ISEP 2022/2023: Engineering cycle - 2nd year

STATISTICS – Exercises SESSION N° 1

About Chapter :

- Chapter 1 : DESCRIPTIVE STATISTICS

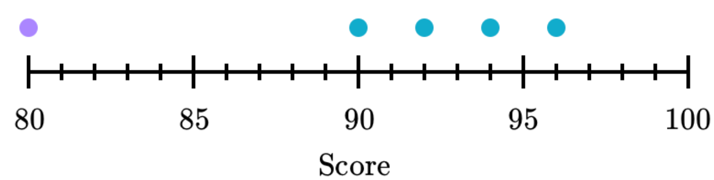
TD1 – 1 :

1. What is the median of the following statistical series? 1, 2, 4, 6, 4
2. The following table gives the distribution of the number of points scored by each player during a Quidditch season. $\frac{23\ 11+12+7+3+x}{5} = 8$

Player	Fred	George	Olivier	Angelina	Harry
Number of points	11	12	7	3	?

If the team scored on average 8 points, how many points did Harry score? 7

3. Anna played 5 rounds of golf this year. During the final round she obtained the lowest score of 80 points.



What is the impact of this last game on the mean and on the median of the scores of the previous games? Interpret the results. both decrease

TD1 – 2 :

均 中 众

Calculate the mean, median and mode for the following samples:

① $\frac{8+8+14+7+14}{10} = 5.1$

② 5 ③ 5

1) (~~3~~ ; ~~5~~ ; ~~2~~ ; 6 ; ~~5~~ ; 9 ; ~~5~~ ; ~~2~~ ; 8 ; 6)

2, 2, 3, 5, 5, 5, 5, 6, 6, 8, 9

2) (~~1.28~~ ; 2.16 ; ~~0.75~~ ; ~~1.44~~ ; 2.05 ; ~~0.65~~ ; ~~1.26~~ ; ~~1.73~~ ; ~~1.81~~ ; ~~0.92~~)

0.65 0.75 0.92 1.26 1.28 1.44 1.73 1.81 2.05 2.16

① $(1.44+2.18+2.72+3.54+4.21)\times\frac{1}{10} = 1.405$

Last update on October 27th, 2022

② $\frac{1.28+1.44}{2} = \frac{2.72}{2} = 1.36$

TD1 – 3 :

The table below represents the distribution of the intelligence quotient (IQ) of 100 students. This distribution is grouped into 9 classes of width 10.

Class mark (mid-value)	59.5	69.5	79.5	89.5	99.5	109.5	119.5	129.5	139.5
frequency	1	2	9	22	33	22	8	2	1
	1	3	12	34	67	89	97	99	100

1) Calculate the mean and standard deviation of this distribution

2) Determine the values of the 1st and 3rd quartile of the distribution. Calculate the Semi-interquartile range.

3) Calculate the coefficient of symmetry $\gamma_1 = \frac{\mu_3}{\sigma^3}$ and the coefficient of kurtosis $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$. Compare these values to the standard normal distribution.

1) $$\left[59.5 + 2\times 69.5 + 9\times 79.5 + 22\times 89.5 + 99.5\times 33 + 109.5\times 22 + 119.5\times 8 + 129.5\times 2 + 139.5\times 1\right]\times\frac{1}{100}$$

$$= 99.3$$

$$\sqrt{\left[(59.5-99.3)^2 + 2\times(69.5-99.3)^2 + 9\times(79.5-99.3)^2 + 22\times(89.5-99.3)^2 + 33(99.5-99.3)^2 + 22\times(109.5-99.3)^2 + 8\times(119.5-99.3)^2 + 2\times(129.5-99.3)^2 + (139.5-99.3)^2\right]\times\frac{1}{100}}$$

$$= \sqrt{(39.8^2 + 2\times 29.8^2 + 9\times 19.8^2 + 22\times 9.8^2 + 33\times 0.2^2 + 22\times 10.2^2 + 8\times 20.2^2 + 2\times 30.2^2 + 40.2^2)\times\frac{1}{100}}$$

$$= \sqrt{(1584.04 + 1776.08 + 3528.36 + 2112.88 + 1.32 + 2288.88 + 3264.32 + 1824.08 + 1616.04)\times\frac{1}{100}}$$

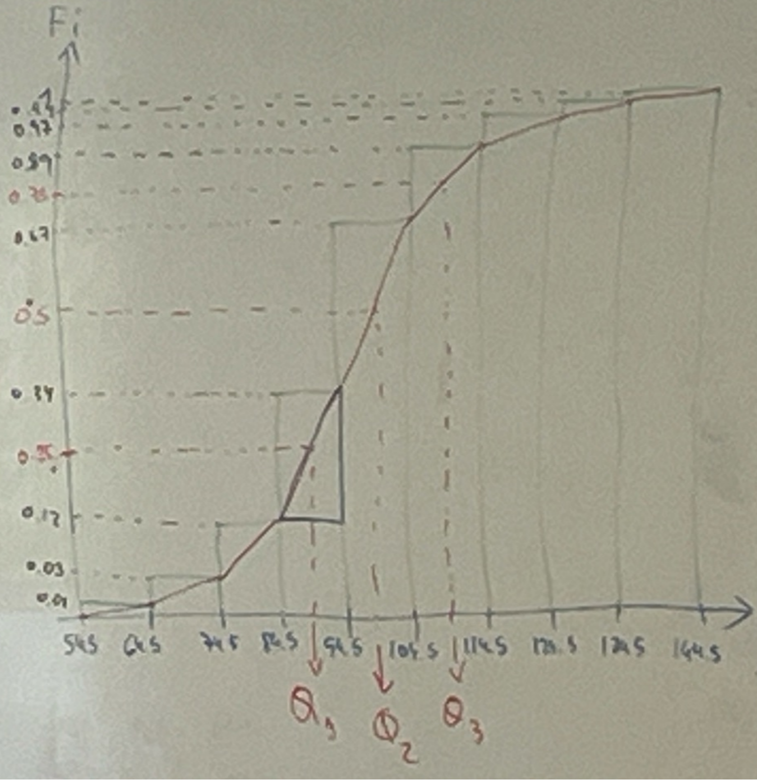
$$= \sqrt{17996\times\frac{1}{100}} = 13.415$$

$$f_i = \frac{n_i}{n}$$

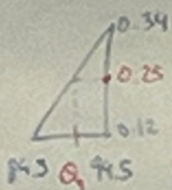
TD1-3

1) $\bar{x} \cong \sum_{i=1}^{9} x_i f_i = \frac{1}{n} \sum_{i=1}^{9} x_i n_i = 99.3 \qquad \hat{\sigma}^2 \cong 179.96 \qquad \hat{\sigma} \cong 13.41$

	$[x_i ; x_{i+1}]$	x_i	n_i	f_i	F_i
	54.5 ; 64.5	59.5	1	0.01	0.01
	64.5 , 74.5	69.5	2	0.02	0.03
	74.5 , 84.5	79.5	9	0.09	0.12
$Q_1 \in$	84.5 , 94.5	89.5	22	0.22	0.34
$Q_2 \in$	94.5 , 104.5	99.5	33	0.33	0.67
$Q_3 \in$	104.5 , 114.5	109.5	22	0.22	0.89
	114.5 , 124.5	119.5	8	0.08	0.97
	124.6 , 134.5	129.5	2	0.02	0.99
	134.5 , 144.5	139.5	1	0.01	1
			100		



$\underline{\underline{Q_1}}$

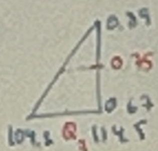


$$\frac{0.34 - 0.12}{94.5 - 84.5} = \frac{0.25 - 0.12}{Q_1 - 84.5} \qquad Q_1 = \frac{0.13 \times 10}{0.22} + 84.5 \cong 90.41$$

$\underline{\underline{Q_2}}$

$Q_2 = 99.35$

$\underline{\underline{Q_3}}$ $\quad Q_3 = 108.13$



$$\frac{0.22}{10} = \frac{0.09}{Q_3 - 104.5}$$

$$Q_3 \cong \frac{0.08 \times 10}{0.22} + 104.5 \cong 108.13$$

Semi-interquartile range $\quad \frac{Q_3 - Q_1}{2} = 8.86$

ISEP 2022/2023: Engineering cycle - 2^{nd} year

STATISTICS – Exercises SESSION N° 2

About chapters :

- Chapter 2 : Statistical theory of estimation

TD2 – 1 :

Let $x_1,\ldots,x_n$, be a sample of size n, results of independent observations of a random variable X whose probability law depends on a parameter. Determine the estimators of this parameter by the Maximum Likelihood method and calculate their precision for the following situations:

1) X follows an exponential law $f_X(x,\lambda) = \lambda e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise.
2) Let us define $a = 1/\lambda$. Calculate the minimum variance (Cramer - Rao bound) for an estimator of the parameter a. Draw a conclusion. $\hat{p}$ is an efficien of P (MVUE) minimum

2) X follows a Bernoulli law with parameter p

Remark : Here you must conclude by indicating whether the estimator is efficient and with minimum variance for the two situations.

TD2 – 2 :

An urn contains an unknown proportion of red balls and white balls. A sampling with replacement of size 60 gave a percentage of 70% for the red balls.

Calculate the 95% and 99.73% confidence limits of the proportion of red balls in the urn.

TD2 – 3 :

We assume that the random variable X that represents the weight of a given object follows a Gaussian distribution of mean *m* and standard deviation σ .

The following table shows the results of 10 weighings of the same object (in grams):

72.20	72.24	72.26	72.30	72.36	72.39	72.42	72.48	72.50	72.54

1) Calculate the point estimates of the mean *m* and of the standard deviation σ of the variable X.

2) The standard deviation is unknown but you can consider the estimated value in question 1), calculate a confidence interval at the threshold of 5% of the mean *m* .

TD2-1

$\ln(b/c) = \ln b - \ln c$
$\ln(bc) = \ln b + \ln c$

1) $X \sim \lambda e^{-\lambda x} \quad (x \geq 0)$

$\lambda = \frac{1}{a} \rightarrow$ sample mean

$X_1, \dots, X_n$

1° $L = f_{X_1 \sim X_n}(x_1, \dots, x_n, a) = \prod_{i=1}^{n} f_X(x_i, a) = \prod_{i=1}^{n} \frac{1}{a} e^{-\frac{1}{a}x}$

MLE

2° $\ln L = \sum_{i=1}^{n} \ln\left(\frac{1}{a} e^{-\frac{1}{a}x_i}\right) = \sum_{i=1}^{n}\left(-\ln a - \frac{1}{a}x_i\right)$

$= -n \ln a - \frac{\Sigma X_i}{a}$

3° $\frac{\partial \ln L}{\partial a} = -\frac{n}{a} + \frac{\Sigma X_i}{a^2} = 0$

$\Rightarrow \begin{cases} -\frac{n}{\hat{a}} + \frac{\Sigma X_i}{\hat{a}^2} = 0 \\ \hat{a} = \frac{\Sigma X_i}{n} = \bar{X} \end{cases}$

$E(\bar{X}) = E(X) = a$

$V(X) = a^2 \qquad V(\bar{X}) = \frac{a^2}{n}$

$E(X) = \int_0^{\infty} x f_X(x, a)\,dx = \int_0^{\infty} \frac{x}{a} e^{-\frac{x}{a}}\,dx$

$$\ln L = \sum_{i=1}^{n}\left(-\ln a - \frac{1}{a}x_i\right) = -n\ln a - \frac{\sum x_i}{a}$$

$X \sim \mathcal{E}(a)$

$$\begin{cases} f_X(x,a) = \frac{1}{a}e^{-\frac{x}{a}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$\ln\left(\frac{b}{c}\right) = \ln b - \ln c$

$\ln(bc) = \ln b - \ln c$

$\ln e^{c} = c$

$$\frac{\partial \ln L}{\partial a} = -\frac{n}{a} + \frac{\sum x_i}{a^2}$$

$$\frac{\partial \ln L}{\partial a} = 0 \qquad -\frac{n}{\hat{a}} + \frac{\sum x_i}{\hat{a}^2} = 0 \qquad \hat{a} = \frac{\sum x_i}{n} = \bar{X}$$

$$E(\bar{X}) = \boxed{E(X) = a} \qquad V(\bar{X}) = \frac{a^2}{n} \qquad V(X) = a^2$$

$$E(X) = \int_0^\infty x f_X(x,a)\,dx = \int_0^\infty \underbrace{\frac{x}{a}}_{u}\underbrace{e^{-\frac{x}{a}}dx}_{dv} = u\int v\,dx - \int\left[u'\int v\,dx\right]dx$$

$$= \frac{x}{a}\int e^{-x/a}dx - \int\left[\frac{1}{a}\int e^{-x/a}dx\right]dx$$

$$= \frac{x}{a}\,\frac{e^{-x/a}}{-1/a} + \int\left[\frac{1}{a}\,\frac{e^{-x/a}}{-1/a}\right]dx$$

$$= -x e^{-x/a} + \int e^{-x/a}dx$$

$$= \left[-x e^{-x/a} + \frac{e^{-x/a}}{-1/a}\right]_0^\infty$$

$$= \left[a e^{-x/a}\right]$$

$$= a$$

$$\boxed{\int u\,dv = uv - \int v\,du}$$

$$\int dv = \int e^{-\frac{x}{a}}dx \qquad du = \frac{dx}{a}$$

$$v = \frac{e^{-\frac{x}{a}}}{-\frac{1}{a}} = -ae^{-\frac{x}{a}}$$

$$= \frac{x}{a}\left(a e^{-\frac{x}{a}}\right)\Big|_0^{\infty} + \int_0^{\infty} \cancel{a} e^{-\frac{x}{a}} \frac{dx}{\cancel{a}} = a$$

r.v X

$$V(X) = E\left(X - E(X)\right)^2$$

$$V(X) = a^2$$

$$E(X^2) = \int_0^{\infty} x^2 f_X(x,a)\,dx = \int_0^{\infty} x^2 \frac{1}{a} e^{-\frac{x}{a}} dx = 2a^2$$

$$V(X) = E(X^2) - E^2(X) = a^2$$

2) $CRB(a) = \dfrac{1}{I(a)}$ $\quad I(a) = nE\left[\left(\dfrac{\partial \ln f_X(x,a)}{\partial a}\right)^2\right]$

$$I(a) = nE\left[\left(\frac{\partial}{\partial a} \ln\left(\frac{1}{a} e^{-\frac{x}{a}}\right)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial a}\left(-\ln a - \frac{x}{a}\right)\right)^2\right] = nE\left[\left(-\frac{1}{a} + \frac{x}{a^2}\right)^2\right]$$

$$= nE\left[\left(\frac{x-a}{a^2}\right)^2\right] = nE\left[\frac{(x-a)^2}{a^4}\right] = \frac{n}{a^4} E\left((X-a)^2\right) = \frac{n\cancel{a}}{a^4} = \frac{n}{a^2}$$

$$CRB(a) = \frac{a^2}{n}$$

$V(\bar{x}) = V(\hat{a}) = \frac{a^2}{n}$ $\quad CRB(a) = \frac{a^2}{n}$

$V(\bar{x}) \geq CRB(a)$

$\bar{X}$ is an efficient estimator, i.e., MVUE (Minimum variance unbiased estimator)

it is also convergent $\lim_{n \to \infty} V(\hat{a}) = 0$

$V(\bar{X}) = V(\hat{a}) = \frac{a^2}{n}$ $\quad CRB(a) = \frac{a^2}{n}$

$V(\bar{X}) \geq CRB(a)$

$\bar{X}$ is an efficient estimator, i.e., MVUE (Minimum variance unbiased estimator)

it is also convergent $\lim_{n\to\infty} V(\hat{a}) = 0$

2) Bernouilli r.v. $\quad X : \begin{cases} 1 & p \\ 0 & (1-p) \end{cases} \quad f_X(x,p) = p^x (1-p)^{1-x}$

$X \sim B(n,p)$

$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

$P(X=x) = p^x (1-p)^{1-x}$

n=1 Bernoulli distribution

$E(X) = p \quad V(X) = np(1-P)$

Estimator of Maximum likelihood (MLE)

$$L = f_{X_1,\dots X_n}(x_1,\dots,x_n,p) = \prod_{i=1}^{n} f_{X_i}(x_i,p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$L = p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}$$

$$\ln L = \ln\left(p^{\sum x_i}(1-p)^{n-\sum x_i}\right)$$

$$= \sum x_i \ln p + (n-\sum x_i)\ln(1-p)$$

$$\frac{\partial \ln L}{\partial p} = \frac{\sum x_i}{p} - \frac{(n-\sum x_i)}{1-p}$$

$$\frac{\partial \ln L}{\partial p} = 0$$

$$\frac{\sum x_i}{\hat p} = \frac{(n-\sum x_i)}{1-\hat p} \iff (1-\hat p)\sum x_i = \hat p(n-\sum x_i)$$

$$\sum x_i - \cancel{\hat p\sum x_i} = \hat p n - \cancel{\hat p\sum x_i}$$

$$\boxed{\hat p = \frac{\sum x_i}{n}} = \bar X$$

$$\boxed{E(\bar X) = E(X) \quad V(\bar X) = \frac{V(X)}{n}}$$

$$\boxed{E(\hat p) = p} \quad \boxed{V(\hat p) = \frac{p(1-p)}{n}}$$

→ $\hat p$ unbiased and convergent $\left(\lim_{n\to\infty} V(\hat p) = 0\right)$

Is $\hat{p}$ efficient?

$$I(p) = nE\left[\left(\frac{\partial}{\partial p}\ln f_X(x,p)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial p}\ln p^x(1-p)^{1-x}\right)^2\right]$$

$$= nE\left[\frac{\partial}{\partial p}\left(x\ln p + (1-x)\ln(1-p)\right)^2\right] = nE\left[\left(\frac{x}{p} - \frac{(1-x)}{1-p}\right)^2\right]$$

$$= nE\left[\left(\frac{x - \cancel{xp} - p + \cancel{px}}{p(1-p)}\right)^2\right] = \frac{n}{p^2(1-p)^2}E(X-p)^2 = \frac{n\,V(X)}{p^2(1-p)^2} = \frac{n\,\cancel{p}\cancel{(1-p)}}{p^{\cancel{2}}(1-p)^{\cancel{2}}} = \frac{n}{p(1-p)}$$

$$CRB(p) = \frac{1}{I(p)} = \frac{p(1-p)}{n} = V(\bar{X}) = \frac{p(1-p)}{n}$$

Conclusion

$$\boxed{\int_0^\infty \frac{x}{a}\cdot e^{-\frac{x}{a}}\,dx}$$

$$u=\frac{x}{a} \Rightarrow x=a\cdot u \Rightarrow dx=adu$$

$$\therefore \int_0^\infty u\cdot e^{-u}\,dx=\int_0^\infty u\cdot e^{-u}\cdot adu$$

令 $u=u$, $dv=e^{-u}\cdot adu$

$$\int dv=\int e^{-u}adu$$

$$v=-e^{-u}+C$$

对分部积分

$$\int u\cdot e^{-u}du=-u\cdot e^{-u}+\int e^{-u}du$$

$$=\left(-\frac{x}{a}\cdot e^{-\frac{x}{a}}\right)\Big|_0^\infty+\left(-a\cdot e^{-\frac{x}{a}}\right)\Big|_0^\infty$$

$$=(0-0)-(0-a)=a$$

TD2 – 2 :

An urn contains an unknown proportion of red balls and white balls. A sampling with replacement of size 60 gave a percentage of 70% for the red balls.

Calculate the 95% and 99.73% confidence limits of the proportion of red balls in the urn.

1. 计算结果：

- 样本均值 m（估计的均值）：72.369 克
- 样本标准差 s（估计的标准差）：0.117 克

接下来，我们将使用这些点估计值来计算均值 m 的95%置信区间。我们将使用t分布因为样本容量小于30并且总体标准差未知。[>-]

1. 使用t分布得出的95%置信区间约为 72.285 克到 72.453 克。

下一步是，在已知方差的情况下，使用z分布重新计算95%置信区间并与之前的区间进行比较。[>-]

使用z分布得出的95%置信区间约为 72.296 克到 72.442 克。

现在，对于第三个问题，我们需要根据给定的置信区间 $[72.31; 72.43]$ 确定显著性水平 α。因为这里给出了标准差 $\sigma = 0.08$，我们将使用这个值来确定z分数，然后找到相应的 α。[>-]

1. 根据给定的置信区间 $[72.31; 72.43]$，计算得出的显著性水平 α 约为 1.59%.

这表示，如果我们设置显著性水平为 1.59%，则我们可以得到一个双尾测试的置信区间 $[72.31; 72.43]$。[>-]

TD2 – 3 :

We assume that the random variable X that represents the weight of a given object follows a Gaussian distribution of mean m and standard deviation σ .

The following table shows the results of 10 weighings of the same object (in grams):

72.20	72.24	72.26	72.30	72.36	72.39	72.42	72.48	72.50	72.54

1) Calculate the point estimates of the mean m and of the standard deviation σ of the variable X.

2) The standard deviation is unknown but you can consider the estimated value in question 1), calculate a confidence interval at the threshold of 5% of the mean m .

Now, consider the variance is known and equal to the value calculated in 1), how does the interval change? Calculate and compare the 2 intervals.

3) The standard deviation of the balance, measured from many previous studies, is indeed, for an object of about this mass, 0.08. In this question, we consider then $\sigma = 0.08$. Determine the level of significance α so that the resulting confidence interval is $[72.31; 72{,}43]$.

$X \sim N(m, \sigma^2)$

for the expected value m : $\bar{X} = \frac{\sum x_i}{n}$ $\bar{X} = 72.37$

for the variance σ^2 : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.0137$

$S \simeq 0.117$

2) CI : $\left[\bar{x} - t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} ; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$

$t_{1-\frac{0.05}{2}} = t_{0.975} = 2.26$

CI: $\left[72.37 - 2.26 \times \frac{0.1172}{\sqrt{10}} ; 72.37 + 2.26 * \frac{0.1172}{\sqrt{10}}\right]$

$[72.29 ; 72{,}45]$

TD2-3

$X \sim \mathcal{N}(m, \sigma^2)$

1) For the expected value m: $\bar{X} = \frac{\sum X_i}{n}$ $\bar{x} = 72.37$

For the variance σ^2: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \cong 0.0137$

$S \cong 0.1172$

VARIANCE UNKNOWN

2) CI: $\left[\bar{x} - t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \, ; \, \bar{x} + t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$

↓

quantile of the Student-t-distribution of order $\left(1-\frac{\alpha}{2}\right)$ with 9 d.f.

$t_{1-\frac{0.05}{2}} = t_{0.975} = 2.26$

CI: $\left[72.37 - 2.26 \times \frac{0.1172}{\sqrt{10}} \; ; \; 72.37 + 2.26 \times \frac{0.1172}{\sqrt{10}} \right]$

$[72.29 \; ; \; 72.45]$

VARIANCE KNOWN $\sigma^2 = 0.0137$

CI: $\left[\bar{x} - t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \; ; \; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$

→ quantile of order $\left(1 - \frac{\alpha}{2}\right)$ of the standard normal distribution

$t_{97.5\%} = 1.96$

$\left[72.37 - 1.96 \times \frac{0.1172}{\sqrt{10}} \; ; \; 72.37 + 1.96 \times \frac{0.1172}{\sqrt{10}} \right] = [72.30 \; ; \; 72.44]$

The 2nd interval is smaller than the 1st one.

TD2-3

$X \sim \mathcal{N}(m, \sigma^2)$

1) For the expected value m: $\bar{X} = \frac{\Sigma X_i}{n}$ $\quad \bar{x} = 72.37$

For the variance σ^2: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \simeq 0.0137$

$S \simeq 0.1172$

3) $[72.31 ; 72.43]$ σ is known $\sigma = 0.08$

$\left[\bar{x} - t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} , \bar{x} + t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$

$72.31 = \bar{x} - t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$t_{1-\frac{\alpha}{2}} = \frac{(\bar{x} - 72.31)}{\frac{\sigma}{\sqrt{n}}} = \frac{(72.37 - 72.31)\sqrt{10}}{0.08} \simeq 2.37$

$1 - \frac{\alpha}{2} = 0.9911 \quad \alpha \simeq 1.7\%$

ChatGPT

To answer the exam questions presented in the image:

1. **Calculate the point estimates of the mean m and of the standard deviation σ of the variable X:**
   - To find the mean m, sum all the weighings and divide by the number of observations.
   - To find the estimated standard deviation s, use the formula for the sample standard deviation.
2. **Calculate a 95% confidence interval for the mean m:**
   - Since the standard deviation is unknown and the sample size is small, use the t-distribution.
   - Find the t-score corresponding to a 95% confidence level for 9 degrees of freedom.
   - Calculate the confidence interval using the formula $\bar{x} \pm t \times \frac{s}{\sqrt{n}}$.
3. **If the variance is known and equal to the value calculated in 1):**
   - Calculate the 95% confidence interval for the mean using the z-score instead of the t-score since the variance is known.
4. **Determine the level of significance α:**
   - Use the known standard deviation $\sigma = 0.08$.
   - Find the z-scores that correspond to the confidence interval $[72.31; 72.43]$.
   - Use the z-scores to determine the level of significance α.

Let's perform the calculations for each step. We'll start with the first one, calculating the mean and standard deviation.

Now, consider the variance is known and equal to the value calculated in 1), how does the interval change? Calculate and compare the 2 intervals.

3) The standard deviation of the balance, measured from many previous studies, is indeed, for an object of about this mass, 0.08. In this question, we consider then $\sigma = 0.08$. Determine the level of significance α so that the resulting confidence interval is $[72.31; 72{,}43]$.

TD2 – 4 :

Consider the following sample:

Sample : (1.28 ; 2.16 ; 0.75 ; 1.44 ; 2.05 ; 0.65 ; 1.26 ; 1.73 ; 1.81 ; 0.92)

Calculate the sample mean and sample variance. Deduce the confidence intervals for the mean, variance and standard deviation of the population. We will assume that the random variable follows a Gaussian law.

TD2 – 4 :
Consider the following sample:

Sample : (1.28 ; 2.16 ; 0.75 ; 1.44 ; 2.05 ; 0.65 ; 1.26 ; 1.73 ; 1.81 ; 0.92)

Calculate the sample mean and sample variance. Deduce the confidence intervals for the mean, variance and standard deviation of the population. We will assume that the random variable follows a Gaussian law.

① $$\frac{1.28+2.16+0.75+1.44+2.05+0.65+1.26+1.73+1.81+0.92}{10}$$

$$= \frac{14.05}{10} = 1.405$$

② $$\big[(1.405-1.28)^2+(2.16-1.405)^2+(0.75-1.405)^2+(1.44-1.405)^2$$
$$+(2.05-1.405)^2+(0.65-1.405)^2+(1.26-1.405)^2+(1.73-1.405)^2$$
$$+(1.81-1.405)^2+(0.92-1.405)^2\big] \div 10$$

TD2-4)

- Sample mean $\overline{x} = 1.405$
- " variance $S^2 = 0.281$, $S \simeq 0.53$

- CI for the expected value m

- $\left[\overline{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \,;\, \overline{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right] = \left[\ 1.026 \ ;\ 1.784 \ \right]$

$t_{1-\frac{\alpha}{2}}$: quantile of order $\left(1-\frac{\alpha}{2}\right) = 97.5\%$ of the Student t-distribution with 9 dof

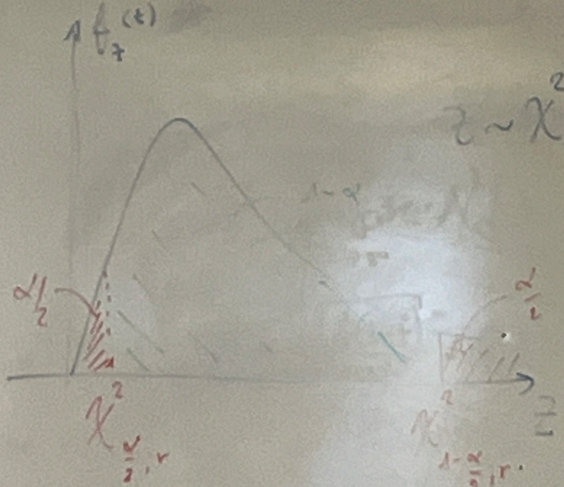
$t_{1-\frac{\alpha}{2}} = 2.26$

- CI for the variance σ^2

r.v. $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$

CI: $\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right]$

TD2-3



$Z \sim \chi^2_r$

$\chi^2_{2.5\%,9} = 2.7$

$\chi^2_{97.5\%,9} = 19.0$

CI for the σ :

$[0.36 \; ; \; 0.97]$

1. **计算样本均值 $\bar{x}$:**
   $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1.28+2.16+0.75+1.44+2.05+0.65+1.26+1.73+1.81+0.92}{10} = 1.405$
2. **计算样本方差 s^2:**
   $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
   其中 x_i 是单个样本观测值，n 是样本大小（此处为10）。
3. **均值的95%置信区间:**
   - 使用t分布找到95%置信水平的t值（自由度为 $n-1$）。
   - 使用公式 $\bar{x} \pm t_{\frac{\alpha}{2},n-1} \times \frac{s}{\sqrt{n}}$ 来计算置信区间。
4. **方差和标准差的95%置信区间:**
   - 使用卡方分布找到95%置信水平的下界和上界卡方值（自由度为 $n-1$）。
   - 使用公式 $\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}$ 和 $\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$ 来计算方差的置信区间。
   - 方差置信区间的平方根即为标准差的置信区间。

具体计算中，样本均值为 1.405，样本方差为 0.2809。均值的95%置信区间为 1.026 到 1.784，方差的95%置信区间为 0.1329 到 0.9361，标准差的95%置信区间为 0.3645 到 0.9675。

STATISTICS – Exercises SESSION N° 3

About Chapters :

- Chapter 3 : Hypothesis testing

TD3 – 1 :

During an extrasensory perception experiment, a person is isolated in a room and tasked to guess the color (red or blue) of a card chosen randomly from a 50-card deck. The cards were previously well shuffled. The person does not know the total number of blue or red cards in the deck.

Suppose the person correctly guessed the color of 32 cards. Determine if it is possible to conclude that the person possesses extrasensory perception (sixth sense) at the 0.05 and 0.01 significance level.

TD3 – 2 :

A manufacturer delivers a batch of bulbs with an average lifetime specification of 1600 hours. After inspection on a sample of 100 bulbs, the average lifespan is 1,578 hours with a standard deviation of 120 hours.

1) Can the delivery be considered to meet specification (that is, the average lifetime is 1600 hours) at the 5% significance level?

2) In fact, what the buyer is interested in is whether the lifetime is equal to or greater than the specification. Can the delivery be considered to meet this condition at the 5% significance level?

TD3 – 3:

You are hired by TOYOTA for a project to design a low-consumption car model. First of all, it is necessary to ensure that fuel consumption is the primary factor behind the purchase of this type of car. After performing a survey of 120 owners of this type of car, 40 considered that the criterion "consumption" was the decisive criterion of purchase.

1) Let *p be* the proportion of owners who favor the “consumption” criterion. You are tasked to test the following hypotheses at a given level of significance α :

H_0: $p = 0.4$

H_1: $p \neq 0.4$

Suppose that the proportion of owners who favored the criterion " consumption" within the total population is exactly 40%. What is the probability of concluding that H_0 should be rejected? Test these hypotheses by considering the level of significance $\alpha = 0.05$.

2) Build a confidence interval of 95% for the proportion of owners whose choice was motivated mainly by the fuel consumption.

3) The company claims that, for this model of cars, the average fuel consumption μ is equal to 7 liters per 100 km on urban roads (50 km/h). On a random sample of 5 cars (the population is assumed to be normal), the consumption was (in liters per 100 km): 8.1; 8.5; 8.9; 9.7; 9.8

At the threshold $\alpha = 0.05$, do you confirm the consumption announced by the manufacturer?

TD3 – 4 *:

Two groups A and B are each composed of 100 people with the same disease. A new drug was given to group A and placebo treatment to group B (referred to as a control group). It was observed that 75 patients from group A and 65 from group B were cured.

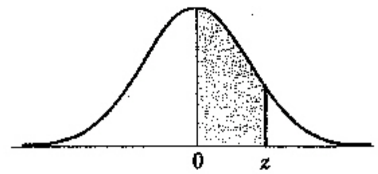
1) Test the hypothesis that the new drug is effective in the cure of the disease by considering a significance level of 1%.

2) Answer the previous question by considering groups of 300 people, and the fact that there were 225 people healed in group A and 195 in group B. Any Comments? Compare both results.

TABLES

Last update on October 27th, 2022

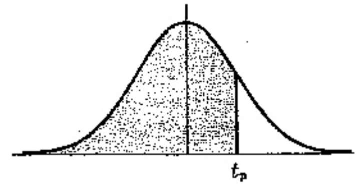
Aires under de density curve
of the STANDARD
NORMAL DISTRIBUTION
From 0 to z



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Cumulative distribution function of the Students t law

Values of the centiles for the STUDENT'S t distribution with ν degrees of freedom (Aire in gray = p)

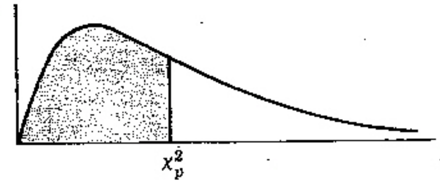


ν	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$	$t_{0,80}$	$t_{0,75}$	$t_{0,70}$	$t_{0,60}$	$t_{0,55}$
1	63,66	31,82	12,71	6,31	3,08	1,376	1,000	0,727	0,325	0,158
2	9,92	6,96	4,30	2,92	1,89	1,061	0,816	0,617	0,289	0,142
3	5,84	4,54	3,18	2,35	1,64	0,978	0,765	0,584	0,277	0,137
4	4,60	3,75	2,78	2,13	1,53	0,941	0,741	0,569	0,271	0,134
5	4,03	3,36	2,57	2,02	1,48	0,920	0,727	0,559	0,267	0,132
6	3,71	3,14	2,45	1,94	1,44	0,906	0,718	0,553	0,265	0,131
7	3,50	3,00	2,36	1,90	1,42	0,896	0,711	0,549	0,263	0,130
8	3,36	2,90	2,31	1,86	1,40	0,889	0,706	0,546	0,262	0,130
9	3,25	2,82	2,26	1,83	1,38	0,883	0,703	0,543	0,261	0,129
10	3,17	2,76	2,23	1,81	1,37	0,879	0,700	0,542	0,260	0,129
11	3,11	2,72	2,20	1,80	1,36	0,876	0,697	0,540	0,260	0,129
12	3,06	2,68	2,18	1,78	1,36	0,873	0,695	0,539	0,259	0,128
13	3,01	2,65	2,16	1,77	1,35	0,870	0,694	0,538	0,259	0,128
14	2,98	2,62	2,14	1,76	1,34	0,868	0,692	0,537	0,258	0,128
15	2,95	2,60	2,13	1,75	1,34	0,866	0,691	0,536	0,258	0,128
16	2,92	2,58	2,12	1,75	1,34	0,865	0,690	0,535	0,258	0,128
17	2,90	2,57	2,11	1,74	1,33	0,863	0,689	0,534	0,257	0,128
18	2,88	2,55	2,10	1,73	1,33	0,862	0,688	0,534	0,257	0,127
19	2,86	2,54	2,09	1,73	1,33	0,861	0,688	0,533	0,257	0,127
20	2,84	2,53	2,09	1,72	1,32	0,860	0,687	0,533	0,257	0,127
21	2,83	2,52	2,08	1,72	1,32	0,859	0,686	0,532	0,257	0,127
22	2,82	2,51	2,07	1,72	1,32	0,858	0,686	0,532	0,256	0,127
23	2,81	2,50	2,07	1,71	1,32	0,858	0,685	0,532	0,256	0,127
24	2,80	2,49	2,06	1,71	1,32	0,857	0,685	0,531	0,256	0,127
25	2,79	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
26	2,78	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
27	2,77	2,47	2,05	1,70	1,31	0,855	0,684	0,531	0,256	0,127
28	2,76	2,47	2,05	1,70	1,31	0,855	0,683	0,530	0,256	0,127
29	2,76	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
30	2,75	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
40	2,70	2,42	2,02	1,68	1,30	0,851	0,681	0,529	0,255	0,126
60	2,66	2,39	2,00	1,67	1,30	0,848	0,679	0,527	0,254	0,126
120	2,62	2,36	1,98	1,66	1,29	0,845	0,677	0,526	0,254	0,126
∞	2,58	2,33	1,96	1,645	1,28	0,842	0,674	0,524	0,253	0,126

D'après R.A. Fisher et F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (5 th édition), Table III, Oliver and Boyd Ltd., Edinburgh.

Cumulative distribution function of the CHI-2 distribution

Values of the centiles (χ_p^2)
for the CHI-2 distribution with ν degrees of freedom
(Aire in gray = p)



ν	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$	$\chi^2_{0,75}$	$\chi^2_{0,50}$	$\chi^2_{0,25}$	$\chi^2_{0,10}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	7,88	6,63	5,02	3,84	2,71	1,32	0,455	0,102	0,0158	0,0039	0,0010	0,0002	0,0000
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	0,575	0,211	0,103	0,0506	0,0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0,216	0,115	0,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	23,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66,8	63,7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79,5	76,2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92,0	88,4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51,7	48,8	45,4	43,3
80	116,3	112,3	106,6	101,9	96,6	88,1	79,3	71,1	64,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3

D'après Catherine M. Thompson, *Table of percentage points of the χ^2 distribution*, Biometrika, Vol. 32 (1941).