RANDOM NUMBER GENERATOR, CHI-SQUARED TEST, CENTRAL LIMIT THEOREM

country roads Driving home via the highway takes $t = 45$ min if there is no traffic jam and 90 min if there is one. The probability of a jam is $p = 0.2$. An alternative drive – via a country road – takes (always) 60 minutes (never any jams). Which route should we prefer?

the expected value of the drive-time via via the highway is
$E(T) = 45 \cdot 0.8 + 90 \cdot 0.2 = 54$ minutes; so the highway saves us 6 minutes.

Illustrate this with the help of the random number generator: create, say, 100 random numbers in the column A, from the $[0, 1]$ interval. Then using the IF function fill in the adjacent cells of the B column – if A(cell#) < 0.8 put B(cell#)= 45; otherwise put 90. In the B(101) compute the mean T – the sum of B(1):B(100)/100. It should be close to 54, right?

Repeat the procedure 10 times. Calculate the average value $\bar{T}$ of the T_i's ($i = 1, 2, \ldots, 10$) and the S^{*2} estimator of the variance. Calculate the standard mean deviation. Construct a 95% confidence interval for the mean time of the trip home. Check whether this interval contains 54.

Pearson's test Ask one of your colleagues to generate for you 200-300 values following the Poisson distribution with λ from, say, 5—15 interval. Store the data in the B column.

1. Construct the frequency histogram; store the bins in the C column, and the frequencies in the D column;
2. compute λ as a regular average of your data and store it in the A1 cell. Calculate – using the POISSON excel function theoretical frequencies, using λ from the A1 cell and store these frequencies in the F column;
3. confront calculated (column E) frequencies with the 'experimental' ones (column D); calculate the χ^2 value and test the hypothesis about the validity of the assumption about the type of distribution (Poisson);
4. **an alternative, more interesting, approach**
   do not calculate the λ; put into the A1 cell an approximate value (from the inspection of your histogram); calculate the theoretical frequencies and the sum:

   $$\sum_{bins} \left(frq^{theor} - frq^{exp} \right)^2$$

   store it in a cell which will be your 'target cell' (i.e. A2).

   Call SOLVER and ask it to minimize the content of the 'target cell' by changing the parameter stored in the A1 cell. Ask your colleague what was the value of λ he/she used for generating data.

central limit theorem Using the random-number generator, generate a set of values of a random variable X_1 uniformly distributed in the [0,1] interval (200-300 values or more). Divide the interval into 15 (or more if you generated more than 300 values) bins; construct the histogram of your data.

Repeat the procedure – generate an analogous set of data for X_2 and calculate a new random variable $Y_1 = X_1 + X_2$; visualize the behaviour of Y_2 in the [0,1] interval.

Carry on – generate X_3, X_4, up to (at least) X_{30} 'new' random variables and – consequently – Y_2, Y_3, $\ldots Y_{29} = \sum_i^{30} X_i$; visualize the behaviour of each Y_i in the [0,1] interval.

an extra task for those aiming at the very good assessment: test the hypothesis that the distribution of Y_{29} is the normal one.