***Exercises using the normal distribution and the chi-square distribution***

A) NORMAL DISTRIBUTION

For simplicity’s sake, let’s assume that we have an IQ scale (mean = 100; standard deviation = 15) that is distributed normally.

a) What is the proportion of individuals with an IQ less than 90?

That is, P(X < 90) or F(90). The best option is to use Excel. Just click on the fx (function) option and select “DISTR.NORM”







So P(X < 90) = F(90) = .25

As for “acum”… if you indicate “1”, you get F(90), which is what we want (i.e., the function of distribution, which is a probability).

But…if you indicate “0”, you get f(90) (i.e., the probability density function; the height of the curve for X=90; this is NOT a probability…)

b) What is the proportion of individuals with an IQ between 80 and 90?

P(80 < X < 90) = P(X<90) – P(X<80). You just have to compute these two terms, and that’s very easy to do.

c) Let’s complicate things a bit. We know that Peter claims that his IQ leaves below 80% of individuals…. Can we find out Peter’s IQ? Yes, but now we have a probability and we want to find the IQ (i.e., the inverse procedure as we just did). Then you go to functions, but not to normal distribution, but to inverse normal distribution…

P(IQ < a) = .80 Which value is “a”?



Peter’s IQ is around 113.

B) CHI-SQUARE DISTRIBUTION (and glance of inferential stats)

Let’s assume that were throw a dice 60 times, and the results are:

1: 10

2: 20

3: 5

4: 5

5: 10

6: 10

Do you think the dice is well balanced (i.e., all 6 options are equally likely)?

To do this type of exercise, we need to check the goodness of fit of the empirical frequencies (i.e., the ones above), with the theoretical/expected frequencies (deduced from one model: all options are equally likely; in red; 1/6 multiplied by 60 is 10 in each case:

1: 10 10

2: 20 10

3: 5 10

4: 5 10

5: 10 10

6: 10 10

and now we computer the following formula of goodness of fit



If we do that, we have the following value of chi-square = 0+10+2.5+2.5+0+0=15

Is that a good fit? Clearly, it is not perfect…

How can we decide that the fit is good or bad?

This is what we do: If our model were TRUE (i.e., the dice is well balanced), then the value of chi-square that we got followed a CHI-SQUARE DISTRIBITUTION with #CELLS-1 degrees of freedom (5 in the example: 6-1).

(Note: once we know the data from 1,2,3,4,5, we can deduce the number of “6s”, that’s why we only have 5 cells that vary freely; in an experiment with coins, the degrees of freedom would be 1, as we only have two cells: head vs. tails, and if you toss a coin 40 times, and we say that the number of heads is 25, the number of tails will necessarily be 15.)

Indeed, if we have a well-balanced dice and we repeat the experiment 4 trillion times, and then we get the chi-square each time, and then we plot a smoothed histogram, we will get something very similar to the CHI-SQUARE distribution with FIVE degrees of freedom.



Obviously, if the model is true, the values of the chi-square statistic will typically be low (i.e., there would be little discrepancy between the observed and the expected values). Conversely, if the model is not true, the fits will be bad and the chi-square statistic will be high.

So the (conventional) criterion is the following:

--If the empirical chi-square test is larger than Percentile 95, we conclude that our model was wrong (i.e., the dice is not balanced)

--If the empirical chi-square test is smaller than Percentile 95, we conclude that our model is correct (i.e., the dice is balanced)

Here is the chi-square distribution:



Percentile 95 when we have 5 degrees of freedom is 11.07

As the empirical chi-square test (15) is larger than Percentile 95 (11.07), we conclude that the dice is not balanced.