

Ex. 5 (practical class) – χ^2 test and linear regression

The following exercises are supposed to be treated using Microsoft EXCEL, which is well adapted to the visualization and the treatment of (potentially large) tables of data. In addition, EXCEL includes the relevant probability distributions such as (among others) the binomial, normal, and χ^2 distributions.

The functions below may be useful when using EXCEL for probability- or statistics-related calculations (left: English names, right: French names):

=BINOM.DIST (k;n;p;c) =BINOM.DIST.RANGE (n;p;a;b) =BINOM.INV (n;p; α) =NORM.DIST (a; μ ; σ ;c) =NORM.INV (p; μ ; σ) =NORM.S.DIST (a;c) =NORM.INV (p) =CHISQ.DIST (x; ν ;c) =CHISQ.INV (p; ν) =LINEST ({Y};{X};c;s)	=LOI.BINOMIALE.N (k;n;p;c) =LOI.BINOMIALE.SERIE.N (n;p;a;b) =LOI.BINOMIALE.INVERSE.N (n;p; α) =LOI.NORMALE.N (a; μ ; σ ;c) =LOI.NORMALE.INVERSE.N (p; μ ; σ) =LOI.NORMALE.STANDARD.N (a;c) =LOI.NORMALE.STANDARD.INVERSE.N (p) =LOI.KHIDEUX.N (x; ν ;c) =LOI.KHIDEUX.INVERSE.N (p; ν) =DROITEREG ({Y};{X};c;s)
---	---

More detailed information on these functions can be found in the internal help and documentation of EXCEL.

Exercise 5.1

A group of people has been set on a weight gain program. The following table summarizes the weight gain (in kg) and the corresponding number of persons.

weight gain x	$x < 0$	$0 < x < 1$	$1 < x < 2$	$2 < x < 3$	$3 < x < 4$	$4 < x < 5$	$5 < x < 6$	$6 < x < 7$	$7 < x < 8$	$x > 8$
persons	4	5	10	13	18	16	9	9	4	2

1. Visualize the data in a histogram.
Which distribution might be adequate to describe the data?
2. Give an estimate of the population's weight gain mean value and standard deviation.
3. Set up a χ^2 test to verify whether or not the data can be described by a normal distribution (at 95% confidence level).

Exercise 5.2

A material science specialist is studying the effect of adding rare earth elements into metallic alloys. He has found that an amount X (in atomic weight) of rare earth increases the mechanical resistance Y according to $Y = aX + b$.

X	1.7	1.8	1.9	2.0	2.1	2.2
Y	0.5	0.7	0.8	0.9	1.2	1.3

1. Visualize the data on a scatter plot.
2. Determine the least-square⁹ estimates of the coefficients a and b .
3. Comment on the quality of the obtained linear fit. Is the equation $Y = aX + b$ adapted to describe the data?
4. Determine the uncertainty (standard deviation) associated to the coefficients a and b of the linear equation.
5. Give 95% confidence intervals for the coefficients a and b . Comment concerning the quality of the linear fit.
6. Obtain an estimate of the expected increase of the mechanical resistance for initial amounts of $X=2.4$ and $X=3.0$.

⁹ method of least squares (*engl.*) = méthode des moindres carrés (*fr.*)