Practical Session 10

Random variables

In this session, we shall see how to illustrate how to use Excel to calculate probabilities for some of the random variables we have seen in class.

1. The geometric distribution.

   Exercise 1a) from the Exercise sheet this week is an example of a geometric probability.

   1. On average, one in every twenty students in Political Science and Journalism says that they are a member of a political party.
   a) Talking to students in these degree programs, what is the chance that the first member of a political party is the tenth student we talk to?

   There is no specific function in Excel for calculating probabilities of this type, but we can simply use the formula = **(1-0.1)^9 * 0.1** to calculate the probability directly.

2. The binomial distribution.

   In contrast to the geometric distribution, Excel has an inbuilt function **BINOM.DIST** (**DISTR.BINOM**) for calculating binomial probabilities. The syntax for the formula is **BINOM.DIST(number_s; trials; probability_s; cumulative)** where **number_s** is the number of successes, **trials** is the number of attempts, **probability_s** is the probability of success and **cumulative** is **TRUE** if we want the probability of number_s or fewer successes or **FALSE** if we want the probability of exactly number_s successes.

   As an example, Exercise 1b) says

   b) One class contains 30 students. What is the chance that exactly one of them is a party member?

   We could solve this directly as 30 * 0.1 * (1-0.1)^29 but we can also use the binomial formula =**BINOM.DIST(1; 30; 0.1; FALSE)**

3. The normal distribution.

   Various functions are available for calculating normal probabilities:

   - **NORM.DIST** (**DISTR.NORM.N**) can be used to calculate P(X ≤ x) for a normal distribution. The syntax is **NORM.DIST(x,mean,standard_dev,cumulative)** where **x** is the value of interest, **mean** and **standard_dev** are the mean and s.d. of the normal distribution and **cumulative** is set to **TRUE** to calculate P(X ≤ x). (In our examples, we should never set this to **FALSE** as this gives the height of the density function which is not relevant to us.

Exercise 3 a) is as follows:

3. The height of women in a certain country follows a normal distribution with mean 162.5 cm and standard deviation 2.5 cm and, independently, the height of men is normal with mean 170 cm and standard deviation 5 cm.

a) What is the proportion of women under 168 cm tall?

We can calculate this using = **NORM.DIST(168;162.5;2.5;TRUE)**

- In older versions of Excel, you can use the functions NORMDIST (DISTR.NORM) with the same syntax to do the same calculations.

- To calculate probabilities for a standard normal distribution (with mean 0 and standard deviation 1), you could use the NORM.DIST function as above, but you could also use **NORM.S.DIST** (**DISTR.NORM.ESTAND.N**) which does not require you to specify the mean

  If we want to calculate the probability P(Z < 2) for a standard normal variable, then we could write = **NORM.DIST(2;0;1;TRUE)** or = **NORM.S.DIST(2;TRUE)** which does not require us to specify the mean or standard deviation.

- In older versions of Excel, you can use the functions **NORMSDIST** (**DISTR.NORM.ESTAND**) to calculate probabilities for a standard normal distribution.

- In order to calculate percentiles of a normal distribution, you can use the function **NORM.INV** (**DISTR.NORM.INV**). To calculate the point x so that P(X ≤ x) = p for a normal distribution with mean m and standard deviation s, we simply write =**NORM.INV(p;m;x)**. In the case of a standard normal distribution, you can use **NORM.S.INV** (**DISTR.NORM.ESTAND.INV**).

  Exercise: For a standard normal distribution, calculate the points z so that P(Z ≤ z) = 0.95, 0.975, 0.99 and 0.995 respectively.

  = **NORM.S.INV(0.95)** or =**NORM.INV(0.95;0;1)**

Exercise 3 c) from this weeks Exercises is as follows:

c) Calculate the third quartile of the female height distribution and the first quartile for the males.

We can calculate the probability for the case of the females via

=NORM.INV(0.75; 162.5; 2.5)