5. LESSON: SOLVED EXERCISES

1. The 8% of the cholesterol control analyses have erroneous results and it is necessary to repeat the analysis.

   a) If an analysis is performed, what is the probability that the result is incorrect?
   b) If 15 random analyses are observed, what is the probability of having to repeat at least 2 analyses?
   c) What number of analyses are expected to be repeated in 200 analyses?
   d) If in one day the laboratory has performed 50 analyses, what is the probability that up to 3 analyses are repeated?

a) First, the random variable and the distribution that follows must be defined.

X : 'Erroneous cholesterol analysis'
$X \sim \text{Binary}(p = 0.08)$

$P(X = 1)$
$> \text{dbinom}(1,1,0.08)$
$[1]\ 0.08$

b) In this case, the binary experiment is repeated several times so a new random variable must be defined as well as the distribution that follows.

X : 'Number of erroneous cholesterol analysis'
$X \sim B(n = 15, p = 0.08)$

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

$> 1 - \text{pbinom}(1,15,0.08)$
$[1]\ 0.3402712$

c) To calculate the number of analysis expected to repeat, the mean is calculated.

$n = 200$
$E(X) = n \cdot p$

```
> n = 200
> p = 0.08
> n*p
[1] 16
```

d) Being in this case $n = 50 > 30$ and $p = 0,08 < 0,1$ the binomial distribution can be approached by Poisson distribution.

X : 'Number of erroneous cholesterol analysis'
$X \sim B(n = 50, p = 0.08)$

So,

$P(X \leq 3)$

```
> pbinom(3,50,0.08)
[1] 0.4252957
```

2. The probability that the tests performed with an ultrasound equipment are effective is 80%. Assuming that the tests carried out are independent, calculate:

   a) Probability that the first effective trial occurs in the fifth trial.
   b) Probability that it is necessary to perform at least four trials to obtain the first effective trial.
   c) Probability that 12 trials are necessary for having 5 effective trials.
   d) Probability of performing maximum 10 and minimum 7 trials for having 3 effective trials.

a) First, the random variable and the distribution that follows must be defined.

X : 'Number of trials performed until the first effective trial'

$X \sim G(p = 0.8)$

For the fifth trial to be effective, the four previous trials must be ineffective, so:

$P(X = 4)$

```
> dgeom(4, 0.8)
[1] 0.00128
```

b) For the minimum number of trials performed to be four, at least three trials must be ineffective.

$P(X \geq 3) = 1 - P(X \leq 2)$

```
> 1-pgeom(2, 0.8)
[1] 0.008
```

c) In this case, the random variable should be modified and the distribution that follows is different, from being a geometric to being a negative binomial. For 5 trials to be effective, 7 must not be effective.

X : 'Number of trials performed until 5 effective trials'

$X \sim BN(n = 5, p = 0.8)$

$P(X=7)$

> dnbinom(7,5,0.8)
[1] 0.00138412

d) In this case, for 3 effective trials, ineffective trials must be between 4 and 7.

X : 'Number of trials performed until 3 effective trials'
$X \sim BN(n=3, p=0.8)$

$P(4 \leq X \leq 7) = P(X \leq 7) - P(X \leq 3)$

> pnbinom(7,3,0.8) – pnbinom(3,3,0.8)
[1] 0.01688207

3. In an electrical company that manufactures fuses, the probability that the fuses are defective is 0.2. A customer buys 15 fuses, but only has to use 5 of them.

   a) What will be the probability that at most 2 of these 5 fuses are defective?
   b) What will be the expected number of defective fuses in these 5 fuses?
   c) Another customer has acquired a box of 200 fuses to use 10 of them. What will be the probability that at most 2 of the 10 fuses are defective in this case?

a) First, the random variable and the distribution that follows must be defined.

X : 'Number of defective fuses'

$X \sim H(N = 15, n = 5, p = 0.2)$

$P(X \leq 2)$

```
> phyper(2,3,12,5)
[1] 0.978022
```

b) In order to calculate the number of fuses expected to be defective, the mean should be calculated.

$n = 5;\ p = 0.2$

$E(X) = n \cdot p$

```
> n = 5
> p = 0.2
> n*p
[1] 1
```

c) In this case $N = 200$ and $n = 10$, so $N > 10 \cdot n$; therefore the hypergeometric distribution can be approached by binomial distribution but with calculating with R Studio this approach should not be made.

X : 'Number of defective fuses'

$X \sim H(N = 200, n = 10, p = 0.2)$

$P(X \leq 2)$

```
> 200*0.2
[1] 40
```

```
> 200*0.8
[1] 160
> phyper(2,40,160,10)
[1] 0.6794051
```

4. In a wool factory of Edinburgh, for every 5 meters of fabric produced a defect appears. Knowing that the number of defects appeared in the fabric follows a Poisson distribution, calculate:

   a) If five meters of wool fabric are bought, the probability that there are more than two defects.
   b) If 50 meters of fabric are bought to make 15 kilts (a typical Scottish skirt), the probability to find seven defects.

a) First, the random variable and the distribution that follows must be defined.

X : 'Number of defects in five meters of wool fabric'

$X \sim P(\lambda = 1)$

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

```
> 1-ppois(2,1)
[1] 0.0803014
```

b) In this case, a new λ parameter must be calculated, since the random variable has changed. Instead of having 5 meters of fabric there are 50 meters so the λ parameter is modified linearly.

X : 'Number of defects in fifty meters of wool fabric'

$X \sim P(\lambda = 1 \cdot 10)$

$P(X = 7)$

```
> dpois(7,10)
[1] 0.09007923
```

5. The time takes a student to get from home to college varies uniformly between 35 and 45 minutes. What time should he leave home with a minimum probability of 0.8 to get to class on time if classes start at 8 am?

First, the random variable and the distribution that follows must be defined.

$X =$ "Minutes from home to college"

The variable follows an uniform distribution. $X \sim U[35,45]$

$$P(X \leq x) \geq 0.8 \Rightarrow \frac{x-35}{45-35} \geq 0.8$$

```
> qunif(0.8,35,45)
[1] 43
```

So that, if the student leave home at 7:17a.m. or earlier, will get to class on time with a probability of 0.8 or higher.

6. In a store, the time we wait from the entry of a customer to the entry of the next customer is distributed exponentially with a mean of 5 minutes. Calculate the probability that we will have to wait between 8 and 10 minutes until the entry of the next customer.

First, the random variable and the distribution that follows must be defined.

X : 'Minutes we have to wait until the entry of the next customer'. $X \sim \varepsilon(1/\lambda)$

The λ parameter should be calculated.

$$E(X) = 5 = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{5}$$

The probability asked is:

$$P(8 \leq X \leq 10) = P(X \leq 10) - P(X < 8)$$

```
> pexp(10,1/5) - pexp(8,1/5)
[1] 0.06656123
```

7. A pot of jam is classified as 'syrup' if the amount of sugar is between 420 and 520 g. The manufacturer when analyzing pots observes that the average weight is 465 g, with a standard deviation of 30 g. Knowing that the weight of the sugar follows a normal distribution,
   a) What percentage of the manufacturer's production cannot be labeled as 'syrup'?
   b) What are the two central values that we have to set so that among them there are 50% of the pots?

First, the random variable and the distribution that follows must be defined.

X : 'Sugar amount in g inside the pot'.

$420 \le X \le 520 \rightarrow$ Pot that could be consider as 'Syrup'

$X < 420$ and $X > 520 \rightarrow$ Pot that could not be consider as 'Syrup'

$E(X) = 465$ and $\sigma = 30$ X ~ N(465;30)

a) This percentage can be calculated by two ways.

1)

$P(420 \le X \le 520) = P(X \le 520) - P(X \le 420)$

```
> pnorm(520,465,30) - pnorm(420,465,30)
[1] 0.8998163
```

$1 - P(420 \le X \le 520) = 0.1004$

So that, the 10.02% could not be labelled as 'syrup'

2)

$P(X \ge 520) + P(X \le 420)$

```
> pnorm(520,465,30,lower.tail = F) + pnorm(420,465,30)
[1] 0.1001837
```

So that 10.02% could not labbeled as 'syrup'.

We will use standardized normal distribution:

$$P(b \le X \le a) = P(X \le a) - P(X \le b)$$

$$Z = \frac{x-465}{30}$$

For solving this section we will apply symmetry many times.

$$P(b \le X \le a) = P(-t \le Z \le t) = P(Z \le \frac{a-465}{30}) - P(Z \le \frac{b-465}{30}) = P(Z \le t) - P(Z \le -t)$$

$$P(b \le X \le a) = P(Z \le t) - P(Z \le -t) = P(Z \le t) - P(Z \ge t) =$$

$$P(Z \le t) - \left(1 - P(Z \le t)\right) = -1 + 2P(Z \le t) = 0.5 \Rightarrow P(Z \le t) = 0.75$$

```
> qnorm(0.75,0,1)
[1] 0.6744898
```

$$P(Z \le t) = 0.75 \Rightarrow t = 0.6744898$$

$$t = \frac{a-465}{30} \Rightarrow a = 485.234694 \quad \text{and}$$

$$-t = \frac{b-465}{30} \Rightarrow b = 444.765306$$

So that a and b central values are 485.234694 and 444.765306.

8. In a factory, two types of pieces are produced, one of them made with renewable materials and another with raw materials derived from oil.
   a) If 2% is defective, what is the probability that when selecting 100 pieces there will be no more than 3 defective?
   b) 40% are pieces based on renewable materials. What is the probability that when selecting 600 pieces, 352 or more pieces are made of raw materials derived from oil?

a) First, the random variable and the distribution that follows must be defined.

X : 'Amount of defective pieces'

n=100; p=0.02 $\quad X \sim B(100;0.02)$

$P(X \leq 3)$

$> \text{pbinom}(3,100,0.02)$

[1] 0.8589616

b) In this case, a new random variable must be defined and consequently the distribution it follows.

X : 'Amount of defective pieces made with raw materals derived from oil'

n=600; p=0.6 $\quad X \sim B(600;0.6)$

$P(X \geq 352) = 1 - P(X \leq 351)$

$> \text{pbinom}(351, 600, 0.6, \text{lower.tail} = \text{F})$

[1] 0.7610461

In this case, the binomial distribution should not be approached by a normal distribution since R Studio can make the calculus very easily.