1.172 How much vitamin C do men consume?

To evaluate whether or not the intake of a vitamin or mineral is adequate, comparisons are made

between the intake distribution and the requirement distribution. Here is some information about the distribution of vitamin C intake, in milligrams per day, for men aged 19 to 30 years:

Table

Description automatically generated

(a) Use the 5th, the 50th, and the 95th percentiles of this distribution to estimate the mean and

standard deviation of this distribution assuming that the distribution is Normal. Explain your method for doing this.

(b) Sketch your normal intake distribution on the same graph with a sketch of the requirement

distribution that you produced in Exercise 1.70.

(c) Do you think that many men aged 19 to 30 years are getting the amount of vitamin C that they

need? Explain your answer.

# A) Solution:

# percent of RDA (p-value)  
pz5 = 0.05  
  
  
# find z-value from p-value  
z5= round(qnorm(pz5),2)  
print(paste("z score for 5th percentile is", z5))

## [1] "z score for 5th percentile is -1.64"

# percent of RDA (p-value)  
pz50 = 0.50  
# RDA  
A50 = 114  
# find z-value from p-value  
z50= round(qnorm(pz50, lower.tail = TRUE),2)  
print(paste("z score for 50th percentile is", z50))

## [1] "z score for 50th percentile is 0"

# percent of RDA (p-value)  
pz95 = 0.95  
  
# find z-value from p-value  
z95= round(qnorm(pz95, lower.tail = TRUE),2)  
print(paste("z score for 95th percentile is", z95))

## [1] "z score for 95th percentile is 1.64"

We know z = (x-mean)/sd

So, mean = x - (sd\*z)

But, for 50th percentile, z is 0 as p(z<=0.00) = .50

So, mean = A50-(sd\*0)

Finally, mean = A50

mean = A50   
print(paste("The value of mean is",mean))

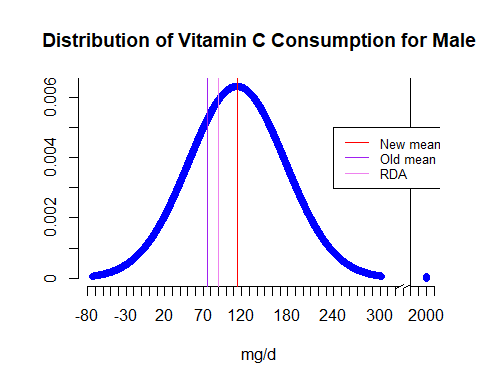
## [1] "The value of mean is 114"

A95 = 217  
sd = (A95-mean)/z95  
print(paste("The value of standard deviation is",sd))

## [1] "The value of standard deviation is 62.8048780487805"

# B) Solution:

ul = 2000  
xmin = mean-sd\*3  
xmax = mean+sd\*3  
# Create a sequence of numbers between xmin and 2000 incrementing by 0.1  
x <- seq(xmin, ul, by = .1)  
xgap <- ifelse(x>xmax,2000, x)  
# Choose the mean as 60 and standard deviation as 7.5.  
y <- dnorm(x, mean, sd)  
par(bty="n") # deleting the box  
gap.plot(xgap,y, gap=c(330,1970), gap.axis="x", pch=16,  
 col="blue", ylim=range(y),  
 xtics=c(seq(-80, 330, by = 10),seq(1970, 2010, by = 10)),   
 xticlab=c(seq(-80, 330, by = 10),seq(1970, 2010, by = 10)), xlab = 'mg/d',   
 ylab ='', main="Distribution of Vitamin C Consumption for Male")  
abline(v=mean, col="red") #add new mean line  
abline(v=75, col="purple") # add old mean line  
abline(v=90, col="violet") # add RDA line  
abline(v=seq(330,1970,100), col="white") # hiding vertical lines  
axis.break(1,330,style="slash") # plotting slashes for breakpoints  
legend(240, .005, legend=c("New mean", "Old mean", "RDA"),  
 col=c("red", "purple", "violet"), lty=1:1, cex=0.8)



# C) Solution:

*As per the new graph, we can see that the new mean (114) is more than the previously estimated mean (75) and the new mean (114) which is at 50th percentile is also more than the RDA (90). So we can confidently say that majority of the male population consumes more vitamin C than the RDA on a daily basis.*

# Problem 1.176 Norms for reading scores.

Raw scores on behavioral tests are often transformed for easier comparison. A test of reading ability has mean 70 and standard deviation 10 when given to third-graders. Sixth-graders have mean score 80 and standard deviation 11 on the same test. To provide separate “norms” for each grade, we want scores in each grade to have mean 100 and standard deviation 20.

(a) What linear transformation will change third-grade scores *x* into new scores *x*new = *a* + *bx* that have the desired mean and standard deviation? (Use *b* > 0 to preserve the order of the scores.)

(b) Do the same for the sixth-grade scores.

(c) David is a third-grade student who scores 72 on the test. Find David’s transformed score. Nancy is a sixth-grade student who scores 78. What is her transformed score? Who scores higher within his or her grade?

(d) Suppose that the distribution of scores in each grade is Normal. Then both sets of transformed scores have the *N*(100, 20) distribution. What percent of third-graders have scores less than 75? What percent of sixth-graders h­­­­­­ave scores less than 75?

# A) Solution:

For third graders’ current scores:

μ=70

σ=10

z = (x-μ)/σ

=>z = (x-70)/10 …..(1)

For desired score

μ=100

σ=20

so X-new =z\*σ +μ

=[(x-70)/10]\*20 +100 ......(from 1)  
  
 =2x-140+100  
  
 =2x-40

So, a=-40 and b=2

# B) Solution:

For six graders’ current scores:

μ=80

σ=11

z = (x-μ)/σ

=>z = (x-80)/11 …..(1)

For desired score

μ=100

σ=20

so X-new =z\*σ +μ

=[(x-80)/11]\*20 +100 ......(from 1)  
  
 =(20x-1600+1100)/11  
   
 =(20x-500)/11  
  
 =1.82x-45.45

So, a=-45.45 and b=1.82

# C) Solution:

David’s transformed score =-40+2\*72=104

Nancy’s transformed score = -45.45+1.82\*78=96.51

David Scores higher as his standardized transformed score is higher.

# D) Solution:

For both third and six graders standardized scoring less than 75 as they have the same distribution:

mean = 100  
sd = 20  
x = 75  
z = (x-mean)/sd  
z

## [1] -1.25

pz = pnorm(z, lower.tail = TRUE)  
print(paste("Percent value is", round(pz\*100,2),"%"))

## [1] "Percent value is 10.56 %"

For non-standardized scoring:

# for third graders  
mean = 70  
sd = 10  
x = 75  
z = (x-mean)/sd  
z

## [1] 0.5

pz = pnorm(z, lower.tail = TRUE)  
print(paste("Percent value of third graders is", round(pz\*100,2),"%"))

## [1] "Percent value of third graders is 69.15 %"

# for sixth graders  
mean = 80  
sd = 11  
x = 75  
z = (x-mean)/sd  
z

## [1] -0.4545455

pz = pnorm(z, lower.tail = TRUE)  
print(paste("Percent value of sixth graders is", round(pz\*100,2),"%"))

## [1] "Percent value of sixth graders is 32.47 %"