Question 1

1. Given Information:
   Total number of balls in a package: 12
   Probability that a ball is overweight from the pack of 12 tennis balls P (X =Overweight): 0.1

   To Find:
   A. Probability that 2 or more ball will be overweight.
   B. Random Variable (RV)
   C. Distribution of the Random Variable

   A. Probability that 2 or more ball will be overweight $P(X \geq 2)$:
      As we know that Random Variable (RV) is the number of overweight balls. So, it's a discrete Random Variable as number of balls can be whole number only.

      As it is a Discrete Random Variable, there can be two types of distribution:
      1. Binomial Distribution
      2. Poisson Distribution
      As this RV does not have interval factor in it so it cannot be Poisson Distribution.
      But to be a Binomial Distribution RV needs to satisfy four conditions:
      1. There should be 'n' identical trials.
      2. There are only two possible outcomes on each trial: Success or Failure.
      3. The probability of success should be same in each trial.
      4. All the 'n' trials should be independent.

      This RV satisfies all the above conditions. So, it is a Binomial Distribution.

      Using Binomial Distribution,
      X~ $Bin(n, p) = X \sim Bin\ (12, 0.1)$

      The probability distribution of X is $P(X = x) = {}^{n}_{x}C * p^{x} * (1-p)^{(n-x)}$, where p is the probability of success and n is the total number of trials.

      As per the given information, p = P (X =Overweight) = 0.1 and n = 12.

      Using R code:

      $P(X \geq 2) = 1 - P(X \leq 1)$

      $= 1 - pbinom\ (q = 1\ , size\ =\ 12\ , prob\ = 0.1)$

      OR

      $= pbinom\ (q = 1\ , size\ =\ 12\ , prob\ = 0.1, lower.tail\ =\ FALSE)$

Output:

```
1 - pbinom(q=1 ,size = 12 , prob =0.1)
pbinom(q=1 ,size = 12 , prob =0.1, lower.tail = FALSE)
```

```
> 1 - pbinom(q=1 ,size = 12 , prob =0.1)
[1] 0.3409977
> pbinom(q=1 ,size = 12 , prob =0.1, lower.tail = FALSE)
[1] 0.3409977
>
```

Answer:

A. Probability that 2 or more ball will be overweight $P(X \geq 2) = 0.3409977$

B. Random Variable (RV) is the number of overweight balls.

C. Distribution used is Binomial Distribution.

Question 2

2.

a. Histogram A is left skewed. So, as per the left skewness Mean < Median. By visualization, the interquartile data interval is from 15 to 25 and 20 seems to be the median data. Considering this analysis, mean should be between 15 and 20 intervals.
Now, Histogram B is slight left skewed. So, even here Mean < Median. By visualisation, Median should be around 10. As Mean < Median, which is equivalent to Mean < 10.
For Histogram A Mean is around 17.5 and for Histogram B Mean is less than 10. **So, Histogram A has larger mean.**
Ans: Histogram A has larger mean.

b. Looking at the x- intervals of the graph it can be deduced that Histogram B has wider intervals than Histogram A i.e., for histogram B the intervals are distributed in multiples of 10 while that of Histogram A are around 2.5. So, histogram A data are not varied much than that of histogram B. **Hence, Standard Deviation of Histogram B is greater than Standard Deviation of Histogram A.**
Ans: Histogram B has larger Standard Deviation.

c. Since histogram is a summary of the variation in data, comparing two histograms can be helpful if they have same interval i.e., the number of bins were same.
Bin Width $= \frac{\max value\ of\ data - \min value\ of\ data}{total\ number\ of\ bins}$
Selecting proper bin width is crucial as if fewer bins are captured then the histogram doesn't catch the distribution well.
If the number of bins were increased in histogram B then the comparison histogram A and B

would have been easy and distribution of data would have been visualized easily.

Question 3

3. For a process to be a Binomial process the random variable must satisfy all following conditions:
   a. There should be a finite number (n) of identical trails.
   b. There should be only two possible outcomes on each trail: Success or Failure.
   c. The probability of success (p) should be same in each trail.
   d. All the trails should be independent i.e., Outcome of one trial should not affect the outcome of other.
   So, using the above rules will validate the below two questions:
   1. One basketball player attempts 10 free throws, and the number of successful attempts is totalled:
      a. There should be a finite number (n) of identical trails: **There will be 10 free throws and 10 is a finite trial.**
      b. There should be only two possible outcomes on each trail: Success or Failure: **In each of the 10 free throws or 10 trials, only two possibilities that free throw is done or not.**
      c. The probability of success (p) should be same in each trail: **As each trial is independent and each trial can be free throw or not a free throw, the probability will be same in each trial or throw.**
      d. All the trails should be independent i.e., Outcome of one trial should not affect the outcome: **Each trial or throw will have its own outcomes and will not have any effect on other outcome. So, it's an independent event.**

      All the conditions of a Binomial Process are satisfied. Hence, **the process is a Binomial Process**.

   2. Ten Different basketball players each attempt 1 free throw and the total number of successful attempts is totalled:
      a. There should be a finite number (n) of identical trails: **There will be 10 free throws but those are not identical as 10 different players are involved. So this rule doesn't hold true.**
      b. There should be only two possible outcomes on each trail: Success or Failure: **In each of the 10 free throws or 10 trials, only two possibilities that free throw is done or not.**
      c. The probability of success (p) should be same in each trail: **As each trial is independent and each trial can be free throw or not a free throw, the probability will be same in each trial or throw.**
      d. All the trails should be independent i.e., Outcome of one trial should not affect the outcome: **Each trial or throw will have its own outcomes and will not have any effect on other outcome. So, it's an independent event.**

      All the conditions of a Binomial Process are satisfied except the identical trial one. Hence, **the process is not a Binomial Process**.

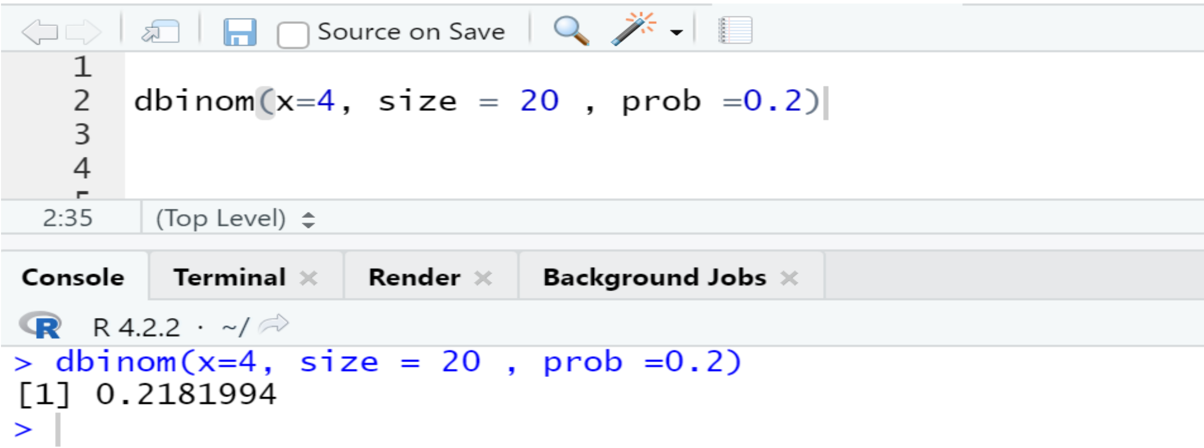
Question 4

4. $B \sim Bin(20{,}0.2)$: This means B follows Binomial Distribution with parameters n = 20 & p = 0.2

a. $P(B = 4)$:
Using R Code:

$P(B = 4) = dbinom(x = 4, size\ = \ 20\ ,prob\ = 0.2) = 0.2181994$

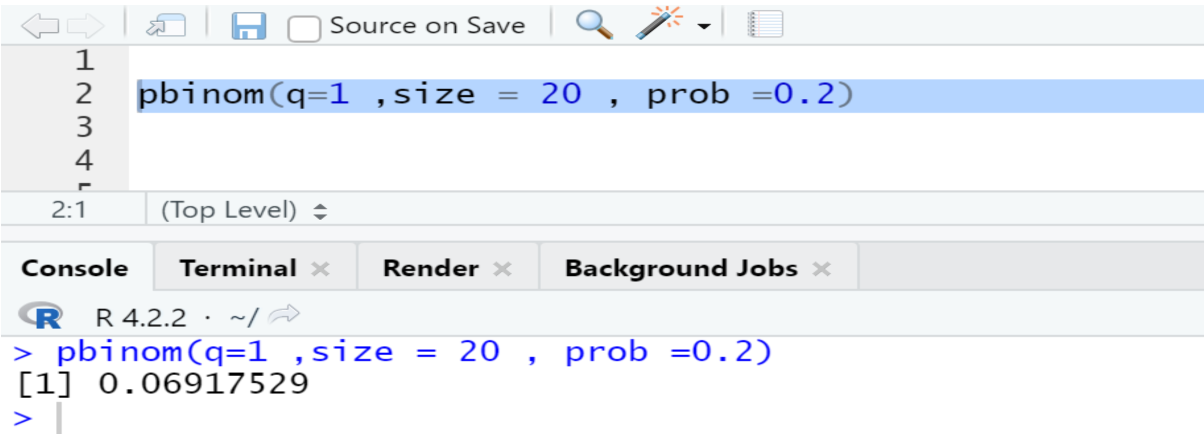
$\mathbf{P(B = 4\ ) = \ 0.2181994}$



b. $P(B \leq 1)$:
Using R Code:

$P(B \leq 1) = pbinom(x = 1, size\ = \ 20\ ,prob\ = 0.2) = 0.06917529$

$\boldsymbol{P(B\ \leq 1) = 0.06917529}$



c. $P\ (B\ > 1)$:

Using R Code:

$P(B > 1) = pbinom(x = 1, size\ = 20\ , prob\ = 0.2\ , lower.tail = FALSE) = 0.9308247$

$\boldsymbol{P(B > 1) = 0.9308247}$

```
2  pbinom(q=1 ,size = 20 , prob =0.2 , lower.tail = FALSE)
3
4
5
2:1   (Top Level) ÷
```

```
Console   Terminal ×   Render ×   Background Jobs ×
R 4.2.2 · ~/
> pbinom(q=1 ,size = 20 , prob =0.2 , lower.tail = FALSE)
[1] 0.9308247
>
```

d. E(B) and Var (B):

$B \sim Bin\ (n, p) \sim Bin\ (20{,}0.2)$

$E(B) = \mu = np$

$E(B) = \ \mu = 20 * 0.2 = 4$

$\boldsymbol{E(B) = \ \mu = 4}$

$Var(B) = \ \sigma^2 = npq = 20 * 0.2 * 0.8 = 3.2$

$\boldsymbol{Var(B) = \ \sigma^2 = npq = 3.2}$

Question 5

5. To find:

   Probability that Hussain receives more than two calls in the next 30 minutes.

   Given Information:

   Number of calls received in two hours i.e., in 120 mins. = 6

This is a Poisson distribution as we are measuring the probability of a certain event occurring in each interval of time.

X: Number of phone calls received in a fixed interval of time.

The Poisson Distribution of X is given by:

$$P(X = x) = \ e^{-\lambda} * \frac{\lambda^x}{x!}$$

X $\sim P(\lambda)\ or\ Po\ (\ \lambda)$

X follows a Poisson distribution with parameter λ.

In this problem statement, we need to find probability of more than 2 calls received in 30 mins. So, the time interval should be 30 mins. But in the information given there were 6 calls received in 2 hours. So, in 30 mins the number of calls received will be the mean λ =$\frac{6}{120} = \frac{6}{4*30} = \frac{1.5}{30}$.

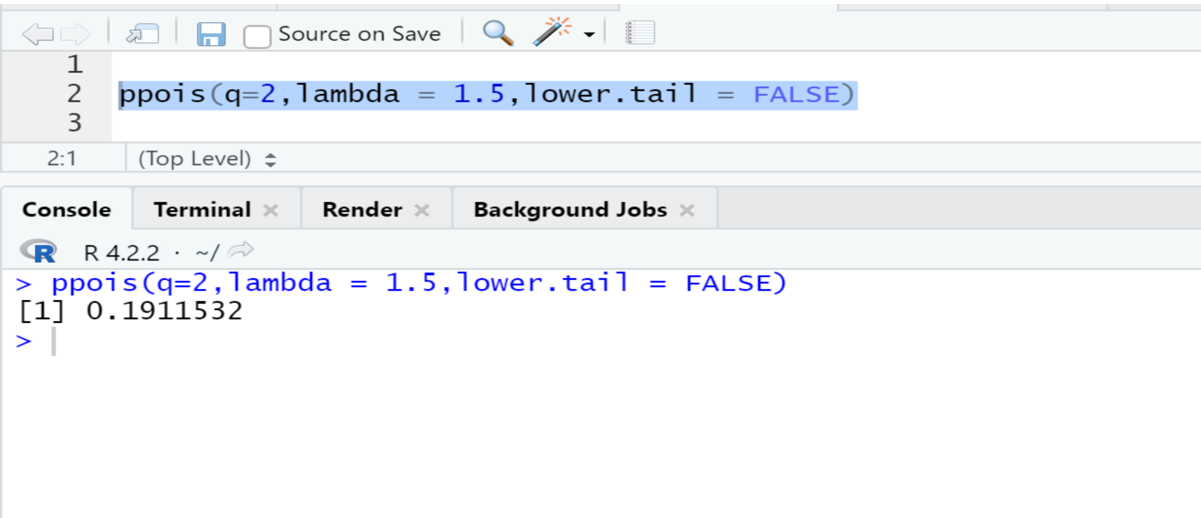
λ = 1.5 i.e., mean 1.5 calls received in 30 mins.

$X \sim P(\lambda = 1.5)$

Using R Code,

$P(X > 2) = 1 - P(X \leq 2)$

$P(X > 2) = ppois(q = 2, lambda = 1.5, lower.tail = FALSE)$



$\boldsymbol{P(X > 2) = 0.1911532}$

Probability that Hussain receives more than two calls in the next 30 minutes = 0.1911532

Question 6

6. Given Information:

   Mean = μ = 5
   Standard Deviation = σ = 3

   To Find: Probability that a single draw from the distribution has value 5 i.e., P (X = 5)

   X: let be the random variable
   $X\sim N\ (\mu\ ,\sigma^2\ ) \sim N(\ 5, 3^2\ )$

   The Probability of a given point or value in a Normal Distribution is always zero as Normal distribution is calculated for a Continuous RV for a range of values and the area under the curve for a point or a single value will always be zero.

   Probability that a single draw from this distribution has the values 5 is 0 Option A.

   Ans: Option A 0.

Question 7

7. Given information:
   $X1 \sim N(\mu\,,\sigma^2\,)$ and $X2 \sim N(\mu\,,\sigma^2\,)$

   To Find: Distributions of
   a. $2\,X1$
   b. $X1 + X2$

   a. $2\,X1$:
      Let Random Variable Q = 2 X1 then Q is also normally distributed as X1.
      Then Q is normally distributed with
      Mean E(Q) = 2 * μ = 2μ
      Variance V(Q) = $2^2 * \sigma^2$ = $4\sigma^2$
      Standard Deviation SD(Q) = 2σ

   b. $X1 + X2$:
      Let N = X1 + X2 be the random variable, which is also normally distributed as X1 & X2.
      Then N is normally distributed with
      Mean E(N) = μ + μ = 2μ (substituted mean of X1 & X2 by Normal Distribution Properties)
      Variance V(N) = $\sigma^2 + \sigma^2 = 2\sigma^2$ (substituted mean of X1 & X2 by Normal Distribution Properties)
      Standard Deviation SD(N) = $\sqrt{2} * \sigma = 1.414\sigma$

      The distributions are similarly distributed. Both the distributions have same mean but are varied slightly with 1st distribution greater variance than the 2nd distribution.

Question 8

8. Given Information:

   F is a RV that denotes operating temperature in Fahrenheit.

   F $\sim N(90{,}25)$ i.e., it is normally distributed at mean = 90 and variance = 25

   C is a RV that denotes operating temperature in Celsius.

   As C is represented C= $\frac{5}{9}\,(F-32)$ and as F is normally distributed, so C is also normally distributed.

   To Find:

   a. Probability that one randomly selected instance of the process will have operating temperature greater than 93.8 Fahrenheit P$(93.8 < F)$

      F $\sim N(90{,}25)$

      μ = 90 and $\sigma = \sqrt{25} = 5$

      Using Normal Transformation:

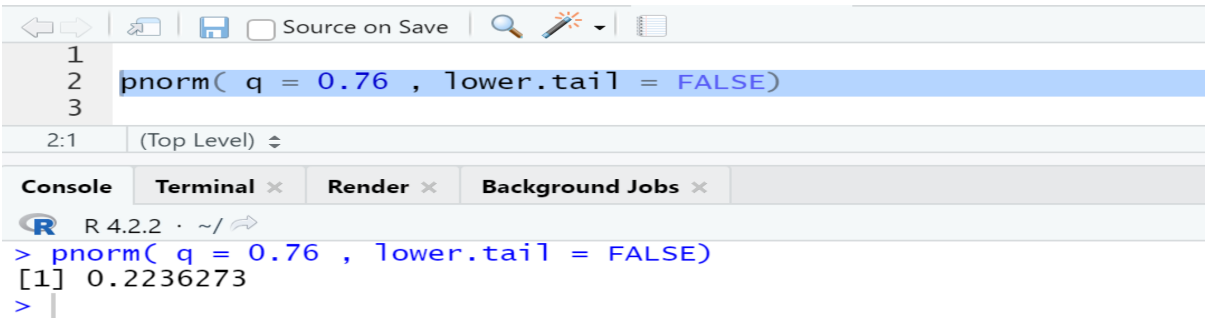
      P$(93.8 < F)$ = $P(\frac{93.8-\mu}{\sigma} < \frac{F-\mu}{\sigma})$

      = $P(\frac{93.8-90}{5} < Z)$

      = $P(0.76 < Z)$

      Finding Probabilities of Standard Normal Distribution using R Code:

      $pnorm(q = 0.76, lower.tail = FALSE)$



      P$(93.8 < F) = 0.2236273$

      Probability that one randomly selected instance of the process will have operating temperature greater than 93.8 Fahrenheit P$(\mathbf{93.8} < \boldsymbol{F})$ is $\mathbf{0.2236273}$

b. As C is normally distributed. Find its mean and variance:

C= $\frac{5}{9}\,(\,F - 32) = \frac{5}{9}\,F - \frac{160}{9}$

We know that F $\sim N(\,90{,}25)$

Mean of C = E(C) = $\frac{5}{9} * 90 - \frac{160}{9}$ (substituted mean of F by Normal Distribution Properties)

= 32.222

E (C) =32.222

Variance of C = Var(C) = $\left(\frac{5}{9}\right)^2 * 25 = 69.444$

Var (C) = 7.71

So, C$\sim \boldsymbol{N(32.22, 7.71)}$

c. Probability that one randomly selected instance of the process will have temperature below P (C < 29):

As we deduced that C$\sim N(32.22, \mathbf{7.71})$

μ = 32.22 and $\sigma = \sqrt{7.71} = 2.776$

Using Normal Transformation:

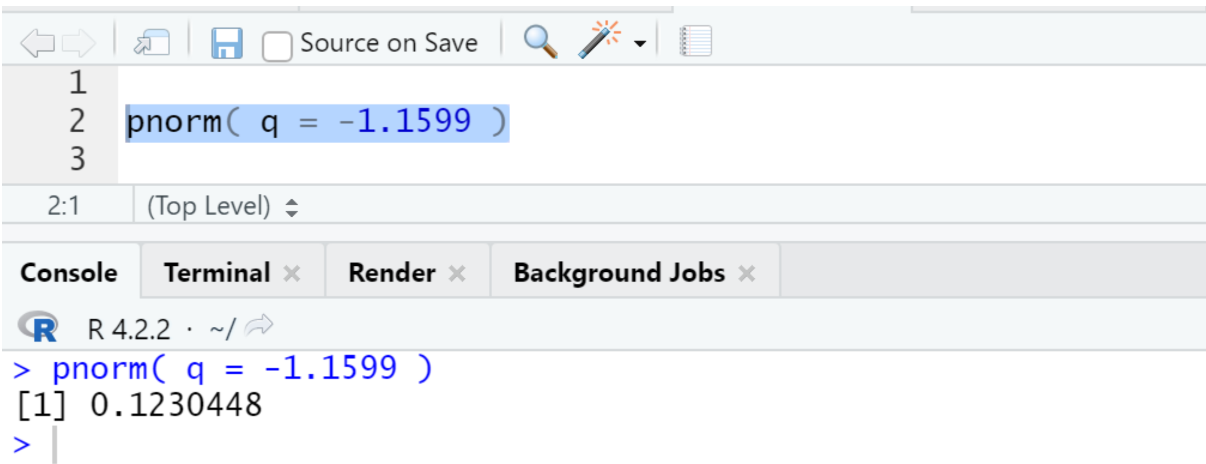
P($C < 29$) = $P(\frac{C-\mu}{\sigma} < \frac{29-\mu}{\sigma})$

= $P(Z < \frac{29-32.22}{2.776})$

= $P(\,Z < -1.1599\,)$

Finding Probabilities of Standard Normal Distribution using R Code:

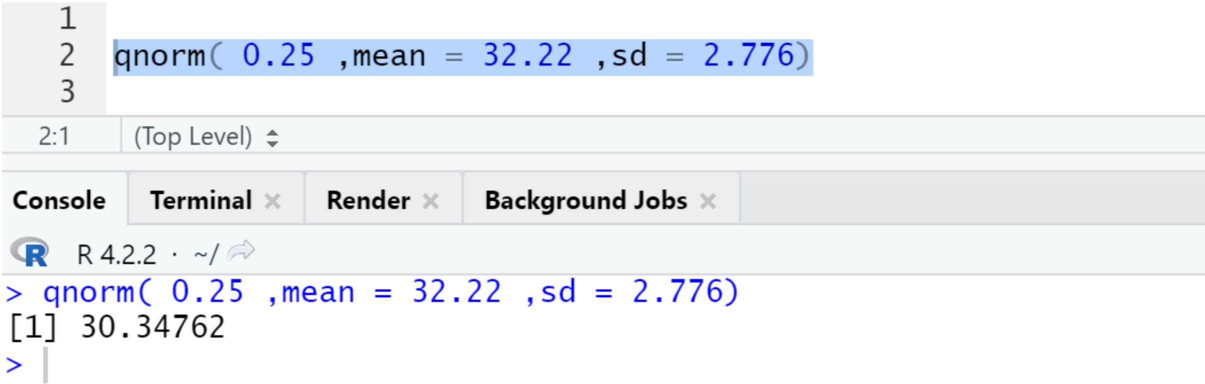
$pnorm(\,q = -1.1599\,)$



P (C < 29) = 0.1230448
Probability that one randomly selected instance of the process will have temperature below P (C < 29) is 0.1230448.

d. Probability that the operating temperature in Celsius of one instance is less than x is 0.25
P (C < x) = 0.25
We know that,
C~ $N(32.22, , \mathbf{7.71})$
μ = 32.22 and $\sigma = \sqrt{7.71} = 2.776$

To find the value of instance x for which the probability is 0.25 we will be using R Code:
$qnorm(\ 0.25\ , mean\ =\ 32.22\ , sd\ =\ 2.776)$



The value of instance x for which the probability is 0.25 = 30.34762

Question 9

9. Let F be the Random Variable RV that represents daily fluctuations of French CAC-40 Stock Index from March to June 1997.
F follows a Normal Distribution: F$\sim N\,(2600\,,2500\,)$
μ = 2600, $\sigma^2 = 2500$ and $\sigma = 50$

To Find: $P(2520 \; < F < \; 2670\,)$

To transform any Normal RV to a Standard RV we need to use standardization.
Z= $\frac{F-\mu}{\sigma} \sim N(0,1)$

Using Normal Transformation,

$$P\,(2520 \;\leq F \leq \; 2670\,) = P(\frac{2520-2600}{50} \leq \frac{F-\mu}{\sigma} \leq \frac{2670-2600}{50})$$

$$= P(\frac{2520-2600}{50} \leq Z \leq \frac{2670-2600}{50})$$

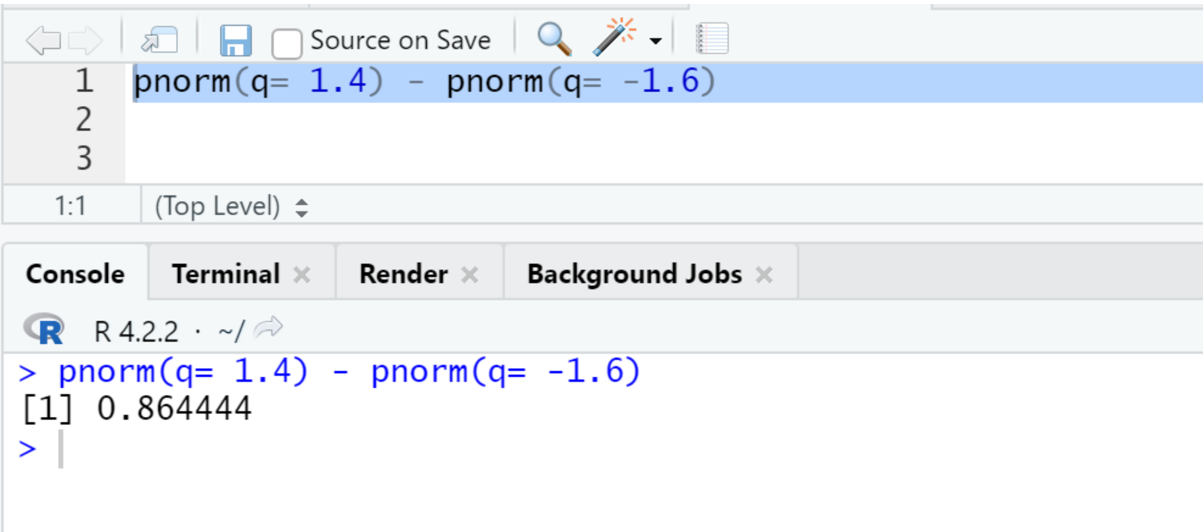
$$= P(-1.6 \;\leq\; Z \;\leq\; 1.4\,)$$
$$= P(Z \;\leq 1.4) - P(\,Z \;\leq\; -1.6)$$

Using Z tables,
= 0.9192 - 0.05480
= 0.8644

$P\,(2520 \;\leq F \leq \; 2670\,) \; = 0.8644$

Using R Code,

$P(-1.6 \;\leq\; Z \;\leq\; 1.4\,) = \; pnorm(q = \; 1.4) \; - \; pnorm(q = \; -1.6)$



OR

$pnorm(q = \; 2670\,, mean \; = \; 2600, sd \; = \; 50\,) \; - \; pnorm(q = \; 2520\,, mean \; = \; 2600, sd \; = \; 50\,)$

```
pnorm(q= 2670 , mean = 2600, sd = 50 ) - pnorm(q= 2520 , mean = 2600, sd = 50 )
```

```
> pnorm(q= 2670 , mean = 2600, sd = 50 ) - pnorm(q= 2520 , mean = 2600, sd = 50 )
[1] 0.864444
>
```

$$\boldsymbol{P}\,(2520\ \leq \boldsymbol{F} \leq\ 2670\,)\ =\ \boldsymbol{P}(-1.6\ \leq\ \boldsymbol{Z}\ \leq\ 1.4\,) =\ 0.864444$$

Probability that the CAC-40 stock index will be between 2520 and 2670 is 0.8644.

Question 10

10.

Using Reference from “Introductory Statistics” By Prem S Mann.

a. In this example, the researcher of the study had not imposed a treatment (A condition or a set of conditions that is imposed by an experimenter is called a treatment) on the elements of the study. Researcher just collected sample of data and observed the results of the study. The conclusion may be invalid as there may be different factors or variables confounding the study. **So, this is an observational study.**

b. In this example, an instructor observed the previous data and deduced a conclusion. This means that instructor had no control over the factors contributing to the previous data. So, the conclusion may be invalid as the correlation between the assignment completion and performance in the final may be confounded with many other variables. **Hence this is an observational study.**

c. The instructor selects a group of students and randomly assigns them to two study group one with assignments and other with readings. By using the randomization to assign students to one of the study groups, instructor has controlled the factors contributing the study i.e., a condition or set of conditions were imposed on a group of students. When the experimenter controls the assignment of elements to different treatment group, the study is experimental. **So, this is an experimental study**.