Statistics. ESIGELEC

Lab Session 2: Probability distributions

Names and surnames	
Date	

TASK 1. In order to arrive at ESIGELEC, a student can choose between two different paths. The time spent to arrive at ESIGELEC using path A follows a normal distribution with mean μ=25 and standard deviation σ=7, whereas the time using path B is normal with μ=30 and σ=3 (units measured in minutes). In your answers, specify the R functions you use.

a) What is the population in this exercise? [2 points]

b) What is/are the random variable of interest in this exercise? [2 points]

c) If classes begin at 8:30, and the student leaves home at 7:45, which path should he/she choose in order to maximise the probability of arriving before class starts? What if he/she leaves home at 8:10? [3 points]

d) Which path should be chosen if he/she wants to leave home as late as possible, but keeping that the probability of arriving before the class starts is at least 0.95? [3 points]

TASK 2. An inspection procedure must be designed and implemented at the end of a production process, consisting on batches of 30,000 units.

a) The customer considers unsuitable any batch with 5% or more defective units. A simple inspection plan is designed, so that n units will be picked at random from a given batch; and the batch will be rejected if more than $c = 4$ defective units are observed. Calculate the minimum value for n for which the probability of *accepting* an unsuitable batch is 0.04 at most (this is usually called the *consumer's risk*). [8 points]

b) The manufacturer considers suitable any batch with at most 1% of defective units. If the inspection plan designed in section (a) was implemented, what would be the probability of *rejecting* a suitable batch? (this is usually called the *producer's risk*) [2 points]

SOLUTION:

TASK 1

a)

POPULATION: All the days in which this student goes to ESIGELEC.

Why? Because we are studying a random variable (see section (b)) that is different each day this student goes to ESIGELEC.

The population is *not* the students of ESIGELEC because we are talking of one particular student, and we are analysing how long it takes him/her to arrive at ESIGELEC each day. We are *not* studying the time to get to ESIGELEC for each student (in one concrete day).

b)

RANDOM VARIABLE(S): Two random variables:

- T_{A}: The time it takes him/her to get to ESIGELEC using path A.
- T_{B}: The time it takes him/her to get to ESIGELEC using path B.

More precisely, $T_{\mathrm{A}} \sim N(\mu_{\mathrm{A}} = 25\,, \sigma_{\mathrm{A}} = 7)$ and $T_{\mathrm{B}} \sim N(\mu_{\mathrm{B}} = 30\,, \sigma_{\mathrm{B}} = 3)$.

c)

			pnorm	pnorm	qnorm
Path	Mean	Std. Dev	45	20	0,95
A	25	7	0.9978626	0.2375253	36.5139754
B	30	3	0.9999997	0.0004290603	34.9345609

The probability of arriving on time if you have 45 minutes is:

```
> pnorm(45,mean=25,sd=7)
[1] 0.9978626
> pnorm(45,mean=30,sd=3)
[1] 0.9999997
```

Choose path B.

The probability of arriving on time if you have 20 minutes is:

```
> pnorm(20,mean=25,sd=7)
[1] 0.2375253
> pnorm(20,mean=30,sd=3)
[1] 0.0004290603
```

Choose path A.

d)

```
> qnorm(0.95,mean=25,sd=7)
[1] 36.51398
> qnorm(0.95,mean=30,sd=3)
[1] 34.93456
```

In path A, if you leave `qnorm(0.95,mean=25,sd=7)=` 36.51398 minutes before class, you have probability 0.95 of arriving on time to class. For path B, you need to leave `qnorm(0.95,mean=30,sd=3) =` 34.93456 minutes before class to keep this probability. Choose path B.

TASK 2

a)

Let $X \sim \mathrm{Bi}(n, p)$ be the number of defective units contained in an n-sized random sample from a batch containing a proportion p of defective units.

A batch is considered unsuitable by the customer if $p = 0.05$ or more. The inspection plan should meet a *consumer's risk* of at most 0.04, which is to say:

$$\Pr(\text{accepting} \mid p = 0.05) \leq 0.04 \;\; \Leftrightarrow \;\; \Pr(X \leq 4 \mid p = 0.05) \leq 0.04.$$

Recall that $\Pr(X \leq c)$ can be calculated with R using **`pbinom(c, size = n, prob = p)`**, which is equivalent to **`pbinom(c, size = n, prob = p , lower.tail = TRUE)`**.

Then, using brute force, we can find out the smallest value of n for which $\Pr(X \leq 4 \mid p = 0.05) \leq 0.04$ holds:

```
> pbinom(4, size = 10, prob = 0.05)

[1] 0.9999363

> pbinom(4, size = 100, prob = 0.05)

[1] 0.4359813

> pbinom(4, size = 186, prob = 0.05)

[1] 0.04188601

> pbinom(4, size = 187, prob = 0.05)

[1] 0.04055467

> pbinom(4, size = 188, prob = 0.05)

[1] 0.03926226
```

Therefore, $n = 188$ is the smallest sample size that would meet the *consumer's risk*.

b)

If a simple inspection plan with $n = 188$ and $c = 4$ is implemented, what would be the *producer's risk* (i.e., the probability of rejecting a suitable batch)?

According to the text, a batch is considered suitable by the manufacturer if the (real) proportion of defective units in the batch is $p \leq 0.01$. Therefore, the probability of the inspection plan rejecting a suitable batch can be expressed as:

$$\Pr(X > 4 \mid p = 0.01) = 1 - \Pr(X \leq 4 \mid p = 0.01).$$

We can calculate this with R:

```
> 1 - pbinom(4, size = 188, prob = 0.01)
```

```
[1] 0.04162965
```

or, equivalently:

```
> pbinom(4, size = 188, prob = 0.01, lower.tail = FALSE)
```

```
[1] 0.04162965
```

Therefore, if we implemented the inspection plan designed in section (a), the probability of rejecting a 'correct' batch would be $0.0416 \equiv 4.16\%$.