1. **What’s the probability of**
   1. rolling a 6 on a die

Ans. 1/6

* 1. rolling three consecutive times a 6

Ans. 1/216

* 1. rolling a 6, a 1 and a 6 again

Ans. 1/216

1. **Define categories of random variables and give at least 2 examples for each group**
   1. Continuous

Ans. It can take infinite many values eg – height, weight

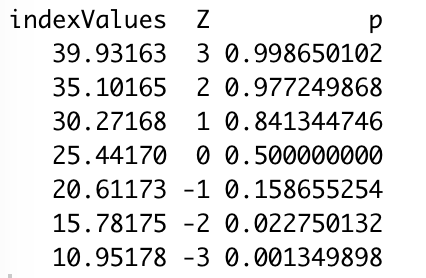
* 1. Categorical

Ans. It can take fixed number of possible values eg- Blood Group, Skin Colour

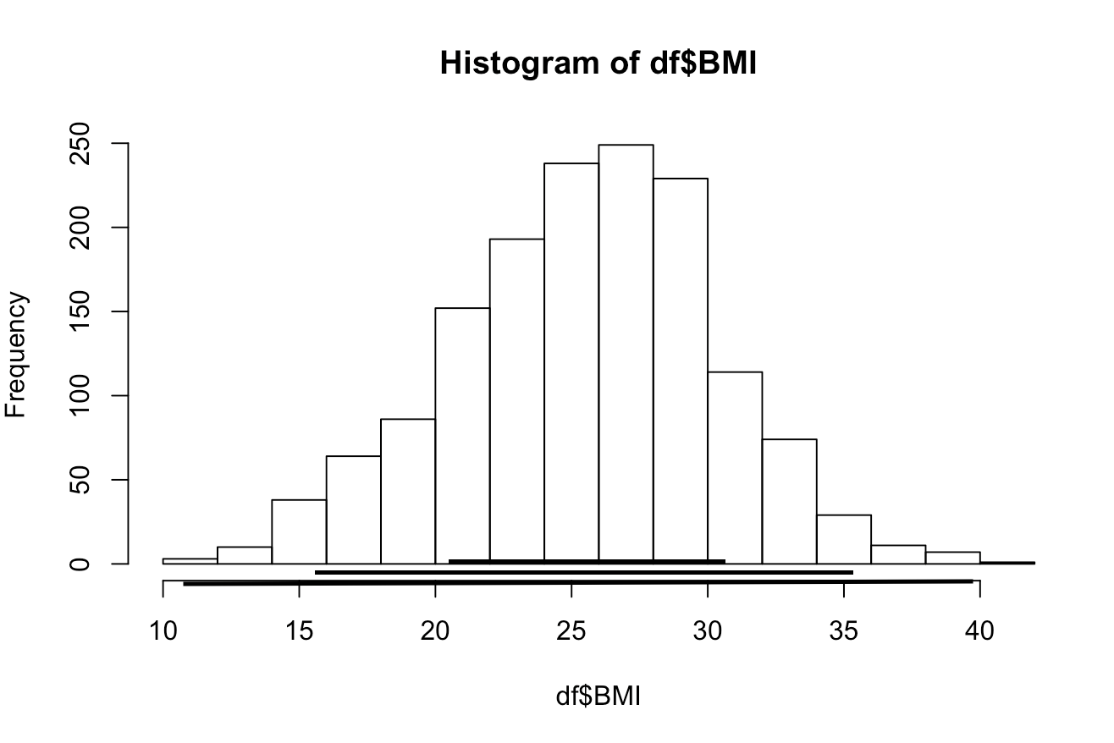
* 1. Discrete

Ans it has countable number of possible values eg – number of hospitalization days, test scores

1. **Normal distribution**
   1. Calculate all index points for the empirical rule



* 1. Plot the normal distribution for BMI as a histogram in R

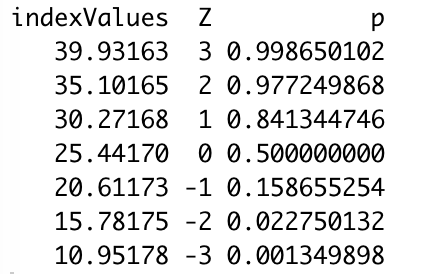


* 1. Indicate +/-1, 2 and 3 standard deviations in the histogram

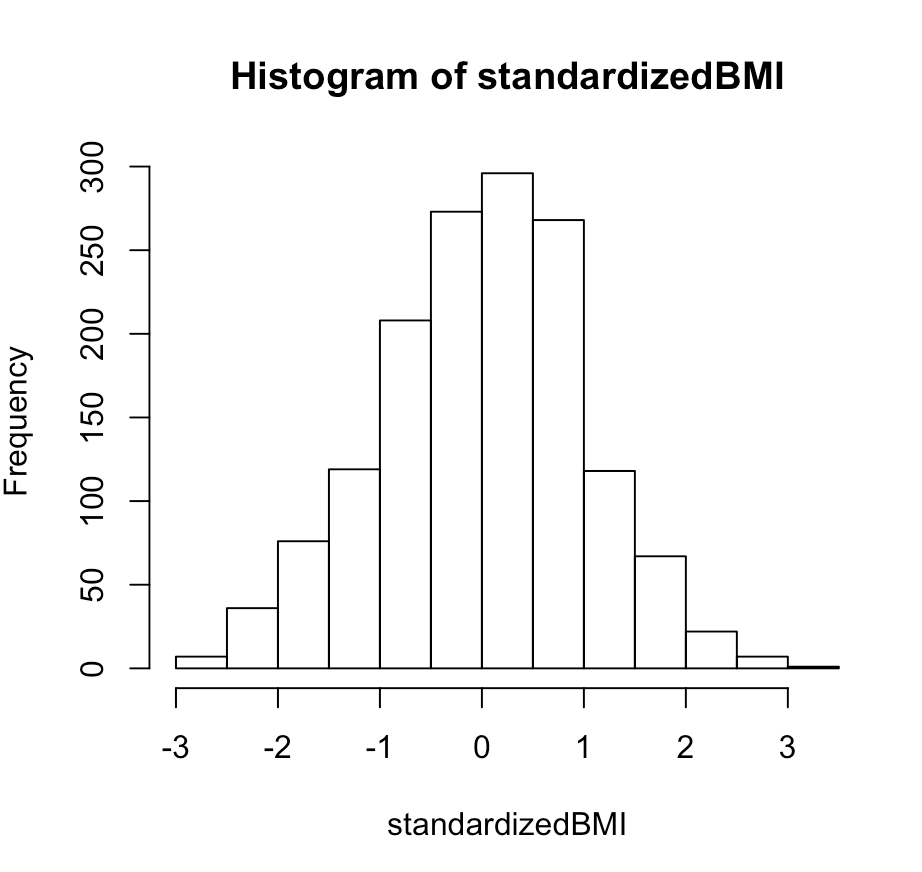
Ans

|  |  |
| --- | --- |
| One | 20.61173 - 30.27168 |
| Two | 15.78175 - 35.10165 |
| Three | 10.95178 - 39.93163 |

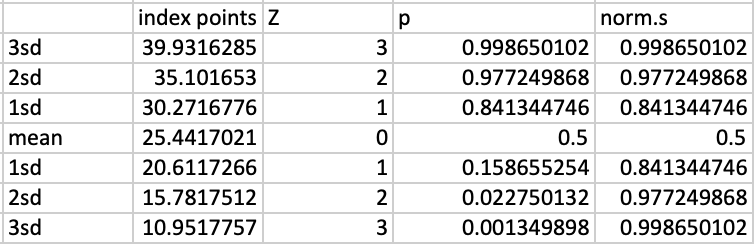
* 1. Calculate the above stated values, the z-scores and the probabilities



* 1. Create a standard normal distribution and plot the z-scores as a histogram



* 1. Use the norm.dist and norm.s.dist functions in Excel to confirm R’s computations of the Empirical rule (3a)



1. **Calculate the following probabilities**
   1. Calculate in R and Excel the probability for the following examples

# 4 individuals enter a clinic, what’s the p that exactly 3 are diabetic (prevalence 10%)

dbinom(3, 4, .1, log = FALSE) # .0036

# 12 individuals enter a hair salon, what’s the p that exactly 4 have alopecia (prevalence 20% in the studied population)

dbinom(4, 12,.2, log = FALSE) # .1328756

# 15 individuals enter a dentist office, what’s the p that exactly 3 have dental cavity (prevalence 30% in the studied population)

dbinom(3, 15, .3, log = FALSE) #.1700402

# 15 individuals enter a dentist office, what’s the p that at least 3 have dental cavity (prevalence 30% in the studied population)

pbinom(3, 15, .3, log = FALSE, lower.tail = FALSE) # .7031321

**CODE**

# Clear the environment

rm(list = ls())

df <-read.csv("C:/BME/Sem 2/Biostat/Class 4/Participation Lab Class 4 - BMI.csv")

# parameters

mu <- mean(df$BMI)

sd <- sd(df$BMI)

# index points

indexPoints <- data.frame("indexValues" = c(mu + 3\*sd, mu + 2\*sd, mu + sd, mu, mu - sd, mu - 2\*sd, mu - 3\*sd),

"Z" = ((indexPoints$indexValues - mu)/sd), "p" = pnorm(indexPoints$Z))

ls(indexPoints)

# histogram of BMI

hist(df$BMI)

# +/- standard deviations

one <- c(mu - sd, mu + sd)

two <- c(mu - 2\*sd, mu + 2\*sd)

three <- c(mu - 3\*sd, mu + 3\*sd)

# standard normal

standardizedBMI <- (df$BMI - mu)/sd

hist(standardizedBMI)