I) A radar records the speed of vehicles on a road. Determine the amount of data required to obtain an accuracy of ± 1 km / h at a confidence level of 97.5 % if the standard deviation is 17 km / h.

$$n=\frac{k^2\sigma^2}{e^2}=\frac{2.24^2 17^2}{1^2}=1450\ vehicles$$

J) The modal split of annual trips in Canada is given in the following table :

Mode of transport	**Split**
Car	81.0%
Public transport	10.4%
Bike	1.3%
Motor bike	0.1%
Taxi	0.2%
Walk	7.0%

We analyze 11 394 displacements and we obtain the following distribution :

Mode of transport	**Number of trips**
Car	9021
Public transport	1367
Bike	141
Motor bike	7
Taxi	29
Walk	829

Is the sample representative at the 95 % confidence level ? What could explain the difference, if there is any ?

The following table shows the expected displacements (E_i) and the observed displacements (O_i) according to the different modes.

Expected	Observed
9229.14	9021
1184.976	1367
148.122	141
11.394	7
22.788	29
797.58	829

We calculate the decision variable of the chi-square goodness-of-fit test :

$$X^2=\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}=37.62$$

that we compare with the value of

$$\chi^2_{6-1}(0.05)=11.07.$$

The value of the decision variable is too large for the distribution to follow the theoretical distribution at a 95 % confidence level .
The sample size may be insufficient : we could take a larger sample of trips and repeat the fit test for this 2 nd data series.

K) The number of accidents on a road is counted in the following table :

Day	Number of accidents
Day 1	1
Day 2	0
Day 3	2
Day 4	1
Day 5	0
Day 6	0
Day 7	2
Day 8	1
Day 9	0
Day 10	0
Day 11	1
Day 12	1
Day 13	0
Day 14	1
Day 15	0

a) Calculate the average number of accidents per dayy

$$\overline{X}=0.667\, accidents$$

b) Calculate the average number of accidents per day

The expression $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ *follows the reduced centered normal distribution, from which we deduce the following confidence interval*

$$\mu=0.667\pm2.576\cdot\frac{0.7}{\sqrt{15}}$$
$$\mu\in[0.201\,;1.132]$$

c) Calculate the number of days of observation necessary to obtain the mean number of accidents within ± 0.1 for a confidence level of 99%.

$$n=\frac{2.576^2 0.7^2}{0.1^2}=325.15$$

We would therefore need 326 days of observation.

L) The 2008 Origin-Destination survey revealed that the 2,213,000 AM trips depending on the pattern were distributed as follows :

Purpose	**% trips**
Work	51%
Study	29%
Other	16%
Return	4%

On a single day in 2009, the reasons for morning trips were investigated and the following data was collected :

Purpose	**Nb trips**
Work	4012
Study	2319
Other	1306
Return	376

Does this distribution agree with the results of the Origin-Destination survey ?

The goal is to determine whether the sample of travel reasons follows the distribution of the OD survey (hypothesis H_0) using the chi-square test. We compare the expected numbers (E_i) and the observed numbers (O_i) :

Motif	Oi	Ei	$\frac{(O_i - E_i)^2}{E_i}$
TRAVAIL	4012	4087	1.36
ÉTUDES	2319	2324	0.01
AUTRES	1306	1282	0.45
RETOUR	376	320	9.60
			11.42

The decision variable :

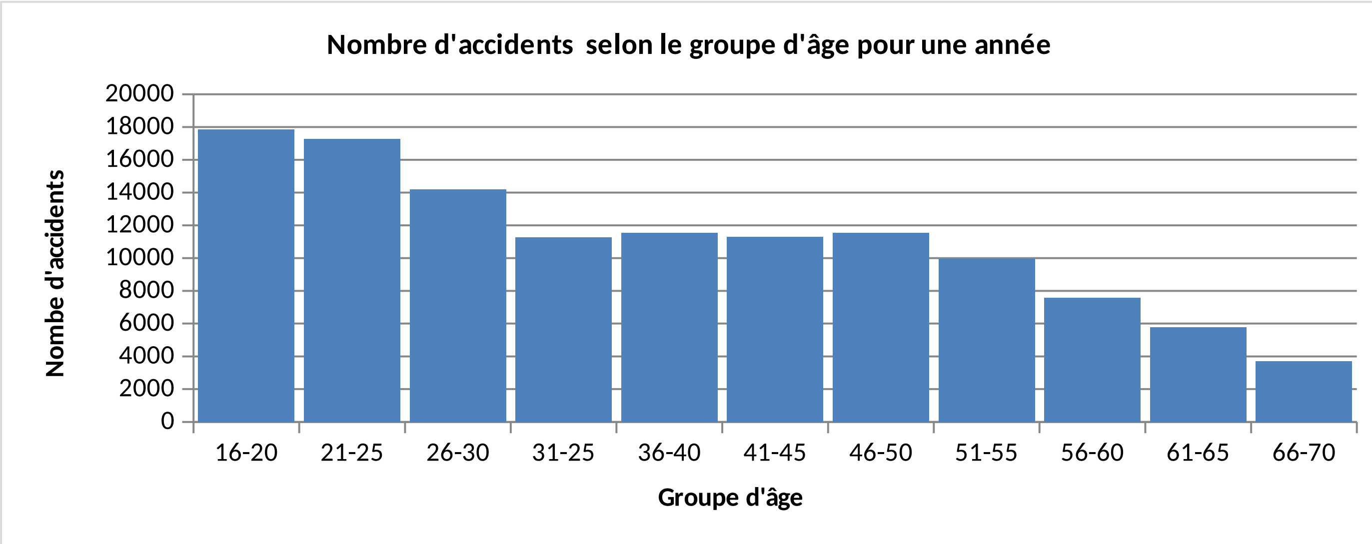
$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 11.42$$

is greater than the value of the chi-square law for a risk of the first kind of 0.05

$\chi^2_{4-1}(0.05) = 7.81$

We can therefore reject the null hypothesis that the sample observed on a single day follows the distribution of trips by reason determined during the OD survey : it is a different day.

M) The number of accidents by age group[2] is given in the following graph :



Let X be the age of a driver involved in an accident. Determine if this random variable follows an exponential distribution using a 95 % confidence level .

The average of the distribution is calculated with the central values of the intervals (18, 23, 28, 33, 38, 43, 48, 53, 58, 63 and 68) :

$$\overline{X}=\frac{18*17839+23*17277+28*14173+33*11261+\ldots}{17839+17277+14173+11261+\ldots}=37.29\,ans$$

We then calculate the theoretical frequencies of each of the intervals with the exponential law :

$$P(16\leq X\leq 20)=F(X=20)-F(X=16)$$

$$P(16\leq X\leq 20)=\left(1-e^{\frac{-20}{37.29}}\right)-\left(1-e^{\frac{-16}{37.29}}\right)=0.0662$$

We perform a chi-square goodness-of-fit test with the following assumptions :

$$H_0: X\exp\left(\lambda=\frac{1}{37.29}\right)$$

$$H_1: X\sim\exp\left(\lambda=\frac{1}{37.29}\right)$$

We then compare the observed values O_i to the expected values E_i using the decision variable of the chi-square goodness-of-fit test:

$$X^2=\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}=111732.7$$

The value of chi-2 at 11-p-1 degrees of freedom at the threshold α = 5 % is $\chi^2_{11-1-1}(0.05)=16.92$. *Since* $X^2 \geqslant \chi^2_9(0.05)$, *we can reject the hypothesis H* $_0$ *and affirm that this distribution does not follow the exponential law.*

N) Observations showed that by 10 a.m., an average of 67 bicycles were found in traffic on a certain street. If the number of bikes X passing on this street follows a Poisson distribution, calculate the probability that we count :

a) 54 bikes in 10 hours on the same street

The value of the parameter λ of a Poisson distribution is equal to the value of the mean. We calculate the probability with

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$$

$$P(X=54)=\frac{67^{54} e^{-67}}{54!}=0.014$$

b) If the average number of bicycles was 50 per 10 hours, what would become of the probability calculated in a) ?

$$P(X=54)=\frac{50^{54} e^{-50}}{54!}=0.0464$$

O) A *Statistics CANADA* survey reveals the use of public transport for work in 1996 and 2001

Age group	1996%	2001 %
15 to 19	12,7	14,5
20 to 24	14,9	16,6
25 to 34	11,0	11,9
35 to 44	8,8	8,8
45 to 54	8,4	8,5
55 to 64	9,1	8,3
65 and older	8,9	7,7

Was there a significant change between the two years at the 95 % confidence level ?

Let us make the following assumptions :

H_0: the distributions are similar
H_1: the distributions are not similar

To compare the two distributions, it is necessary to calculate the decision variable of the chi-square goodness-of-fit test, taking the observations of the year 1996 as a reference.

$$X^2 = \sum_{i=1}^{n} \frac{\left(O_{2001} - 0_{1996}\right)^2}{O_{1996}} = 0.756$$

and compare it with

$$\chi^2_{7-1}(0.05) = 12.59$$

The value of the decision variable is less than the value of $\chi^2_6(0.05)$. *We cannot reject the hypothesis H* $_0$*: the two samples follow the same law at the 95 % confidence level .*

P) Daily flows between 6 a.m. and 6 p.m. in one direction of a motorway are classified in the following table :

Hour	Flow (veh)
6h	1519
7h	2267
8h	2036
9h	1902
10h	2025
11h	2197
12h	2371
13h	2669
14h	3617
15h	5486
16h	5843
17h	6024
18h	5012

a) Calculate the average flow during this time of day.

$$\overline{X} = 3305.23\, vehicles$$

b) Calculate a 95 % confidence interval on the mean if the standard deviation is 1673 vehicles.

$$\mu = 3305.23 \pm 1.96 \cdot \frac{1673}{\sqrt{13}}$$

$$\mu \in \left[2395.8\, ; 4214.7\right]$$

Q) Morning peak flows (in number of vehicles) on a motorway are counted every day of a week :

Hour	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
06:00	511	1659	1492	1647	1658	1689	892
07:00	817	2553	2322	2526	2483	2447	1584
08:00	1154	2316	2039	2319	2367	2291	2173
09:00	1698	1974	1903	2208	2123	2314	2705
10:00	2302	2132	2011	2189	2234	2632	3236

Are the following flow rates significantly similar? Consider a 95% confidence level..

a) Tuesday, compared to Monday
b) Thursday, compared to Wednesday

We calculate the decision variables

a)

$$X^2 = \sum_{i=1}^{n} \frac{\left(O_{tue} - 0_{mon}\right)^2}{O_{mon}} = 80.263$$

b)

$$X^2 = \sum_{i=1}^{n} \frac{\left(O_{thu} - 0_{wed}\right)^2}{O_{wed}} = 5.996$$

We compare the variables X^2*with* $\chi^2_{5-1}(0.05) = 9.49$ *and we conclude that*

a) *The flow rates are not the same, because* $X^2 > \chi^2_4(0.05)$
b) *The flow rates are similar because* $X^2 < \chi^2_4(0.05)$

R) A bus line has 25 stops. The average travel times between each stop are given in the table below.

# Depart	# Destination	avg length. (min)	Std Dev (min)
1	2	4	0.8
2	3	3	0.1
3	4	2	0.7
4	5	7	0.7
5	6	6	0.5
6	7	5	0.8
7	8	6	0.2

8	9	5	0.8
9	10	2	0.7
10	11	4	0.1
11	12	2	0.5
12	13	7	0.7
13	14	3	0.3
14	15	4	0.1
15	16	5	0.6
16	17	5	0.5
17	18	4	0.1
18	19	6	0.1
19	20	7	0.0
20	21	3	0.4
21	22	7	0.5
22	23	6	0.2
23	24	4	0.2
24	25	5	0.7

a) Assuming that the travel time between each stop follows a Normal law, calculate the probability that a user traveling from stop 3 to stop 17 will complete their journey in less than an hour.

The average time between stops # 3 and # 17 is $\overline{X} = 63\,minutes$*. The variance is* $\sigma^2 = 4.5\,minutes^2$*. Standard deviation is thus* $\sigma = 2.12\,minutes$*. We calculate the probability as :*

$$P(X \le 60) = P\left(Z \le \frac{60-63}{2.12}\right) = \Phi(-1{,}42) = 1 - 0.92219 = 0{,}07781$$

b) Calculate a 95% confidence interval on the average time the line has traveled.

The average time between stops # 1 and # 25 is $\overline{X} = 112\,minutes$*. The variance is* $\sigma^2 = 6.21\,minutes^2$*; the standard deviation is therefore* $\sigma = 2.49\,minutes$*. The confidence interval is calculated with*

$$\mu = 112 \pm 1.96 \cdot \frac{2.49}{\sqrt{24}}$$

$$\mu \in [111.0\,;113.3]$$

c) Calculate the number of time measurements between stops 7 and 8 needed to obtain the mean time to ± 0.1 minutes if the standard deviation remains at 0.2 at a 95% confidence level.

The number of measurements to be performed is calculated using

$$n=\frac{k^2\sigma^2}{e^2}=\frac{1.96^2\,0.2^2}{0.1^2}=15.36\approx 16\,mesures$$

U) The average number of vehicles on a road in 2 hours is 5040. Assuming that the number of vehicles arriving at a point follows a Poisson distribution, calculate :

a) The probability of counting at least 6 vehicles in an interval of 20 seconds.

We have $\lambda=\overline{X}=\frac{5040\,véhicules}{7200\,secondes}=\frac{14\,véhicules}{20\,secondes}$

$$P(X\geq 6)=1-P(X=0)-P(X=1)-P(X=2)-P(X=3)-P(X=4)-P(X=5)$$

$$P(X\geq 6)=1-\frac{e^{-14}14^0}{0!}-\frac{e^{-14}14^1}{1!}-\frac{e^{-14}14^2}{2!}-\frac{e^{-14}14^3}{3!}-\frac{e^{-14}14^4}{4!}-\frac{e^{-14}14^5}{5!}=0.994$$

b) The probability of counting 10 vehicles in an interval of 30 seconds.

We have $\lambda=\overline{X}=\frac{5040\,vehicles}{7200\,secondes}=\frac{21\,vehicles}{20\,secondes}$

$$P(X=10)=\frac{e^{-21}21^{10}}{10!}=0.00349$$

c) The probability of counting 80 vehicles in a two-minute interval.

We have $\lambda=\overline{X}=\frac{5040\,vehicles}{120\,minutes}=\frac{84\,vehicles}{2\,minutes}$

$$P(X=10)=\frac{e^{-84}84^{80}}{80!}=0.0404$$

V) An attempt is made to validate a traffic simulation model by comparing the actual speeds of vehicles at a point with the speeds generated by the simulation. Determine if there is a significant difference between the two speed sets using a 95% confidence level.

Speeds measured (km/h)	Model (km/h)
51.1	50.6
51.3	50.7
51.8	50.9
52.1	51.1

	52.2	51.6
	52.2	52.0
	52.3	52.6
	52.7	52.9
	53.1	53.1
	53.2	53.2
	53.8	54.5
	55.2	54.9
	57.4	55.2
	57.6	58.4
	58.2	59.3
MEAN	53.6	53.4
STD DEV	2.4	2.7

We perform a statistical test to compare the means :

H0 : The means are equal

H1 : The means are different

We calculate the decision variable

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{53.6 - 53.4}{\sqrt{\frac{2.4^2}{15} + \frac{2.7^2}{15}}} = 0.214$$

We compare the decision variable with $z_{\alpha/2} = 1.96$

Since $|0.214| < 1.96$ *, we do not reject the hypothesis H0; there is therefore no significant difference between the two distributions.*

W) The number of vehicles passing a given point was counted for 190 5-minute intervals. The results are as follows :

Number of vehicles in 5 minutes of observation	Frequency observed
0	0
1	0
2	0
3	1
4	2
5	4
6	6
7	6

8	11
9	15
10	18
11	20
12	21
13	18
14	17
15	15
16	13
17	9
18	6
19	5
20	3

a) Calculate the average number of vehicles per 5 minute interval.

$$\overline{X} = \frac{0*0+1*0+2*1+3*2+4*2+5*4+\ldots}{0+0+1+2+2+4+\ldots} = 12.1\,\textit{vehicles per}\,5\,\textit{min}$$

b) Does this distribution follow a Poisson law ? Use a 95 % confidence level .

Number of vehicles in 5 minutes of observation	Oi	Ei	$\frac{(O_i - E_i)^2}{E_i}$
0	0	0.0	0.001
1	0	0.0	0.012
2	0	0.1	0.075
3	1	0.3	1.588
4	2	0.9	1.254
5	4	2.2	1.380
6	6	4.5	0.476
7	6	7.9	0.437
8	11	11.9	0.069
9	15	16.1	0.069
10	18	19.5	0.112
11	20	21.5	0.102
12	21	21.7	0.024
13	18	20.3	0.253
14	17	17.6	0.018
15	15	14.2	0.045
16	13	10.8	0.463

17	9	7.7	0.225
18	6	5.2	0.130
19	5	3.3	0.867
20	3	2.0	0.493
			8.093

$X^2 = 8.093 \chi^2_{21-1}(0.05) = 31.41$

$X^2 < \chi^2_{20}(0.05)$ *We do not reject the hypothesis that the distribution can follow a Poisson distribution at a confidence level of 95%.*