Essentials of Statistics

David Brink

David Brink

Statistics

Statistics

ISBN 978-87-7681-408-3

Contents

1 Preface

Many students find that the obligatory Statistics course comes as a shock. The set textbook is difficult, the curriculum is vast, and secondary-school maths feels infinitely far away.

"Statistics" offers friendly instruction on the core areas of these subjects. The focus is overview. And the numerous examples give the reader a "recipe" for solving all the common types of exercise. You can download this book free of charge.

2 Basic concepts of probability theory

2.1 Probability space, probability function, sample space, event

A **probability space** is a pair (Ω, P) consisting of a set Ω and a function P which assigns to each subset A of Ω a real number $P(A)$ in the interval $[0,1]$. Moreover, the following two axioms are required to hold:

1. $P(\Omega) = 1$,

2. $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ if $A_1, A_2, \ldots$ is a sequence of pairwise disjoint subsets of Ω.

The set Ω is called a **sample space**. The elements $\omega \in \Omega$ are called **sample points** and the subsets $A \subseteq \Omega$ are called **events**. The function P is called a **probability function**. For an event A, the real number $P(A)$ is called the **probability** of A.

From the two axioms the following consequences can be deduced:

3. $P(\emptyset) = 0$,

4. $P(A \backslash B) = P(A) - P(B)$ if $B \subseteq A$,

5. $P(\complement A) = 1 - P(A)$,

6. $P(A) \geqq P(B)$ if $B \subseteq A$,

7. $P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$ if $A_1, \ldots, A_n$ are pairwise disjoint events,

8. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for arbitrary events A and B.

EXAMPLE. Consider the set $\Omega = \{1,2,3,4,5,6\}$. For each subset A of Ω, define

$$P(A) = \frac{\#A}{6} \ ,$$

where $\#A$ is the number of elements in A. Then the pair (Ω, P) is a probability space. One can view this probability space as a model for the for the situation "throw of a dice".

EXAMPLE. Now consider the set $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$. For each subset A of Ω, define

$$P(A) = \frac{\#A}{36} \ .$$

Now the probability space (Ω, P) is a model for the situation "throw of two dice". The subset

$$A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$

is the event "a pair".

2.2 Conditional probability

For two events A and B the **conditional probability of** A **given** B is defined as

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)} \ .$$

We have the following theorem called *computation of probability by division into possible causes*: Suppose $A_1, \ldots, A_n$ are pairwise disjoint events with $A_1 \cup \cdots \cup A_n = \Omega$. For every event B it then holds that

$$P(B) = P(A_1) \cdot P(B \mid A_1) + \cdots + P(A_n) \cdot P(B \mid A_n) \, .$$

EXAMPLE. In the French Open final, Nadal plays the winner of the semifinal between Federer and Davydenko. A bookmaker estimates that the probability of Federer winning the semifinal is 75%. The probability that Nadal can beat Federer is estimated to be 51%, whereas the probability that Nadal can beat Davydenko is estimated to be 80%. The bookmaker therefore computes the probability that Nadal wins the French Open, using division into possible causes, as follows:

$$\begin{aligned} P(\text{Nadal wins the final}) &= P(\text{Federer wins the semifinal}) \times \\ & P(\text{Nadal wins the final} | \text{Federer wins the semifinal}) + \\ & P(\text{Davydenko wins the semifinal}) \times \\ & P(\text{Nadal wins the final} | \text{Davydenko wins the semifinal}) \\ &= 0.75 \cdot 0.51 + 0.25 \cdot 0.8 \\ &= 58.25\% \end{aligned}$$

2.3 Independent events

Two events A and B are called **independent**, if

$$P(A \cap B) = P(A) \cdot P(B) .$$

Equivalent to this is the condition $P(A \mid B) = P(A)$, i.e. that the probability of A is the same as the conditional probability of A given B.

> **Remember:** Two events are independent if the probability of one of them is not affected by knowing whether the other has occurred or not.

EXAMPLE. A red and a black dice are thrown. Consider the events

A: red dice shows 6,
B: black dice show 6.

Since

$$P(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(A) \cdot P(B) ,$$

A and B are independent. The probability that the red dice shows 6 is not affected by knowing anything about the black dice.

EXAMPLE. A red and a black dice are thrown. Consider the events

A: the red and the black dice show the same number,
B: the red and the black dice show a total of 10.

Since

$$P(A) = \frac{1}{6} , \text{ but } P(A \mid B) = \frac{1}{3} ,$$

A and B are not independent. The probability of two of a kind increases if one knows that the sum of the dice is 10.

2.4 The Inclusion-Exclusion Formula

Formula **8** on page 12 has the following generalization to three events A, B, C:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) .$$

This equality is called the *Inclusion-Exclusion Formula* for three events.

EXAMPLE. What is the probability of having at least one 6 in three throws with a dice? Let A_1 be the event that we get a 6 in the first throw, and define A_2 and A_3 similarly. Then, our probability can be computed by inclusion-exclusion:

$$\begin{aligned} P &= P(A_1 \cup A_2 \cup A_3) \\ &= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{1}{6^2} - \frac{1}{6^2} - \frac{1}{6^2} + \frac{1}{6^3} \\ &\approx 41\% \end{aligned}$$

The following generalization holds for n events $A_1, A_2, \ldots, A_n$ with union $A = A_1 \cup \cdots \cup A_n$:

$$P(A) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots \pm P(A_1 \cap \cdots \cap A_n)\ .$$

This equality is called the *Inclusion-Exclusion Formula* for n events.

EXAMPLE. Pick five cards at random from an ordinary pack of cards. We wish to compute the probability $P(B)$ of the event B that all four suits appear among the 5 chosen cards.

For this purpose, let A_1 be the event that none of the chosen cards are spades. Define A_2, A_3, and A_4 similarly for hearts, diamonds, and clubs, respectively. Then

$$\complement B = A_1 \cup A_2 \cup A_3 \cup A_4\ .$$

The Inclusion-Exclusion Formula now yields

$$P(\complement B) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - P(A_1 \cap A_2 \cap A_3 \cap A_4)\ ,$$

that is

$$P(\complement B) = 4 \cdot \frac{\binom{39}{5}}{\binom{52}{5}} - 6 \cdot \frac{\binom{26}{5}}{\binom{52}{5}} + 4 \cdot \frac{\binom{13}{5}}{\binom{52}{5}} - 0 \approx 73.6\%$$

We thus obtain the probability

$$P(B) = 1 - P(\complement B) = 26.4\%$$

EXAMPLE. A school class contains n children. The teacher asks all the children to stand up and then sit down again on a random chair. Let us compute the probability $P(B)$ of the event B that each pupil ends up on a new chair.

We start by enumerating the pupils from 1 to n. For each i we define the event

$$A_i : \text{pupil number } i \text{ gets his or her old chair}$$

Then

$$\complement B = A_1 \cup \cdots \cup A_n\ .$$

Now $P(\complement B)$ can be computed by the Inclusion-Exclusion Formula for n events:

$$P(\complement B) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \cdots \pm P(A_1 \cap \cdots \cap A_n)\ ,$$

thus

$$\begin{aligned} P(\complement B) &= \binom{n}{1}\frac{1}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \cdots \pm \binom{n}{n}\frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \cdots \pm \frac{1}{n!} \end{aligned}$$

We conclude

$$P(B) = 1 - P(\complement B) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots \pm \frac{1}{n!}$$

It is a surprising fact that this probability is more or less independent of n: $P(B)$ is very close to 37% for all $n \geq 4$.

2.5 Binomial coefficients

The binomial coefficient $\begin{pmatrix} n \\ k \end{pmatrix}$ (read as "n over k") is defined as

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1 \cdot 2 \cdot 3 \cdots n}{1 \cdot 2 \cdots k \cdot 1 \cdot 2 \cdots (n-k)}$$

for integers n and k with $0 \leqq k \leqq n$. (Recall the convention $0! = 1$.)

The reason why binomial coefficients appear again and again in probability theory is the following theorem:

The number of ways of choosing k elements from a set of n elements is $\begin{pmatrix} n \\ k \end{pmatrix}$.

For example, the number of subsets with 5 elements (poker hands) of a set with 52 elements (a pack of cards) is equal to

$$\binom{52}{5} = 2598960\ .$$

An easy way of remembering the binomial coefficients is by arranging them in **Pascal's triangle** where each number is equal to the sum of the numbers immediately above:

$$\begin{array}{c}
\binom{0}{0} \\
\binom{1}{0}\ \binom{1}{1} \\
\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2} \\
\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3} \\
\binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4} \\
\binom{5}{0}\ \binom{5}{1}\ \binom{5}{2}\ \binom{5}{3}\ \binom{5}{4}\ \binom{5}{5} \\
\binom{6}{0}\ \binom{6}{1}\ \binom{6}{2}\ \binom{6}{3}\ \binom{6}{4}\ \binom{6}{5}\ \binom{6}{6} \\
\vdots
\end{array}
\qquad
\begin{array}{c}
1 \\
1\ 1 \\
1\ 2\ 1 \\
1\ 3\ 3\ 1 \\
1\ 4\ 6\ 4\ 1 \\
1\ 5\ 10\ 10\ 5\ 1 \\
1\ 6\ 15\ 20\ 15\ 6\ 1 \\
\vdots
\end{array}$$

One notices the rule

$$\binom{n}{n-k} = \binom{n}{k}\ , \text{ e.g. } \binom{10}{7} = \binom{10}{3}\ .$$

2.6 Multinomial coefficients

The multinomial coefficients are defined as

$$\binom{n}{k_1\ \cdots\ k_r} = \frac{n!}{k_1!\cdots k_r!}$$

for integers n and $k_1, \dots, k_r$ with $n = k_1 + \cdots + k_r$. The multinomial coefficients are also called *generalized binomial coefficients* since the binomial coefficient

$$\binom{n}{k}$$

is equal to the multinomial coefficient

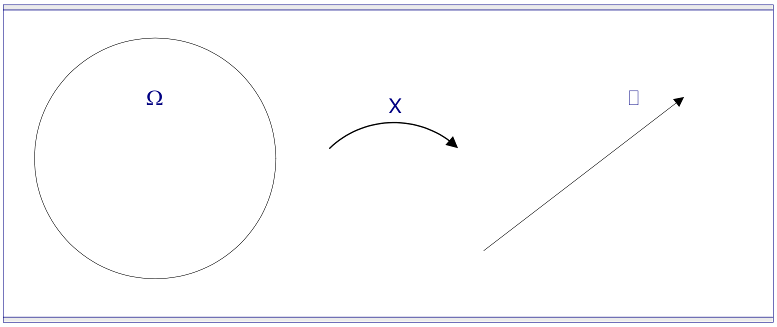
$$\binom{n}{k\quad l}$$

with $l = n - k$.

3 Random variables

3.1 Random variables, definition

Consider a probability space (Ω, P). A **random variable** is a map X from Ω into the set of real numbers $\mathbb{R}$.



Normally, one can forget about the probability space and simply think of the following rule of thumb:

Remember: A random variable is a function taking different values with different probabilities.

The probability that the random variable X takes certain values is written in the following way:

$P(X = x)$: the probability that X takes the value $x \in \mathbb{R}$,
$P(X < x)$: the probability that X takes a value smaller than x,
$P(X > x)$: the probability that X takes a value greater than x,
etc.

One has the following rules:

$$\begin{aligned} P(X \leq x) &= P(X < x) + P(X = x) \\ P(X \geq x) &= P(X > x) + P(X = x) \\ 1 &= P(X < x) + P(X = x) + P(X > x) \end{aligned}$$

3.2 The distribution function

The **distribution function** of a random variable X is the function $F : \mathbb{R} \to \mathbb{R}$ given by

$$F(x) = P(X \leq x) \, .$$

$F(x)$ is an increasing function with values in the interval $[0, 1]$ and moreover satisfies $F(x) \to 1$ for $x \to \infty$, and $F(x) \to 0$ for $x \to -\infty$.

By means of $F(x)$, all probabilities of X can be computed:

$$\begin{aligned} P(X < x) &= \lim_{\varepsilon \to 0} F(x - \varepsilon) \\ P(X = x) &= F(x) - \lim_{\varepsilon \to 0} F(x - \varepsilon) \\ P(X \geq x) &= 1 - \lim_{\varepsilon \to 0} F(x - \varepsilon) \\ P(X > x) &= 1 - F(x) \end{aligned}$$

3.3 Discrete random variables, point probabilities

A random variable X is called **discrete** if it takes only finitely many or countably many values. For all practical purposes, we may define a discrete random variable as a random variable taking only values in the set $\{0, 1, 2, \dots\}$. The **point probabilities**

$$P(X = k)$$

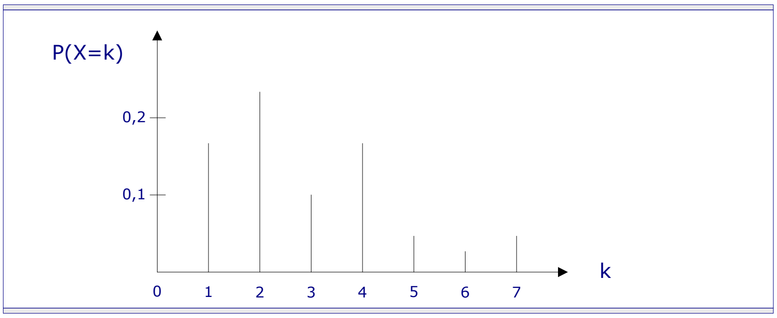
determine the distribution of X. Indeed,

$$P(X \in A) = \sum_{k \in A} P(X = k)$$

for any $A \subseteq \{0, 1, 2, \dots\}$. In particular we have the rules

$$\begin{aligned} P(X \leq k) &= \textstyle\sum_{i=0}^{k} P(X = i) \\ P(X \geq k) &= \textstyle\sum_{i=k}^{\infty} P(X = i) \end{aligned}$$

The point probabilities can be graphically illustrated by means of a **pin diagram**:



3.4 Continuous random variables, density function

A random variable X is called **continuous** if it has a **density function** $f(x)$. The density function, usually referred to simply as the **density**, satisfies

$$P(X \in A) = \int_{t \in A} f(t)dt$$

for all $A \subseteq \mathbb{R}$. If A is an interval $[a, b]$ we thus have

$$P(a \leq X \leq b) = \int_a^b f(t)dt \; .$$

One should think of the density as the continuous analogue of the point probability function in the discrete case.

3.5 Continuous random variables, distribution function

For a continuous random variable X with density $f(x)$ the distribution function $F(x)$ is given by

$$F(x) = \int_{-\infty}^x f(t)dt \; .$$

The distribution function satisfies the following rules:

$$\begin{aligned} P(X \leq x) &= F(x) \\ P(X \geq x) &= 1 - F(x) \\ P(|X| \leq x) &= F(x) - F(-x) \\ P(|X| \geq x) &= F(-x) + 1 - F(x) \end{aligned}$$

3.6 Independent random variables

Two random variables X and Y are called **independent** if the events $X \in A$ and $Y \in B$ are independent for any subsets $A, B \subseteq \mathbb{R}$. Independence of three or more random variables is defined similarly.

Remember: X and Y are independent if nothing can be deduced about the value of Y from knowing the value of X.

EXAMPLE. Throw a red dice and a black dice and consider the random variables

X: number of pips of red dice,
Y: number of pips of black dice,
Z: number of pips of red and black dice in total.

X and Y are independent since we can deduce nothing about X by knowing Y. In contrast, X and Z are not independent since information about Z yields information about X (if, for example, Z has the value 10, then X necessarily has one of the values 4, 5 and 6).

3.7 Random vector, simultaneous density, and distribution function

If $X_1, \dots, X_n$ are random variables defined on the same probability space (Ω, P) we call $\mathbf{X} = (X_1, \dots, X_n)$ an (n-dimensional) **random vector**. It is a map

$$\mathbf{X} : \Omega \to \mathbb{R}^n \ .$$

The **simultaneous** (n-dimensional) distribution function is the function $\mathbf{F} : \mathbb{R}^n \to [0, 1]$ given by

$$\mathbf{F}(x_1, \dots, x_n) = P(X_1 \leq x_1 \wedge \dots \wedge X_n \leq x_n) \ .$$

Suppose now that the X_i are continuous. Then $\mathbf{X}$ has a **simultaneous** (n-dimensional) density $\mathbf{f} : \mathbb{R}^n \to [0, \infty[$ satisfying

$$P(\mathbf{X} \in A) = \int_{\mathbf{x} \in A} \mathbf{f}(\mathbf{x})\, d\mathbf{x}$$

for all $A \subseteq \mathbb{R}^n$. The individual densities f_i of the X_i are called **marginal** densities, and we obtain them from the simultaneous density by the formula

$$f_1(x_1) = \int_{\mathbb{R}^{n-1}} \mathbf{f}(x_1, \dots, x_n)\, dx_2 \dots dx_n$$

stated here for the case $f_1(x_1)$.

Remember: The marginal densities are obtained from the simultaneous density by “integrating away the superfluous variables”.

4 Expected value and variance

4.1 Expected value of random variables

The **expected value** of a **discrete** random variable X is defined as

$$E(X) = \sum_{k=1}^{\infty} P(X = k) \cdot k \ .$$

The expected value of a **continuous** random variable X with density $f(x)$ is defined as

$$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x \, dx \ .$$

Often, one uses the Greek letter μ (“mu”) to denote the expected value.

4.2 Variance and standard deviation of random variables

The **variance** of a random variable X with expected value $E(X) = \mu$ is defined as

$$\mathrm{var}(X) = E((X - \mu)^2) \; .$$

If X is discrete, the variance can be computed thus:

$$\mathrm{var}(X) = \sum_{k=0}^{\infty} P(X = k) \cdot (k - \mu)^2 \; .$$

If X is continuous with density $f(x)$, the variance can be computed thus:

$$\mathrm{var}(X) = \int_{-\infty}^{\infty} f(x)(x - \mu)^2 \, dx \; .$$

The **standard deviation** σ ("sigma") of a random variable X is the square root of the variance:

$$\sigma(X) = \sqrt{\mathrm{var}(X)} \; .$$

4.3 Example (computation of expected value, variance, and standard deviation)

EXAMPLE 1. Define the discrete random variable X as the number of pips shown by a certain dice. The point probabilities are $P(X = k) = 1/6$ for $k = 1, 2, 3, 4, 5, 6$. Therefore, the expected value is

$$E(X) = \sum_{k=1}^{6} \frac{1}{6} \cdot k = \frac{1+2+3+4+5+6}{6} = 3.5 \,.$$

The variance is

$$\text{var}(X) = \sum_{k=1}^{6} \frac{1}{6} \cdot (k-3.5)^2 = \frac{(1-3.5)^2 + (2-3.5)^2 + \cdots + (6-3.5)^2}{6} = 2.917 \,.$$

The standard deviation thus becomes

$$\sigma(X) = \sqrt{2.917} = 1.708 \,.$$

EXAMPLE 2. Define the continuous random variable X as a random real number in the interval $[0, 1]$. X then has the density $f(x) = 1$ on $[0, 1]$. The expected value is

$$E(X) = \int_0^1 x\,dx = 0.5 \,.$$

The variance is

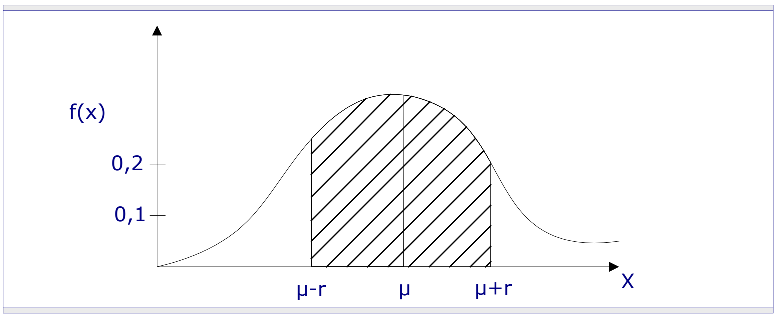
$$\text{var}(X) = \int_0^1 (x-0.5)^2\,dx = 0.083 \,.$$

The standard deviation is

$$\sigma = \sqrt{0.083} = 0.289 \,.$$

4.4 Estimation of expected value μ and standard deviation σ by eye

If the density function (or a pin diagram showing the point probabilities) of a random variable is given, one can estimate μ and σ by eye. The expected value μ is approximately the "centre of mass" of the distribution, and the standard deviation σ has a size such that more or less two thirds of the "probability mass" lie in the interval $\mu \pm \sigma$.



4.5 Addition and multiplication formulae for expected value and variance

Let X and Y be random variables. Then one has the formulae

$$\begin{aligned} E(X+Y) &= E(X)+E(Y) \\ E(aX) &= a\cdot E(X) \\ \text{var}(X) &= E(X^2)-E(X)^2 \\ \text{var}(aX) &= a^2\cdot \text{var}(X) \\ \text{var}(X+a) &= \text{var}(X) \end{aligned}$$

for every $a \in \mathbb{R}$. If X and Y are **independent**, one has moreover

$$\begin{aligned} E(X\cdot Y) &= E(X)\cdot E(Y) \\ \text{var}(X+Y) &= \text{var}(X)+\text{var}(Y) \end{aligned}$$

Remember: The expected value is additive. For independent random variables, the expected value is multiplicative and the variance is additive.

4.6 Covariance and correlation coefficient

The **covariance** of two random variables X and Y is the number

$$\text{Cov}(X,Y) = E((X-EX)(Y-EY))\ .$$

One has

$$\begin{aligned} \text{Cov}(X,X) &= \text{var}(X) \\ \text{Cov}(X,Y) &= E(X\cdot Y) - EX\cdot EY \\ \text{var}(X+Y) &= \text{var}(X)+\text{var}(Y)+2\cdot\text{Cov}(X,Y) \end{aligned}$$

The **correlation coefficient** ρ (“rho”) of X and Y is the number

$$\rho = \frac{\text{Cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}\ ,$$

where $\sigma(X) = \sqrt{\text{var}(X)}$ and $\sigma(Y) = \sqrt{\text{var}(Y)}$ are the standard deviations of X and Y. It is here assumed that neither standard deviation is zero. The correlation coefficient is a number in the interval $[-1,1]$. If X and Y are independent, both the covariance and ρ equal zero.

Remember: A positive correlation coefficient implies that normally X is large when Y large, and vice versa. A negative correlation coefficient implies that normally X is small when Y is large, and vice versa.

EXAMPLE. A red and a black dice are thrown. Consider the random variables

X: number of pips of red dice,
Y: number of pips of red and black dice in total.

If X is large, Y will normally be large too, and vice versa. We therefore expect a positive correlation coefficient. More precisely, we compute

$$\begin{aligned} E(X) &= 3.5 \\ E(Y) &= 7 \\ E(X \cdot Y) &= 27.42 \\ \sigma(X) &= 1.71 \\ \sigma(Y) &= 2.42 \end{aligned}$$

The covariance thus becomes

$$\text{Cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 27.42 - 3.5 \cdot 7 = 2.92 \; .$$

As expected, the correlation coefficient is a positive number:

$$\rho = \frac{\text{Cov}(X,Y)}{\sigma(X) \cdot \sigma(Y)} = \frac{2.92}{1.71 \cdot 2.42} = 0.71 \; .$$

5 The Law of Large Numbers

5.1 Chebyshev's Inequality

For a random variable X with expected value μ and variance σ^2, we have **Chebyshev's Inequality**:

$$P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

for every $a > 0$.

5.2 The Law of Large Numbers

Consider a sequence $X_1, X_2, X_3, \dots$ of independent random variables with the same distribution and let μ be the common expected value. Denote by S_n the sums

$$S_n = X_1 + \cdots + X_n \ .$$

The Law of Large Numbers then states that

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \to 0 \ \text{ for } n \to \infty$$

for every $\varepsilon > 0$. Expressed in words:

The mean value of a sample from any given distribution converges to the expected value of that distribution when the size n of the sample approaches ∞.

5.3 The Central Limit Theorem

Consider a sequence $X_1, X_2, X_3, \dots$ of independent random variables with the same distribution. Let μ be the common expected value and σ^2 the common variance. It is assumed that σ^2 is positive. Denote by S'_n the normed sums

$$S'_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \ .$$

By "normed" we understand that the S'_n have expected value 0 and variance 1. The **Central Limit Theorem** now states that

$$P(S'_n \leq x) \to \Phi(x) \ \text{ for } n \to \infty$$

for all $x \in \mathbb{R}$, where Φ is the distribution function of the standard normal distribution (see section 15.4):

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \ .$$

The distribution function of the normed sums S'_n thus converges to Φ when n converges to ∞.

This is a quite amazing result and the absolute climax of probability theory! The surprising thing is that the limit distribution of the normed sums is independent of the distribution of the X_i.

5.4 Example (distribution functions converge to Φ)

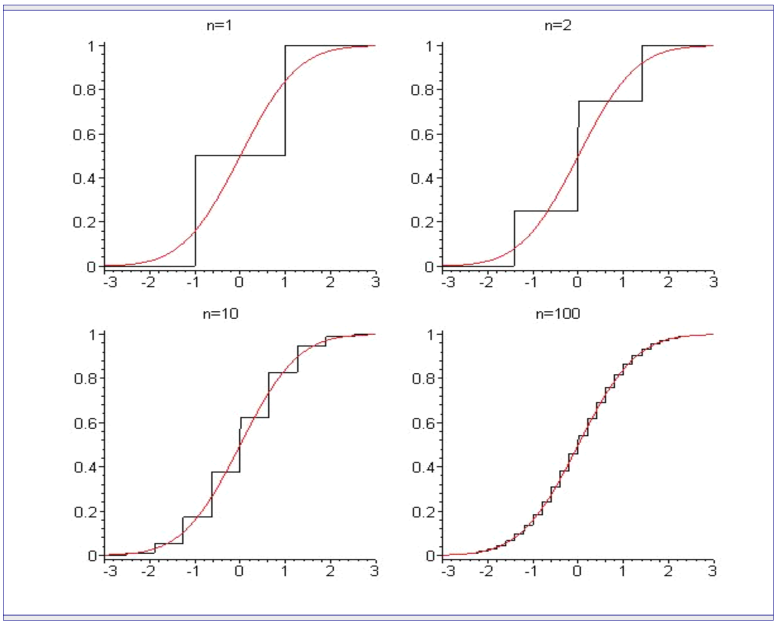
Consider a sequence of independent random variables $X_1, X_2, \dots$ all having the same point probabilities

$$P(X_i = 0) = P(X_i = 1) = \frac{1}{2}.$$

The sums $S_n = X_1 + \cdots + X_n$ are binomially distributed with expected value $\mu = n/2$ and variance $\sigma^2 = n/4$. The normed sums thus become

$$S'_n = \frac{X_1 + \cdots + X_n - \mu/2}{\sqrt{n}/2} \ .$$

The distribution of the S'_n is given by the distribution function F_n. The Central Limit Theorem states that F_n converges to Φ for $n \to \infty$. The figure below shows F_n together with Φ for $n = 1, 2, 10, 100$. It is a moment of extraordinary beauty when one watches the F_n slowly approaching Φ:



6 Descriptive statistics

6.1 Median and quartiles

Suppose we have n observations $x_1, \dots, x_n$. We then define the **median** $x(0.5)$ of the observations as the "middle observation". More precisely,

$$x(0.5) = \begin{cases} x_{(n+1)/2} & \text{if } n \text{ is odd} \\ (x_{n/2} + x_{n/2+1})/2 & \text{if } n \text{ is even} \end{cases}$$

where the observations have been sorted according to size as

$$x_1 \leq x_2 \leq \cdots \leq x_n \ .$$

Similarly, the **lower quartile** $x(0.25)$ is defined such that 25% of the observations lie below $x(0.25)$, and the **upper quartile** $x(0.75)$ is defined such that 75% of the observations lie below $x(0.75)$.

The **interquartile range** is the distance between $x(0.25)$ and $x(0.75)$, i.e. $x(0.75) - x(0.25)$.

6.2 Mean value

Suppose we have n observations $x_1, \ldots, x_n$. We define the **mean** or **mean value** of the observations as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

6.3 Empirical variance and empirical standard deviation

Suppose we have n observations $x_1, \ldots, x_n$. We define the **empirical variance** of the observations as

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \, .$$

The **empirical standard deviation** is the square root of the empirical variance:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \ .$$

The greater the empirical standard deviation s is, the more "dispersed" the observations are around the mean value $\bar{x}$.

6.4 Empirical covariance and empirical correlation coefficient

Suppose we have n pairs of observations $(x_1, y_1), \ldots, (x_n, y_n)$. We define the **empirical covariance** of these pairs as

$$\mathrm{Cov}_{\mathrm{emp}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \ .$$

Alternatively, $\mathrm{Cov}_{\mathrm{emp}}$ can be computed as

$$\mathrm{Cov}_{\mathrm{emp}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{n-1} \ .$$

The **empirical correlation coefficient** is

$$r = \frac{\text{empirical covariance}}{(\text{empirical standard deviation of the } x)(\text{empirical standard deviation of the } y)} = \frac{\mathrm{Cov}_{\mathrm{emp}}}{s_x s_y} \ .$$

The empirical correlation coefficient r always lies in the interval $[-1, 1]$.

Understanding of the empirical correlation coefficient. If the x-observations are independent of the y-observations, then r will be equal or close to 0. If the x-observations and the y-observations are dependent in such a way that large x-values usually correspond to large y-values, and vice versa, then r will be equal or close to 1. If the x-observations and the y-observations are dependent in such a way that large x-values usually correspond to small y-values, and vice versa, then r will be equal or close to –1.

7 Statistical hypothesis testing

7.1 Null hypothesis and alternative hypothesis

A **statistical test** is a procedure that leads to either **acceptance** or **rejection** of a **null hypothesis** $\mathbf{H}_0$ given in advance. Sometimes $\mathbf{H}_0$ is tested against an explicit **alternative hypothesis** $\mathbf{H}_1$.

At the base of the test lie one or more **observations**. The null hypothesis (and the alternative hypothesis, if any) concern the question which distribution these observations were taken from.

7.2 Significance probability and significance level

One computes the **significance probability** P, that is *the probability* – if $\mathbf{H}_0$ is true – *of obtaining an observation which is as extreme, or more extreme, than the one given.* The smaller P is, the less plausible $\mathbf{H}_0$ is.

Often, one chooses a **significance level** α in advance, typically $\alpha = 5\%$. One then rejects $\mathbf{H}_0$ if P is smaller than α (and one says, "$\mathbf{H}_0$ is rejected at significance level α"). If P is greater than α, then $\mathbf{H}_0$ is accepted (and one says, "$\mathbf{H}_0$ is accepted at significance level α" or "$\mathbf{H}_0$ cannot be rejected at significance level α").

7.3 Errors of type I and II

We speak about a **type I error** if we reject a true null hypothesis. If the significance level is α, then the risk of a type I error is at most α.

We speak about a **type II error** if we accept a false null hypothesis.

The **strength** of a test is the probability of rejecting a false $\mathbf{H}_0$. The greater the strength, the smaller the risk of a type II error. Thus, the strength should be as great as possible.

7.4 Example

Suppose we wish to investigate whether a certain dice is fair. By "fair" we here only understand that the probability p of a six is 1/6. We test the null hypothesis

$$\mathbf{H}_0 : p = \frac{1}{6} \text{ (the dice is fair)}$$

against the alternative hypothesis

$$\mathbf{H}_1 : p > \frac{1}{6} \text{ (the dice is biased)}$$

The observations on which the test is carried out are the following ten throws of the dice:

$$2, 6, 3, 6, 5, 2, 6, 6, 4, 6\ .$$

Let us in advance agree upon a significance level $\alpha = 5\%$. Now the significance probability P can be computed. By "extreme observations" is understood that there are many sixes. Thus, P is the probability of having at least five sixes in 10 throws with a fair dice. We compute

$$P = \sum_{k=5}^{10} \binom{10}{k} (1/6)^k (5/6)^{10-k} = 0.015$$

(see section 8 on the binomial distribution). Since $P = 1.5\%$ is smaller than $\alpha = 5\%$, we reject $\mathbf{H}_0$. If the same test was performed with a fair dice, the probability of committing a type I error would be 1.5%.

8 The binomial distribution Bin(n, p)

8.1 Parameters

n: number of tries
p: probability of success

In the formulae we also use the "probability of failure" $q = 1 - p$.

8.2 Description

We carry out n independent tries that each result in either success or failure. In each try the probability of success is the same, p. Consequently, the total number of successes X is binomially distributed, and we write $X \sim \text{Bin}(n, p)$. X is a discrete random variable and takes values in the set $\{0, 1, \ldots, n\}$.

8.3 Point probabilities

For $k \in \{0, 1, \ldots, n\}$, the point probabilities in a $\text{Bin}(n, p)$ distribution are

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \, .$$

See section 2.5 regarding the **binomial coefficients** $\left(\begin{array}{c} n \\ k \end{array}\right)$.

EXAMPLE. If a dice is thrown twenty times, the total number of sixes, X, will be binomially distributed with parameters $n = 20$ and $p = 1/6$. We can list the point probabilities $P(X = k)$

and the **cumulative probabilities** $P(X \geq k)$ in a table (expressed as percentages):

k	0	1	2	3	4	5	6	7	8	9
$P(X = k)$	2.6	10.4	19.8	23.8	20.2	12.9	6.5	2.6	0.8	0.2
$P(X \geq k)$	100	97.4	87.0	67.1	43.3	23.1	10.2	3.7	1.1	0.3

8.4 Expected value and variance

Expected value: $E(X) = np$.

Variance: $\text{var}(X) = npq$.

8.5 Significance probabilities for tests in the binomial distribution

We perform n independent experiments with the same probability of success p and count the number k of successes. We wish to test the null hypothesis $\mathbf{H}_0 : p = p_0$ against an alternative hypothesis $\mathbf{H}_1$.

$\mathbf{H}_0$	$\mathbf{H}_1$	Significance probability
$p = p_0$	$p > p_0$	$P(X \geq k)$
$p = p_0$	$p < p_0$	$P(X \leq k)$
$p = p_0$	$p \neq p_0$	$\sum_l P(X = l)$

where in the last line we sum over all l for which $P(X = l) \leq P(X = k)$.

EXAMPLE. A company buys a machine that produces microchips. The manufacturer of the machine claims that at most one sixth of the produced chips will be defective. The first day the machine produces 20 chips of which 6 are defective. Can the company reject the manufacturer's claim on this background?

SOLUTION. We test the null hypothesis $\mathbf{H}_0 : p = 1/6$ against the alternative hypothesis $\mathbf{H}_1 : p > 1/6$. The significance probability can be computed as $P(X \geq 6) = 10.2\%$ (see e.g. the table in section 8.3). We conclude that the company *cannot* reject the manufacturer's claim at the 5% level.

8.6 The normal approximation to the binomial distribution

If the parameter n (the number of tries) is large, a binomially distributed random variable X will be approximately normally distributed with expected value $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$. Therefore, the point probabilities are approximately

$$P(X = k) \approx \varphi\left(\frac{k - np}{\sqrt{npq}}\right) \cdot \frac{1}{\sqrt{npq}}$$

where φ is the density of the standard normal distribution, and the tail probabilities are approximately

$$P(X \leq k) \approx \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{npq}}\right)$$

$$P(X \geq k) \approx 1 - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{npq}}\right)$$

where Φ is the distribution function of the standard normal distribution (Table B.2).

Rule of thumb. One may use the normal approximation if np and nq are both greater than 5.

EXAMPLE (continuation of the example in section 8.5). After 2 weeks the machine has produced 200 chips of which 46 are defective. Can the company now reject the manufacturer's claim that the probability of defects is at most one sixth?

SOLUTION. Again we test the null hypothesis $H_0 : p = 1/6$ against the alternative hypothesis $H_1 : p > 1/6$. Since now $np \approx 33$ and $nq \approx 167$ are both greater than 5, we may use the normal approximation in order to compute the significance probability:

$$P(X \geq 46) \approx 1 - \Phi\left(\frac{46 - \frac{1}{2} - 33.3}{\sqrt{27.8}}\right) \approx 1 - \Phi(2.3) \approx 1.1\%$$

Therefore, the company may now *reject* the manufacturer's claim at the 5% level.

8.7 Estimators

Suppose k is an observation from a random variable $X \sim \text{Bin}(n, p)$ with known n and unknown p. The **maximum likelihood estimate (ML estimate)** of p is

$$\hat{p} = \frac{k}{n} .$$

This estimator is **unbiased** (i.e. the expected value of the estimator is p) and has **variance**

$$\text{var}(\hat{p}) = \frac{pq}{n} .$$

The expression for the variance is of no great practical value since it depends on the true (unknown) probability parameter p. If, however, one plugs in the estimated value $\hat{p}$ in place of p, one gets the **estimated variance**

$$\frac{\hat{p}(1-\hat{p})}{n} .$$

EXAMPLE. We consider again the example with the machine that has produced twenty microchips of which the six are defective. What is the maximum likelihood estimate of the probability parameter? What is the estimated variance?

SOLUTION. The maximum likelihood estimate is

$$\hat{p} = \frac{6}{20} = 30\%$$

and the variance of $\hat{p}$ is estimated as

$$\frac{0.3 \cdot (1 - 0.3)}{20} = 0.0105 .$$

The standard deviation is thus estimated to be $\sqrt{0.0105} \approx 0.10$. If we presume that $\hat{p}$ lies within two standard deviations from p, we may conclude that p is between 10% and 50%.

8.8 Confidence intervals

Suppose k is an observation from a binomially distributed random variable $X \sim \text{Bin}(n, p)$ with known n and unknown p. The confidence interval with confidence level $1 - \alpha$ around the point estimate $\hat{p} = k/n$ is

$$\left[\hat{p} - u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ , \ \hat{p} + u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] .$$

Loosely speaking, the true value p lies in the confidence interval with the probability $1 - \alpha$.

The number $u_{1-\alpha/2}$ is determined by $\Phi(u_{1-\alpha/2}) = 1 - \alpha/2$ where Φ is the distribution function of the standard normal distribution. It appears e.g. from Table B.2 that with confidence level 95% one has

$$u_{1-\alpha/2} = u_{0.975} = 1.96 .$$

EXERCISE. In an opinion poll from the year 2015, 62 out of 100 persons answer that they intend to vote for the Green Party at the next election. Compute the confidence interval with confidence level 95% around the true percentage of Green Party voters.

SOLUTION. The point estimate is $\hat{p} = 62/100 = 0.62$. A confidence level of 95% yields $\alpha = 0.05$. Looking up in the table (see above) gives $u_{0.975} = 1.96$. We get

$$1.96\sqrt{\frac{0.62 \cdot 0.38}{100}} = 0.10 \ .$$

The confidence interval thus becomes

$$[0.52\ ,\ 0.72]\ .$$

So we can say with a certainty of 95% that between 52% and 72% of the electorate will vote for the Green Party at the next election.

9 The Poisson distribution Pois(λ)

9.1 Parameters

λ: Intensity

9.2 Description

Certain events are said to occur *spontaneously*, i.e. they occur at random times, independently of each other, but with a certain constant *intensity* λ. The intensity is the average number of spontaneous events per time interval. The number of spontaneous events X in any given concrete time interval is then Poisson distributed, and we write $X \sim \text{Pois}(\lambda)$. X is a discrete random variable and takes values in the set $\{0, 1, 2, 3, \dots\}$.

9.3 Point probabilities

For $k \in \{0, 1, 2, 3 \dots\}$ the point probabilities in a Pois(λ) distribution are

$$P(X = k) = \frac{\lambda^k}{k!}\exp(-\lambda)\ .$$

Recall the convention $0! = 1$.

EXAMPLE. In a certain shop an average of three customers per minute enter. The number of customers X entering during any particular minute is then Poisson distributed with intensity $\lambda = 3$. The point probabilities (as percentages) can be listed in a table as follows:

k	0	1	2	3	4	5	6	7	8	9	≥ 10
$P(X = k)$	5.0	14.9	22.4	22.4	16.8	10.1	5.0	2.2	0.8	0.3	0.1

9.4 Expected value and variance

Expected value: $E(X) = \lambda$.

Variance: $\text{var}(X) = \lambda$.

9.5 Addition formula

Suppose that $X_1, \ldots, X_n$ are independent Poisson distributed random variables. Let λ_i be the intensity of X_i, i.e. $X_i \sim \text{Pois}(\lambda_i)$. Then the sum

$$X = X_1 + \cdots + X_n$$

will be Poisson distributed with intensity

$$\lambda = \lambda_1 + \cdots + \lambda_n \ ,$$

i.e. $X \sim \text{Pois}(\lambda)$.

9.6 Significance probabilities for tests in the Poisson distribution

Suppose that k is an observation from a Pois(λ) distribution with unknown intensity λ. We wish to test the null hypothesis $\mathbf{H}_0 : \lambda = \lambda_0$ against an alternative hypothesis $\mathbf{H}_1$.

$\mathbf{H}_0$	$\mathbf{H}_1$	Significance probability
$\lambda = \lambda_0$	$\lambda > \lambda_0$	$P(X \geq k)$
$\lambda = \lambda_0$	$\lambda < \lambda_0$	$P(X \leq k)$
$\lambda = \lambda_0$	$\lambda \neq \lambda_0$	$\sum_l P(X = l)$

where the summation in the last line is over all l for which $P(X = l) \leq P(X = k)$.

If n independent observations $k_1, \ldots, k_n$ from a Pois(λ) distribution are given, we can treat the sum $k = k_1 + \cdots + k_n$ as an observation from a Pois($n \cdot \lambda$) distribution.

9.7 Example (significant increase in sale of Skodas)

EXERCISE. A Skoda car salesman sells on average 3.5 cars per month. The month after a radio campaign for Skoda, seven cars are sold. Is this a significant increase?

SOLUTION. The sale of cars in the given month may be assumed to be Poisson distributed with a certain intensity λ. We test the null hypothesis

$$\mathbf{H}_0 : \lambda = 3.5$$

against the alternative hypothesis

$$\mathbf{H}_1 : \lambda > 3.5 \ .$$

The significance probability, i.e. the probability of selling at least seven cars given that $\mathbf{H}_0$ is true, is

$$P = \sum_{k=7}^{\infty} \frac{(3.5)^k}{k!} \exp(-3.5) = 0.039 + 0.017 + 0.007 + 0.002 + \cdots = 0.065 \ .$$

Since P is greater than 5%, we cannot reject $\mathbf{H}_0$. In other words, the increase is not significant.

9.8 The binomial approximation to the Poisson distribution

The Poisson distribution with intensity λ is the limit distribution of the binomial distribution with parameters n and $p = \lambda/n$ when n tends to ∞. In other words, the point probabilities satisfy

$$P(X_n = k) \to P(X = k) \quad \text{for} \quad n \to \infty$$

for $X \sim \text{Pois}(\lambda)$ and $X_n \sim \text{Bin}(n, \lambda/n)$. In real life, however, one almost always prefers to use the normal approximation instead (see the next section).

9.9 The normal approximation to the Poisson distribution

If the intensity λ is large, a Poisson distributed random variable X will to a good approximation be normally distributed with expected value $\mu = \lambda$ and standard deviation $\sigma = \sqrt{\lambda}$. The point probabilities therefore are

$$P(X = k) \approx \varphi\left(\frac{k-\lambda}{\sqrt{\lambda}}\right) \cdot \frac{1}{\sqrt{\lambda}}$$

where $\varphi(x)$ is the density of the standard normal distribution, and the tail probabilities are

$$P(X \leq k) \approx \Phi\left(\frac{k + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$$

$$P(X \geq k) \approx 1 - \Phi\left(\frac{k - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$$

where Φ is the distribution function of the standard normal distribution (Table B.2).

Rule of thumb. The normal approximation to the Poisson distribution applies if λ is greater than nine.

9.10 Example (significant decrease in number of complaints)

EXERCISE. The ferry *Deutschland* between Rødby and Puttgarten receives an average of 180 complaints per week. In the week immediately after the ferry's cafeteria was closed, only 112 complaints are received. Is this a significant decrease?

SOLUTION. The number of complaints within the given week may be assumed to be Poisson distributed with a certain intensity λ. We test the null hypothesis

$$\mathbf{H}_0 : \lambda = 180$$

against the alternative hypothesis

$$\mathbf{H}_1 : \lambda < 180 \; .$$

The significance probability, i.e. the probability of having at most 112 complaints given $\mathbf{H}_0$, can be approximated by the normal distribution:

$$P = \Phi\left(\frac{112 + \frac{1}{2} - 180}{\sqrt{180}}\right) = \Phi(-5.03) < 0.0001 \; .$$

Since P is very small, we can clearly reject $\mathbf{H}_0$. The number of complaints has significantly decreased.

9.11 Estimators

Suppose $k_1, \ldots k_n$ are independent observations from a random variable $X \sim \text{Pois}(\lambda)$ with unknown intensity λ. The **maximum likelihood estimate (ML estimate)** of λ is

$$\hat{\lambda} = (k_1 + \cdots + k_n)/n \; .$$

This estimator is unbiased (i.e. the expected value of the estimator is λ) and has **variance**

$$\text{var}(\hat{\lambda}) = \frac{\lambda}{n} \; .$$

More precisely, we have

$$n\hat{\lambda} \sim \text{Pois}(n\lambda) \; .$$

If we plug in the estimated value $\hat{\lambda}$ in λ's place, we get the **estimated variance**

$$\widehat{\text{var}}(\hat{\lambda}) = \frac{\hat{\lambda}}{n} \; .$$

9.12 Confidence intervals

Suppose $k_1, \dots, k_n$ are independent observations from a Poisson distributed random variable $X \sim$ Pois(λ) with unknown λ. The confidence interval with confidence level $1-\alpha$ around the point estimate $\hat{\lambda} = (k_1 + \cdots + k_n)/n$ is

$$\left[\hat{\lambda} - u_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \ , \ \hat{\lambda} + u_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} \ \right] .$$

Loosely speaking, the true value λ lies in the confidence interval with probability $1-\alpha$.

The number $u_{1-\alpha/2}$ is determined by $\Phi(u_{1-\alpha/2}) = 1 - \alpha/2$, where Φ is the distribution function of the standard normal distribution. It appears from, say, Table B.2 that

$$u_{1-\alpha/2} = u_{0.975} = 1.96$$

for confidence level 95%.

EXAMPLE (continuation of the example in section 9.10). In the first week after the closure of the ferry's cafeteria, a total of 112 complaints were received. We consider $k = 112$ as an observation from a Pois(λ) distribution and wish to find the confidence interval with confidence level 95% around the estimate

$$\hat{\lambda} = 112 .$$

Looking up in the table gives $u_{0.975} = 1.96$. The confidence interval thus becomes

$$\left[112 - 1.96\sqrt{112} \ , \ 112 + 1.96\sqrt{112} \ \right] \approx [91 \ , \ 133]$$

10 The geometrical distribution Geo(p)

10.1 Parameters

p: probability of success

In the formulae we also use the "probability of failure" $q = 1 - p$.

10.2 Description

A series of experiments are carried out, each of which results in either success or failure. The probability of success p is the same in each experiment. The number W of failures before the first success is then geometrically distributed, and we write $W \sim$ Geo(p). W is a discrete random variable and takes values in the set $\{0, 1, 2, \dots\}$. The "wait until success" is $V = W + 1$.

10.3 Point probabilities and tail probabilities

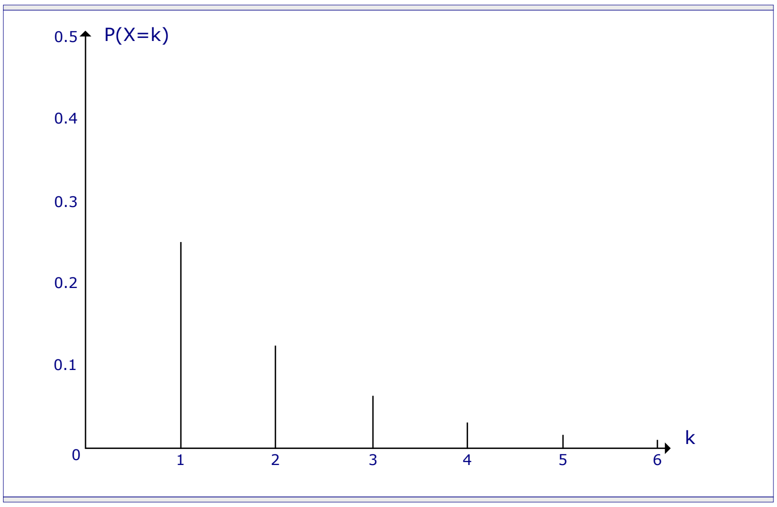
For $k \in \{0, 1, 2 \dots\}$ the point probabilities in a Geo(p) distribution are

$$P(X = k) = q^k p .$$

In contrast to most other distributions, we can easily compute the tail probabilities in the geometrical distribution:

$$P(X \geq k) = q^k \, .$$

EXAMPLE. Pin diagram for the point probabilities in a geometrical distribution with probability of success $p = 0.5$:



10.4 Expected value and variance

Expected value: $E(W) = q/p$.

Variance: $\text{var}(W) = q/p^2$.

Regarding the "wait until success" $V = W + 1$, we have the following useful rule:

Rule. The expected wait until success is the reciprocal probability of success: $E(V) = 1/p$.

EXAMPLE. A gambler plays lotto each week. The probability of winning in lotto, i.e. the probability of guessing correctly seven numbers picked randomly from a pool of 36 numbers, is

$$p = \binom{36}{7}^{-1} \approx 0.0000001198\ .$$

The expected wait until success is thus

$$E(V) = p^{-1} = \binom{36}{7}\ \text{weeks}\ = 160532\ \text{years}\ .$$

11 The hypergeometrical distribution HG(n, r, N)

11.1 Parameters

r: number of red balls

s: number of black balls

N: total number of balls ($N = r + s$)

n: number of balls picked out ($n \leq N$)

11.2 Description

In an urn we have r red balls and s black balls, in total $N = r + s$ balls. We now pick out n balls from the urn, randomly and without returning the chosen balls to the urn. Necessarily $n \leq N$. The number of red balls Y amongst the balls chosen is then hypergeometrically distributed and we write $Y \sim$ HG(n, r, N). Y is a discrete random variable with values in the set $\{0, 1, \ldots, \min\{n, r\}\}$.

11.3 Point probabilities and tail probabilities

For $k \in \{0, 1, \ldots, \min\{n, r\}\}$ the point probabilities of a HG(n, r, N) distribution are

$$P(Y = k) = \frac{\binom{r}{k} \cdot \binom{s}{n-k}}{\binom{N}{n}}\ .$$

EXAMPLE. A city council has 25 members of which 13 are Conservatives. A cabinet is formed by picking five council members at random. What is the probability that the Conservatives will have absolute majority in the cabinet?

SOLUTION. We have here a hypergeometrically distributed random variable $Y \sim \text{HG}(5, 13, 25)$ and have to compute $P(Y \geq 3)$. Let us first compute all point probabilities (as percentages):

k	0	1	2	3	4	5
$P(Y = k)$	1.5	12.1	32.3	35.5	16.1	2.4

The sought-after probability thus becomes

$$P(Y \geq 3) = 35.5\% + 16.1\% + 2.4\% = 54.0\%$$

11.4 Expected value and variance

Expected value: $E(Y) = nr/N$.

Variance: $\text{var}(Y) = nrs(N-n)/(N^2(N-1))$.

11.5 The binomial approximation to the hypergeometrical distribution

If the number of balls picked out, n, is small compared to both the number of red balls r and the number of black balls s, it becomes irrelevant whether the balls picked out are returned to the urn or not. We can thus approximate the hypergeometrical distribution by the binomial distribution:

$$P(Y = k) \approx P(X = k)$$

for $Y \sim \text{HG}(n, r, N)$ and $X \sim \text{Bin}(n, r/N)$. In practice, this approximation is of little value since it is as difficult to compute $P(X = k)$ as $P(Y = k)$.

11.6 The normal approximation to the hypergeometrical distribution

If n is small compared to both r and s, the hypergeometrical distribution can be approximated by the normal distribution with the same expected value and variance.

The point probabilities thus become

$$P(Y = k) \approx \varphi\left(\frac{k - nr/N}{\sqrt{nrs(N-n)/(N^2(N-1))}}\right) \cdot \frac{1}{\sqrt{nrs(N-n)/(N^2(N-1))}}$$

where φ is the density of the standard normal distribution. The tail probabilities become

$$P(Y \leq k) \approx \Phi\left(\frac{k + \frac{1}{2} - nr/N}{\sqrt{nrs(N-n)/(N^2(N-1))}}\right)$$

$$P(Y \geq k) \approx 1 - \Phi\left(\frac{k - \frac{1}{2} - nr/N}{\sqrt{nrs(N-n)/(N^2(N-1))}}\right)$$

where Φ is the distribution function of the standard normal distribution (Table B.2).

12 The multinomial distribution Mult$(n, p_1, \ldots, p_r)$

12.1 Parameters

n: number of tries

p_1: 1st probability parameter

$\vdots$

p_r: rth probability parameter

It is required that $p_1 + \cdots + p_r = 1$.

12.2 Description

We carry out n independent experiments each of which results in one out of r possible outcomes. The probability of obtaining an outcome of type i is the same in each experiment, namely p_i. Let

S_i denote the total number of outcomes of type i. The random vector $S = (S_1, \ldots, S_r)$ is then multinomially distributed and we write $S \sim \text{Mult}(n, p_1, \ldots, p_r)$. S is discrete and takes values in the set $\{(k_1, \ldots k_r) \in \mathbb{Z}^r \mid k_i \geq 0\ ,\ k_1 + \cdots + k_r = n\}$.

12.3 Point probabilities

For $k_1 + \cdots + k_r = n$ the point probabilities of a $\text{Mult}(n, p_1, \ldots, p_r)$ distribution are

$$P(S = (k_1, \ldots, k_r)) = \begin{pmatrix} n \\ k_1 \ \cdots \ k_r \end{pmatrix} \cdot \prod_{i=1}^{r} p_i^{k_i}$$

EXAMPLE. Throw a dice six times and, for each i, let S_i be the total number of i's. Then $S = (S_1, \ldots, S_6)$ is a multinomially distributed random vector: $S \sim \text{Mult}(6, 1/6, \ldots, 1/6)$. The probability of obtaining, say, exactly one 1, two 2s, and three sixes is

$$P(S = (1,2,0,0,0,3)) = \begin{pmatrix} 6 \\ 1\ 2\ 0\ 0\ 0\ 3 \end{pmatrix} \cdot (1/6)^1 \cdot (1/6)^2 \cdot (1/6)^3 \approx 0.13\%$$

Here, the **multinomial coefficient** (see also section 2.6) is computed as

$$\begin{pmatrix} 6 \\ 1\ 2\ 0\ 0\ 0\ 3 \end{pmatrix} = \frac{6!}{1!2!0!0!0!3!} = \frac{720}{12} = 60\ .$$

12.4 Estimators

Suppose $k_1, \ldots, k_r$ is an observation from a random vector $S \sim \text{Mult}(n, p_1, \ldots, p_r)$ with known n and unknown p_i. The **maximum likelihood estimate** of p_i is

$$\hat{p}_i = \frac{k_i}{n}\ .$$

This estimator is unbiased (i.e. the estimator's expected value is p_i) and has variance

$$\text{var}(\hat{p}_i) = \frac{p_i(1-p_i)}{n}\ .$$

13 The negative binomial distribution NB(n, p)

13.1 Parameters

n: number of tries

p: probability of success

In the formulae we also use the letter $q = 1 - p$.

13.2 Description

A series of independent experiments are carried out each of which results in either success or failure. The probability of success p is the same in each experiment. The total number X of failures before the n'th success is then negatively binomially distributed and we write $X \sim \text{NB}(n,p)$. The random variable X is discrete and takes values in the set $\{0,1,2,\dots\}$.

The geometrical distribution is the special case $n = 1$ of the negative binomial distribution.

13.3 Point probabilities

For $k \in \{0,1,2\dots\}$ the point probabilities of a NB(k,p) distribution are

$$P(X = k) = \binom{n+k-1}{n-1} \cdot p^n \cdot q^k \,.$$

13.4 Expected value and variance

Expected value: $E(X) = nq/p$.

Variance: $\text{var}(X) = nq/p^2$.

13.5 Estimators

The negative binomial distribution is sometimes used as an alternative to the Poisson distribution in situations where one wishes to describe a random variable taking values in the set $\{0,1,2,\dots\}$.

Suppose $k_1,\dots,k_m$ are independent observations from a NB(n,p) distribution with unknown parameters n and p. We then have the following estimators:

$$\hat{n} = \frac{\bar{k}^2}{s^2 - \bar{k}} \quad , \quad \hat{p} = \frac{\bar{k}}{s^2}$$

where $\bar{k}$ and s^2 are the mean value and empirical variance of the observations.

14 The exponential distribution Exp(λ)

14.1 Parameters

λ: Intensity

14.2 Description

In a situation where events occur spontaneously with the intensity λ (and where the number of spontaneous events in any given time interval thus is Pois(λ) distributed), the *wait* T between two spontaneous events is exponentially distributed and we write $T \sim \text{Exp}(\lambda)$. T is a continuous random variable taking values in the interval $[0,\infty[$.

14.3 Density and distribution function

The density of the exponential distribution is

$$f(x) = \lambda \cdot \exp(-\lambda x) \ .$$

The distribution function is

$$F(x) = 1 - \exp(-\lambda x) \ .$$

14.4 Expected value and variance

Expected value: $E(T) = 1/\lambda$.

Variance: $\text{var}(T) = 1/\lambda^2$.

15 The normal distribution

15.1 Parameters

μ: expected value

σ^2: variance

Remember that the standard deviation σ is the square root of the variance.

15.2 Description

The normal distribution is a continuous distribution. If a random variable X is normally distributed, then X can take any values in $\mathbb{R}$ and we write $X \sim N(\mu, \sigma^2)$.

The normal distribution is the most important distribution in all of statistics. Countless naturally occurring phenomena can be described (or approximated) by means of a normal distribution.

15.3 Density and distribution function

The density of the normal distribution is the function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) .$$

It is symmetric, i.e.

$$f(-x) = f(x) .$$

The distribution function of the normal distribution

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

is difficult to compute. Instead, one uses the formula

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

where Φ is the distribution function of the standard normal distribution which can be looked up in Table B.2. From the table the following fact appears:

Fact: In a normal distribution, about 68% of the probability mass lies within one standard deviation from the expected value, and about 95% of the probability mass lies within two standard deviations from the expected value.

15.4 The standard normal distribution

A normal distribution with expected value $\mu = 0$ and variance $\sigma^2 = 1$ is called a **standard normal distribution**. The standard deviation in a standard normal distribution equals 1 (obviously). The density $\varphi(t)$ of a standard normal distribution is

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) .$$

The distribution function Φ of a standard normal distribution is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt .$$

On can look up Φ in Table B.2.

15.5 Properties of Φ

The distribution function Φ of a standard normally distributed random variable $X \sim N(0,1)$ satisfies

$$\begin{aligned}
P(X \leq x) &= \Phi(x) \\
P(X \geq x) &= \Phi(-x) \\
P(|X| \leq x) &= \Phi(x) - \Phi(-x) \\
P(|X| \geq x) &= 2 \cdot \Phi(-x) \\
\Phi(-x) &= 1 - \Phi(x)
\end{aligned}$$

15.6 Estimation of the expected value μ

Suppose $x_1, x_2, \ldots, x_n$ are independent observations of a random variable $X \sim N(\mu, \sigma^2)$. The **maximum likelihood estimate (ML estimate)** of μ is

$$\hat{\mu} = \frac{x_1 + \cdots + x_n}{n} .$$

This is simply the **mean value** and is written $\bar{x}$. The mean value is an unbiased estimator of μ (i.e. the estimator's expected value is μ). The **variance** of the mean value is

$$\text{var}^2(\bar{x}) = \frac{\sigma^2}{n} .$$

More precisely, $\bar{x}$ is itself normally distributed:

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}) .$$

15.7 Estimation of the variance σ^2

Suppose $x_1, \ldots, x_n$ are independent observations of a random variable $X \sim N(\mu, \sigma^2)$. Normally, the variance σ^2 is estimated by the **empirical variance**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} .$$

The empirical variance s^2 is an unbiased estimator of the true variance σ^2.

Warning: The empirical variance is *not* the maximum likelihood estimate of σ^2. The maximum likelihood estimate of σ^2 is

$$\frac{\sum (x_i - \bar{x})^2}{n}$$

but this is seldom used since it is biased and usually gives estimates which are too small.

15.8 Confidence intervals for the expected value μ

Suppose $x_1, \ldots, x_n$ are independent observations of a normally distributed random variable $X \sim N(\mu, \sigma^2)$ and that we wish to estimate the expected value μ. If the variance σ^2 is known, the confidence interval for μ with confidence level $1-\alpha$ is as follows:

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

The number $u_{1-\alpha/2}$ is determined by $\Phi(u_{1-\alpha/2}) = 1-\alpha/2$ where Φ is the distribution function of the standard normal distribution. It appears from, say, Table B.2 that

$$u_{1-\alpha/2} = u_{0.975} = 1.96$$

for confidence level 95%.

If the variance σ^2 is unknown, the confidence interval for μ with confidence level $1-\alpha$ is

$$\left[\ \bar{x} - t_{1-\alpha/2}(n-1)\sqrt{\frac{s^2}{n}}\ ,\ \bar{x} + t_{1-\alpha/2}(n-1)\sqrt{\frac{s^2}{n}}\ \right]$$

where s^2 is the empirical variance (section 6.3). The number $t_{1-\alpha/2}$ is determined by $F(u_{1-\alpha/2}) = 1-\alpha/2$, where F is the distribution function of Student's t distribution with $n-1$ degrees of freedom. It appears from, say, Table B.4 that

n	2	3	4	5	6	7	8	9	10	11	12
$t_{1-\alpha/2}$	12.7	4.30	3.18	2.78	2.57	2.45	2.36	2.31	2.26	2.23	2.20

for confidence level 95%.

15.9 Confidence intervals for the variance σ^2 and the standard deviation σ

Suppose $x_1, \ldots, x_n$ are independent observations of a normally distributed random variable $X \sim N(\mu, \sigma^2)$. The confidence interval for the variance σ^2 with confidence level $1 - \alpha$ is:

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \; , \; \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right]$$

where s^2 is the empirical variance (section 6.3). The numbers $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are determined by $F(\chi^2_{\alpha/2}) = \alpha/2$ and $F(\chi^2_{1-\alpha/2}) = 1 - \alpha/2$ where F is the distribution function of the χ^2 distribution with $n-1$ degrees of freedom (Table B.3).

Confidence intervals for the standard deviation σ with confidence level $1 - \alpha$ are computed simply by taking the square root of the limits of the confidence intervals for the variance:

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \; , \; \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}} \right]$$

15.10 Addition formula

A linear function of a normally distributed random variable is itself normally distributed. If, in other words, $X \sim N(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$ ($a \neq 0$), then

$$aX + b \sim N(a\mu + b, a^2\sigma^2) \; .$$

The sum of independent normally distributed random variables is itself normally distributed. If, in other words, $X_1, \ldots, X_n$ are independent with $X_i \sim N(\mu_i, \sigma_i^2)$, then we have the *addition formula*

$$X_1 + \cdots + X_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2) \; .$$

16 Distributions connected with the normal distribution

16.1 The χ^2 distribution

Let $X_1, \ldots, X_n \sim N(0, 1)$ be independent standard normally distributed random variables. The distribution of the sum of squares

$$Q = X_1^2 + \cdots + X_n^2$$

is called the χ^2 **distribution** with n **degrees of freedom**. The number of degrees of freedom is commonly symbolized as *df*.

A χ^2 distributed random variable Q with df degrees of freedom has **expected value**

$$E(Q) = df$$

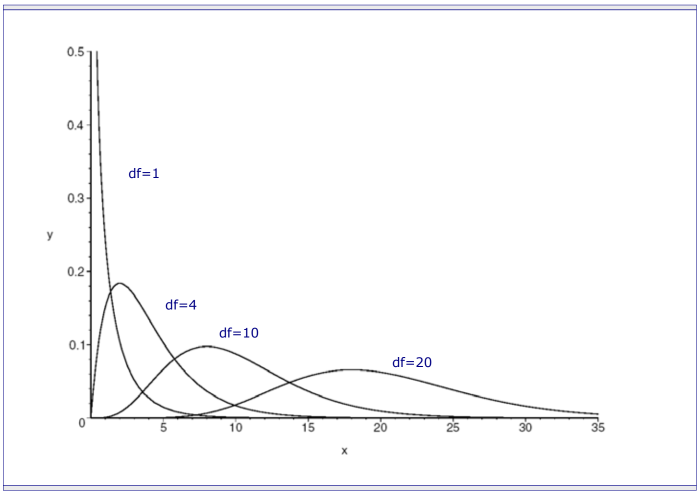
and **variance**

$$\text{var}(Q) = 2 \cdot df\ .$$

The **density** of the χ^2 distribution is

$$f(x) = K \cdot x^{\frac{df}{2}-1} \cdot e^{-\frac{x}{2}}$$

where df is the number of degrees of freedom and K is a constant. In practice, one doesn't use the density, but rather looks up the distribution function in Table B.3. The graph below shows the density function with $df = 1, 4, 10, 20$ degrees of freedom.



16.2 Student's t distribution

Let X be a normally distributed random variable with expected value μ and variance σ^2. Let the random variables $\bar{X}$ and S^2 be the mean value and empirical variance, respectively, of a sample consisting of n observations from X. The distribution of

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}}$$

is then independent of both μ and σ^2 and is called **Student's t distribution** with $n-1$ **degrees of freedom**.

A t distributed random variable T with df degrees of freedom has **expected value**

$$E(T) = 0$$

for $df \geq 2$, and **variance**

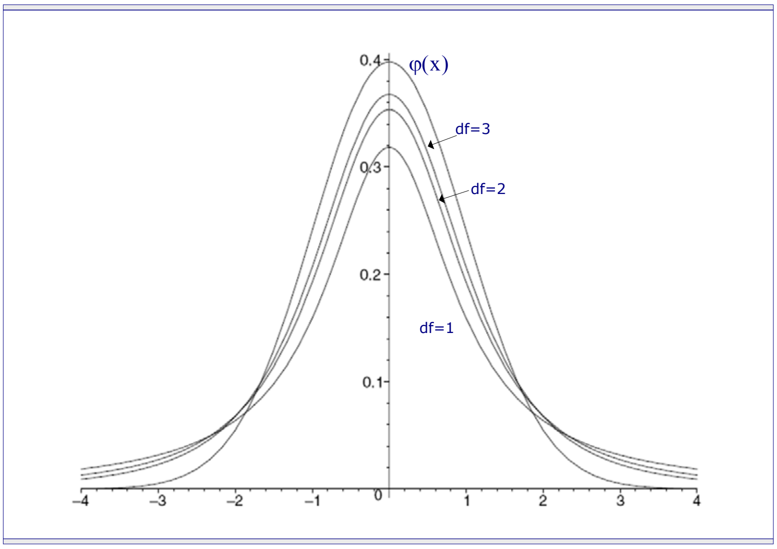
$$\text{var}(T) = \frac{df}{df - 2}$$

for $df \geq 3$.

The **density** of the t distribution is

$$f(x) = K \cdot \left(1 + \frac{x^2}{df}\right)^{-(df+1)/2}$$

where df is the number of degrees of freedom and K is a constant. In practice, one doesn't use the density, but rather looks up the distribution function in Table B.4. The graph below shows the density of the t distribution with $df = 1, 2, 3$ degrees of freedom and additionally the density $\varphi(x)$ of the standard normal distribution. As it appears, the t distribution approaches the standard normal distribution when $df \to \infty$.



16.3 Fisher's F distribution

Let X_1 and X_2 be independent normally distributed random variables with the same variance. For $i = 1, 2$ let the random variable S_i^2 be the empirical variance of a sample of size n_i from X_i. The distribution of the quotient

$$V = \frac{S_1^2}{S_2^2}$$

is called **Fisher's F distribution** with $n_1 - 1$ **degrees of freedom in the numerator** and $n_2 - 1$ **degrees of freedom in the denominator.**

The **density** of the F distribution is

$$f(x) = K \cdot \frac{x^{df_1/2-1}}{(df_2 + df_1 x)^{df/2}}$$

where K is a constant, df_1 the number of degrees of freedom in the numerator, df_2 the number of degrees of freedom in the denominator, and $df = df_1 + df_2$. In practice, one doesn't use the density, but rather looks up the distribution function in Table B.5.

17 Tests in the normal distribution

17.1 One sample, known variance, $\mathrm{H}_0 : \mu = \mu_0$

Let there be given a sample $x_1, \ldots, x_n$ of n independent observations from a normal distribution with **unknown expected value** μ and **known variance** σ^2. We wish to test the null hypothesis

$$\mathbf{H}_0 : \ \mu = \mu_0 \ .$$

For this purpose, we compute the **statistic**

$$u = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} = \frac{\sum_{i=1}^n x_i - n\mu_0}{\sqrt{n\sigma^2}} \ .$$

The **significance probability** now appears from the following table, where Φ is the distribution function of the standard normal distribution (Table B.2).

Alternative hypothesis	Significance probability
$\mathbf{H}_1 : \ \mu > \mu_0$	$\Phi(-u)$
$\mathbf{H}_1 : \ \mu < \mu_0$	$\Phi(u)$
$\mathbf{H}_1 : \ \mu \neq \mu_0$	$2 \cdot \Phi(-\lvert u\rvert)$

Normally, we reject $\mathbf{H}_0$ if the significance probability is less than 5%.

17.2 One sample, unknown variance, $\mathrm{H}_0 : \mu = \mu_0$ (Student's t test)

Let there be given a sample $x_1, \ldots, x_n$ of n independent observations from a normal distribution with **unknown expected value** μ and **unknown variance** σ^2. We wish to test the null hypothesis

$$\mathbf{H}_0 : \ \mu = \mu_0 \ .$$

For this purpose, we compute the statistic

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} = \frac{\sum_{i=1}^n x_i - n\mu_0}{\sqrt{n \cdot s^2}} \ ,$$

where s^2 is the empirical variance (see section 6.3).

The significance probability now appears from the following table where F_{Student} is the distribution function of Student's t distribution with $df = n - 1$ degrees of freedom (Table B.4).

Alternative hypothesis	Significance probability
$\mathbf{H}_1 : \mu > \mu_0$	$1 - F_{\text{Student}}(t)$
$\mathbf{H}_1 : \mu < \mu_0$	$1 - F_{\text{Student}}(-t)$
$\mathbf{H}_1 : \mu \neq \mu_0$	$2 \cdot (1 - F_{\text{Student}}(\lvert t \rvert))$

Normally, we reject $\mathbf{H}_0$ if the significance probability is less than 5%.

EXAMPLE. The headmaster of a school wishes to confirm statistically that his students have performed significantly miserably in the 2008 final exams. For this purpose, $n = 10$ students are picked at random. Their final scores are

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
7.6	7.7	7.5	5.8	5.7	7.9	5.4	6.7	7.9	9.4

The national average for 2008 is 8.27. It is reasonable to assume that the final scores are normally distributed. However, the variance is unknown. Therefore, we apply Student's t test to test the null hypothesis

$$\mathbf{H}_0 : \mu = 8.27$$

against the alternative hypothesis

$$\mathbf{H}_1 : \mu < 8.27 \, .$$

We compute the mean value of the observations as $\bar{x} = 7.17$ and the empirical standard deviation as $s = 1.26$. We obtain the statistic

$$t = \frac{\sqrt{10}(7.17 - 8.27)}{1.26} = -2.76 \, .$$

Looking up in Table B.4 under $df = n - 1 = 9$ degrees of freedom gives a significance probability

$$1 - F_{\text{Student}}(-t) = 1 - F_{\text{Student}}(2.76)$$

between 1% and 2.5%. We may therefore reject $\mathbf{H}_0$ and confirm the headmaster's assumption that his students have performed significantly poorer than the rest of the country.

17.3 One sample, unknown expected value, $\mathbf{H}_0 : \sigma^2 = \sigma_0^2$

THEOREM. Let there be given n independent observations $x_1, \ldots, x_n$ from a normal distribution with variance σ^2. The statistic

$$q = \frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}$$

is then χ^2 distributed with $df = n - 1$ degrees of freedom (here s^2 is the empirical variance).

Let there be given a sample $x_1, \ldots, x_n$ of n independent observations from a normal distribution with **unknown expected value** μ and **unknown variance** σ^2. We wish to test the null hypothesis

$$\mathbf{H}_0 : \ \sigma^2 = \sigma_0^2 \,.$$

For this purpose, we compute the statistic

$$q = \frac{(n-1)s^2}{\sigma_0^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2}$$

where s^2 is the empirical variance.

The significance probability can now be read from the following table where F_{χ^2} is the distribution function of the χ^2 distribution with $df = n-1$ degrees of freedom (Table B.3).

Alternative hypothesis	Significance probability
$\mathbf{H}_1 : \ \sigma^2 > \sigma_0^2$	$1 - F_{\chi^2}(q)$
$\mathbf{H}_1 : \ \sigma^2 < \sigma_0^2$	$F_{\chi^2}(q)$
$\mathbf{H}_1 : \ \sigma^2 \neq \sigma_0^2$	$2 \cdot \min\{F_{\chi^2}(q), 1 - F_{\chi^2}(q)\}$

Normally, $\mathbf{H}_0$ is rejected if the significance probability is smaller than 5%.

Note: In practice, we always test against the alternative hypothesis $\mathbf{H}_1 : \ \sigma^2 > \sigma_0^2$.

17.4 Example

Consider the following twenty observations originating from a normal distribution with unknown expected value and variance:

91	97	98	112	91	97	116	108	108	100
107	98	92	103	100	99	98	104	104	97

We wish to test the null hypothesis

$\mathbf{H}_0$: the standard deviation is at most 5 (i.e. the variance is at most 25)

against the alternative hypothesis

$\mathbf{H}_1$: the standard deviation is greater than 5 (i.e. the variance is greater than 25).

The empirical variance is found to be $s^2 = 45.47$ and we thus find the statistic

$$q = \frac{(20-1)\cdot 45.47}{5^2} = 34.56 \ .$$

By looking up in Table B.3 under $df = 19$ degrees of freedom, we find a significance probability around 2%. We can thus reject $\mathbf{H}_0$.

(Actually, the observations came from a normal distribution with expected value $\mu = 100$ and standard deviation $\sigma = 6$. The test is thus remarkably sensitive.)

17.5 Two samples, known variances, $\mathrm{H}_0 : \mu_1 = \mu_2$

Let there be given a sample $x_1, \ldots, x_n$ from a normal distribution with **unknown expected value** μ_1 and **known variance** σ_1^2. Let there in addition be given a sample $y_1, \ldots, y_m$ from a normal distribution with **unknown expected value** μ_2 and **known variance** σ_2^2. It is assumed that the two samples are independent of each other.

We wish to test the null hypothesis

$$\mathbf{H}_0: \ \mu_1 = \mu_2 \ .$$

For this purpose, we compute the statistic

$$u = \frac{\bar{x}-\bar{y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \ .$$

The significance probability is read from the following table where Φ is the distribution function of the standard normal distribution (Table B.3).

Alternative hypothesis	Significance probability
$\mathbf{H}_1: \ \mu_1 > \mu_2$	$\Phi(-u)$
$\mathbf{H}_1: \ \mu_1 < \mu_2$	$\Phi(u)$
$\mathbf{H}_1: \ \mu_1 \neq \mu_2$	$2\cdot\Phi(-\lvert u\rvert)$

Normally, we reject $\mathbf{H}_0$ if the significance probability is smaller than 5%.

Note: In real life, the preconditions of this test are rarely met.

17.6 Two samples, unknown variances, $\mathrm{H}_0 : \mu_1 = \mu_2$ (Fisher-Behrens)

Let the situation be as in section 17.5, but suppose that the variances σ_1^2 and σ_2^2 are unknown. The problem of finding a suitable statistic to test the null hypothesis

$$\mathbf{H}_0 : \ \mu_1 = \mu_2$$

is called the **Fisher-Behrens problem** and has no satisfactory solution.

If $n, m > 30$, one can re-use the test from section 17.5 with the alternative statistic

$$u^* = \frac{\bar{x} - \bar{y}}{\sqrt{s_1^2/n + s_2^2/m}}$$

where s_1^2 and s_2^2 are the empirical variances of the x's and y's, respectively.

17.7 Two samples, unknown expected values, $\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2$

Let there be given a sample $x_1, \ldots, x_n$ from a normal distribution with **unknown expected value** μ_1 and **unknown variance** σ_1^2. In addition, let there be given a sample $y_1, \ldots, y_m$ from a normal distribution with **unknown expected value** μ_2 and **unknown variance** σ_2^2. It is assumed that the two samples are independent of each other.

We wish to test the null hypothesis

$$\mathbf{H}_0 : \ \sigma_1 = \sigma_2 \ .$$

For this purpose, we compute the statistic

$$v = \frac{s_1^2}{s_2^2} = \frac{\text{empirical variance of the } x\text{'s}}{\text{empirical variance of the } y\text{'s}} \ .$$

Further, put

$$v^* = \max\left\{v, \frac{1}{v}\right\} \ .$$

The significance probability now appears from the following table where F_{Fisher} is the distribution function of Fisher's F distribution with $n-1$ degrees of freedom in the numerator and $m-1$ degrees of freedom in the denominator (Table B.5).

Alternative hypothesis	Significance probability
$\mathbf{H}_1 : \ \sigma_1^2 > \sigma_2^2$	$1 - F_{\text{Fisher}}(v)$
$\mathbf{H}_1 : \ \sigma_1^2 < \sigma_2^2$	$1 - F_{\text{Fisher}}(1/v)$
$\mathbf{H}_1 : \ \sigma_1^2 \neq \sigma_2^2$	$2 \cdot (1 - F_{\text{Fisher}}(v^*))$

Normally, $\mathbf{H}_0$ is rejected if the significance probability is smaller than 5%.

If $\mathbf{H}_0$ is accepted, the common variance $\sigma_1^2 = \sigma_2^2$ is estimated by the **"pooled" variance**

$$s_{\text{pool}}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2}{n+m-2} = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2} \ .$$

17.8 Two samples, unknown common variance, $\mathrm{H}_0 : \mu_1 = \mu_2$

Let there be given a sample $x_1, \ldots, x_n$ from a normal distribution with **unknown expected value** μ_1 and **unknown variance** σ^2. In addition, let there be given a sample $y_1, \ldots, y_m$ from a normal distribution with **unknown expected value** μ_2 and **the same variance** σ^2. It is assumed that the two samples are independent of each other.

We wish to test the null hypothesis

$$\mathbf{H}_0 : \ \mu_1 = \mu_2 \ .$$

For this purpose, we compute the statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(1/n + 1/m)s^2_{\text{pool}}}}$$

where s^2_{pool} is the "pooled" variance as given in section 17.7.

The significance probability now appears from the following table where F_{Student} is the distribution function of Student's t distribution with $n + m - 2$ degrees of freedom (Table B.4).

Alternative hypothesis	Significance probability
$\mathbf{H}_1 : \ \mu_1 > \mu_2$	$1 - F_{\text{Student}}(t)$
$\mathbf{H}_1 : \ \mu_1 < \mu_2$	$1 - F_{\text{Student}}(-t)$
$\mathbf{H}_1 : \ \mu_1 \neq \mu_2$	$2 \cdot (1 - F_{\text{Student}}(\|t\|))$

Normally, $\mathbf{H}_0$ is rejected if the significance probability is smaller than 5%.

17.9 Example (comparison of two expected values)

Suppose we are given seven independent observations from a normally distributed random variable X:

$$x_1 = 26 \ , \ x_2 = 21 \ , \ x_3 = 15 \ , \ x_4 = 7 \ , \ x_5 = 15 \ , \ x_6 = 28 \ , \ x_7 = 21$$

and also four independent observations from a normally distributed random variable Y:

$$y_1 = 29 \ , \ y_2 = 31 \ , \ y_3 = 17 \ , \ y_4 = 22 \ .$$

We wish to test the hypothesis

$$\mathbf{H}_0 : E(X) = E(Y) \ .$$

In order to be able to test this, we need to test first whether X and Y have the same variance. We therefore test the **auxiliary hypothesis**

$$\mathbf{H}_0^* : \text{var}(X) = \text{var}(Y)$$

against the alternative

$$\mathbf{H}_1^* : \text{var}(X) \neq \text{var}(Y) \ .$$

For this purpose, we compute the statistic

$$v = \frac{s_1^2}{s_2^2} = \frac{52.3}{41.6} = 1.26$$

as in section 17.7, as well as

$$v^* = \max\left\{v, \frac{1}{v}\right\} = 1.26 \ .$$

Looking up in Table B.5 with $7 - 1 = 6$ degrees of freedom in the numerator and $4 - 1 = 3$ degrees of freedom in the denominator shows that the significance probability is clearly greater than 20%, and we may therefore accept the auxiliary hypothesis $\mathbf{H}_0^*$.

Now we return to the test of $\mathbf{H}_0$ against the alternative hypothesis

$$\mathbf{H}_1 : E(X) \neq E(Y) \ .$$

The "pooled" variance is found to be

$$s^2_{\text{pool}} = \frac{6s_1^2 + 3s_2^2}{9} = 48.8 \ .$$

The statistic thereby becomes

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(1/7 + 1/4)s^2_{\text{pool}}}} = \frac{19 - 24.8}{\sqrt{(1/7 + 1/4)48.8}} = -1.31 \ .$$

Therefore, the significance probability is found to be

$$2 \cdot (1 - F_{\text{Student}}(|t|)) = 2 \cdot (1 - F_{\text{Student}}(1.31)) \approx 2 \cdot (1 - 0.90) = 20\%$$

by looking up Student's t distribution with $7 + 4 - 2 = 9$ degrees of freedom in Table B.4. Consequently, we *cannot* reject $\mathbf{H}_0$.

18 Analysis of variance (ANOVA)

18.1 Aim and motivation

Analysis of variance, also known as ANOVA, is a clever method of comparing the mean values from more than two samples. Analysis of variance is a natural extension of the tests in the previous chapter.

18.2 k samples, unknown common variance, $\mathrm{H}_0 : \mu_1 = \cdots = \mu_k$

Let $X_1, \ldots, X_k$ be k independent, normally distributed random variables, with expected values $\mu_1, \ldots, \mu_k$ and **common variance** σ^2. From each X_i, let there be given a sample consisting of n_i observations. Let $\bar{x}_j$ and s_j^2 be mean value and empirical variance of the sample from X_j.

We wish to test the null hypothesis

$$\mathbf{H}_0 : \ \mu_1 = \cdots = \mu_k$$

against all alternatives. For this purpose, we estimate the common variance σ^2 in two different ways.

The **variance estimate within the samples** is

$$s_I^2 = \frac{1}{n-k} \sum_{j=1}^{k} (n_j - 1) s_j^2 \ .$$

The **variance estimate between the samples** is

$$s_M^2 = \frac{1}{k-1} \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 \ .$$

s_I^2 estimates σ^2 regardless of whether $\mathbf{H}_0$ is true or not. s_M^2 only estimates σ^2 correctly if $\mathbf{H}_0$ is true. If $\mathbf{H}_0$ is false, then s_M^2 estimates too high.

Now consider the statistic

$$v = \frac{s_M^2}{s_I^2} \ .$$

The significance probability is

$$1 - F_{\text{Fisher}}(v)$$

where F_{Fisher} is the distribution function of Fisher's F distribution with $k-1$ degrees of freedom in the numerator and $n-k$ degrees of freedom in the denominator (Table B.5).

18.3 Two examples (comparison of mean values from three samples)

Let three samples be given:

sample 1: 29, 28, 29, 21, 28, 22, 22, 29, 26, 26

sample 2: 22, 21, 18, 28, 23, 25, 25, 28, 23, 26

sample 3: 24, 23, 26, 20, 33, 23, 26, 24, 27, 22

It is assumed that the samples originate from independent normal distributions with common variance. Let μ_i be the expected value of the i'th normal distribution. We wish to test the null hypothesis

$$\mathbf{H}_0 : \ \mu_1 = \mu_2 = \mu_3 \ .$$

(As a matter of fact, all the observations originate from a normal distribution with expected value 25 and variance 10, so the test shouldn't lead to a rejection of $\mathbf{H}_0$.) We thus have $k = 3$ samples each consisting of $n_i = 10$ observations, a total of $n = 30$ observations. A computation gives the following variance estimate **within** the samples:

$$s_I^2 = 10.91$$

and the following variance estimate **between** the samples:

$$s_M^2 = 11.10$$

(Since we know that $\mathbf{H}_0$ is true, both s_I^2 and s_M^2 should estimate $\sigma^2 = 10$ well, which they also indeed do.) Now we compute the statistic:

$$v = \frac{s_M^2}{s_I^2} = \frac{11.10}{10.91} = 1.02 \ .$$

Looking up in Table B.5 under $k-1 = 2$ degrees of freedom in the numerator and $n-k = 27$ degrees of freedom in the denominator shows that the significance probability is more than 10%. The null hypothesis $\mathbf{H}_0$ cannot be rejected.

Somewhat more carefully, the computations can be summed up in a table as follows:

Sample number	**1**	**2**	**3**
	29	22	24
	28	21	23
	29	18	26
	21	28	20
	28	23	33
	22	25	23
	22	25	26
	29	28	24
	26	23	27
	26	26	22
Mean value $\bar{x}_j$	26.0	23.9	24.8
Empirical variance s_j^2	10.22	9.88	12.62

$\bar{x} = 24.9$	(grand mean value)
$s_I^2 = (s_1^2 + s_2^2 + s_3^2)/3 = 10.91$	(variance within samples)
$s_M^2 = 5\sum(\bar{x}_j - \bar{x})^2 = 11.10$	(variance between samples)
$v = s_M^2/s_I^2 = 1.02$	(statistic)

If we add 5 to all the observations in sample 3, we get the following table instead:

Sample number	**1**	**2**	**3**
	29	22	29
	28	21	28
	29	18	31
	21	28	25
	28	23	38
	22	25	28
	22	25	31
	29	28	29
	26	23	32
	26	26	27
Mean value $\bar{x}_j$	26.0	23.9	29.8
Empirical variance s_j^2	10.22	9.88	12.62
$\bar{x} = 26.6$			(grand mean value)
$s_I^2 = (s_1^2 + s_2^2 + s_3^2)/3 = 10.91$			(variance within samples)
$s_M^2 = 5\sum(\bar{x}_j - \bar{x})^2 = 89.43$			(variance between samples)
$v = s_M^2/s_I^2 = 8.20$			(statistic)

Note how the variance within the samples doesn't change, whereas the variance between the samples is now far too large. Thus, the statistic $v = 8.20$ also becomes large and the significance probability is seen in Table B.5 to be less than 1%. Therefore, we reject the null hypothesis $\mathbf{H}_0$ of equal expected values (which was also to be expected, since $\mathbf{H}_0$ is now manifestly false).

19 The chi-squared test (or χ^2 test)

19.1 χ^2 test for equality of distribution

The reason why the χ^2 distribution is so important is that it can be used to test whether a given set of observations comes from a certain distribution. In the following sections, we shall see many examples of this. The test, which is also called *Pearson's χ^2 test* or *χ^2 test for goodness of fit*, is carried out as follows:

1. First, divide the observations into categories. Let us denote the number of categories by k and the number of observations in the i'th category by O_i. The total number of observations is thus $n = O_1 + \cdots + O_k$.

2. Formulate a null hypothesis $\mathbf{H}_0$. This null hypothesis must imply what the probability p_i is that an observation belongs to the i'th category.

3. Compute the statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \,.$$

As mentioned, O_i is the *observed* number in the i'th category. Further, E_i is the *expected* number in the i'th category (expected according to the null hypothesis, that is): $E_i = np_i$. Incidentally, the statistic $\chi = \sqrt{\chi^2}$ is sometimes called the *discrepancy*.

4. Find the **significance probability**

$$P = 1 - F(\chi^2)$$

where $F = F_{\chi^2}$ is the distribution function of the χ^2 distribution with df degrees of freedom (look up in Table B.3). $\mathbf{H}_0$ is rejected if P is smaller than 5% (or whatever significance level one chooses). The number of degrees of freedom is normally $df = k-1$, i.e. one less than the number of categories. If, however, one uses the observations to estimate the probability parameters p_i of the null hypothesis, df becomes smaller.

Remember: Each estimated parameter costs one degree of freedom.

Note: It is logical to reject H_0 if χ^2 is large, because this implies that the difference between the observed and the expected numbers is large.

19.2 The assumption of normal distribution

Since the χ^2 test rests upon a normal approximation, it only applies provided there are not too few observations.

Remember: The χ^2 test applies if the expected number E_i is at least five in each category. If, however, there are more than five categories, an expected number of at least three in each category suffices.

19.3 Standardized residuals

If the null hypothesis regarding equality of distribution is rejected by a χ^2 test, this was because some of the observed numbers deviated widely from the expected numbers. It is then interesting to investigate exactly which observed numbers are extreme. For this purpose, we compute the **standardized residuals**

$$r_i = \frac{O_i - np_i}{\sqrt{np_i(1-p_i)}} = \frac{O_i - E_i}{\sqrt{E_i(1-p_i)}}$$

for each category. If the null hypothesis were true, each r_i would be normally distributed with expected value $\mu = 0$ and standard deviation $\sigma = 1$. Therefore:

Remember: Standardized residuals numerically greater than 2 are signs of an extreme observed number.

It can very well happen that standardized residuals numerically greater than 2 occur even though the χ^2 test does not lead to rejection of the null hypothesis. This does **not** mean that the null hypothesis should be rejected after all. In particular when one has a large number of categories, it will not be unusual to find some large residuals.

Warning: Only compute the standardized residuals if the null hypothesis has been *rejected* by a χ^2 test.

19.4 Example (women with five children)

EXERCISE. A hospital has registered the sex of the children of 1045 women who each have five children. Result:

	O_i
5 girls	58
4 girls + 1 boy	149
3 girls + 2 boys	305
2 girls + 3 boys	303
1 girl + 4 boys	162
5 boys	45

Test the hypothesis $\mathbf{H}_0$ that, at every birth, the probability of a boy is the same as the probability of a girl.

SOLUTION. If $\mathbf{H}_0$ is true, the above table consists of 1045 observations from a $\text{Bin}(5, 1/2)$ distribution. The point probabilities in a $\text{Bin}(5, 1/2)$ distribution are

	p_i
5 girls	0.0313
4 girls + 1 boy	0.1563
3 girls + 2 boys	0.3125
2 girls + 3 boys	0.3125
1 girl + 4 boys	0.1563
5 boys	0.0313

The expected numbers $E_i = 1045 \cdot p_i$ then become

	E_i
5 girls	32.7
4 girls + 1 boy	163.3
3 girls + 2 boys	326.6
2 girls + 3 boys	326.6
1 girl + 4 boys	163.3
5 boys	32.7

The statistic is computed:

$$\chi^2 = \frac{(58-32.7)^2}{32.7} + \frac{(149-163.3)^2}{163.3} + \frac{(305-326.6)^2}{326.6} + \frac{(303-326.6)^2}{326.6} + \frac{(162-163.3)^2}{163.3} + \frac{(45-32.7)^2}{32.7} = 28.6\,.$$

Since the observations are divided into six categories, we compare the statistic with the χ^2 distribution with $df = 6 - 1 = 5$ degrees of freedom. Table B.3 shows that the significance probability is well below 0.5%. We can therefore with great confidence reject the hypothesis that the boy-girl ratio is $\text{Bin}(5, 1/2)$ distributed.

Let us finally compute the standardized residuals:

	r_i
5 girls	4.5
4 girls + 1 boy	–1.2
3 girls + 2 boys	–1.4
2 girls + 3 boys	–1.6
1 girl + 4 boys	–0.1
5 boys	2.2

We note that it is the numbers of women with five children of the same sex which are extreme and make the statistic large.

19.5 Example (election)

EXERCISE. At the election for the Danish parliament in February 2005, votes were distributed among the parties as follows (as percentages):

A	B	C	F	O	V	Ø	others
25.8	9.2	10.3	6.0	13.3	29.0	3.4	3.0

In August 2008, an opinion poll was carried out in which 1000 randomly chosen persons were asked which party they would vote for now. The result was:

A	B	C	F	O	V	Ø	others
242	89	98	68	141	294	43	25

Has the popularity of the different parties changed since the election?

SOLUTION. We test the null hypothesis $\mathbf{H}_0$ that the result of the opinion poll is an observation from a multinomial distribution with $k = 8$ categories and probability parameters p_i as given in the table above. The expected observations (given the null hypothesis) are:

A	B	C	F	O	V	Ø	others
258	92	103	60	133	290	34	30

Now we compute the statistic χ^2:

$$\chi^2 = \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} = \frac{(242-258)^2}{258} + \cdots + \frac{(25-30)^2}{30} = 6.15\ .$$

By looking up in Table B.3 under the χ^2 distribution with $df = 8-1 = 7$ degrees of freedom, it is only seen that the significance probability is below 50%. Thus, we have no statistical evidence to conclude that the popularity of the parties has changed.

Let us ignore the warning in section 19.3 and compute the standardized residuals. For category A, for example, we find

$$r = \frac{242 - 1000 \cdot 0.258}{\sqrt{1000 \cdot 0.258 \cdot 0.742}} = -1.16\ .$$

Altogether we get

A	B	C	F	O	V	Ø	others
–1.16	–0.33	–0.52	1.06	0.74	0.28	1.57	–0.93

Not surprisingly, all standardized residuals are numerically smaller than 2.

19.6 Example (deaths in the Prussian cavalry)

In the period 1875–1894 the number of deaths caused by horse kicks was registered in 10 of the regiments of the Prussian cavalry. Of the total of 200 "regiment-years", there were 109 years with no deaths, 65 years with one death, 22 years with two deaths, three years with three deaths, and one year with four deaths. We wish to investigate whether these numbers come from a Poisson distribution $\text{Pois}(\lambda)$.

In order to get expected numbers greater than five (or at least to come close to that), we group the years with three and four deaths into a single category and thus obtain the following observed numbers O_i of years with i deaths:

i	O_i
0	109
1	65
2	22
≥ 3	4

The intensity λ is estimated as $\hat{\lambda} = 122/200 = 0.61$, since there were a total of 122 deaths during the 200 regiment-years. The point probabilities of a $\text{Pois}(0.61)$ distribution are

i	p_i
0	0.543
1	0.331
2	0.101
≥ 3	0.024

The expected numbers thus become

i	E_i
0	108.7
1	66.3
2	20.2
≥ 3	4.8

The reader should let himself be impressed by the striking correspondence between expected and observed numbers. It is evidently superfluous to carry the analysis any further, but let us compute the statistic anyway:

$$\chi^2 = \frac{(109-108.7)^2}{108.7} + \frac{(65-66.3)^2}{66.3} + \frac{(22-20.2)^2}{20.2} + \frac{(4-4.8)^2}{4.8} = 0.3\ .$$

Since there are four categories and we have estimated one parameter using the data, the statistic should be compared with the χ^2 distribution with $df = 4-1-1 = 2$ degrees of freedom. As expected, Table B.3 shows a significance probability well above 50%.

Incidentally, the example comes from Ladislaus von Bortkiewicz's 1898 book *Das Gesetz der kleinen Zahlen*.

20 Contingency tables

20.1 Definition, method

Suppose that a number of observations are given and that the observations are divided into categories according to two different criteria. The number of observations in each category can then be displayed in a **contingency table**. The purpose of the test presented here is to test whether there is **independence** between the two criteria used to categorize the observations.

METHOD. Let there be given an $r \times s$ table, i.e. a table with r rows and s columns:

a_{11}	a_{12}	$\dots$	$\dots$	a_{1s}
a_{21}	a_{22}	$\dots$	$\dots$	a_{2s}
$\vdots$	$\vdots$			$\vdots$
$\vdots$	$\vdots$			$\vdots$
a_{r1}	a_{r2}	$\dots$	$\dots$	a_{rs}

It has row sums $R_i = \sum_{j=1}^{s} a_{ij}$, column sums $S_j = \sum_{i=1}^{r} a_{ij}$, and total sum

$$N = \sum_{i,j} a_{ij} \ .$$

These are the *observed* numbers O. The **row probabilities** are estimated as

$$\hat{p}_{i\cdot} = \frac{R_i}{N} \ ,$$

and the **column probabilities** as

$$\hat{p}_{\cdot j} = \frac{S_j}{N} \ .$$

If there is independence between rows and columns, the **cell probabilities** can be estimated as

$$\hat{p}_{ij} = \hat{p}_{i\cdot}\hat{p}_{\cdot j} = \frac{R_i S_j}{N^2} \ .$$

We can thus compute the *expected* numbers E:

$\frac{R_1 S_1}{N}$	$\frac{R_1 S_2}{N}$	$\dots$	$\dots$	$\frac{R_1 S_s}{N}$
$\frac{R_2 S_1}{N}$	$\frac{R_2 S_2}{N}$	$\dots$	$\dots$	$\frac{R_2 S_s}{N}$
$\vdots$	$\vdots$			$\vdots$
$\vdots$	$\vdots$			$\vdots$
$\frac{R_r S_1}{N}$	$\frac{R_r S_2}{N}$	$\dots$	$\dots$	$\frac{R_r S_s}{N}$

since the expected number in the (i,j)'th cell is

$$E = N\hat{p}_{ij} = R_i S_j / N \ .$$

Now we compute the statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \sum \frac{(a_{ij} - R_i S_j/N)^2}{R_i S_j/N}$$

where the summation is carried out over each cell of the table. If the independence hypothesis holds true and the expected number is at least 5 in each cell, then the statistic is χ^2 distributed with

$$df = (r-1)(s-1)$$

degrees of freedom.

Important! If the data are given as percentages, they must be expressed as absolute numbers before insertion into the contingency table.

20.2 Standardized residuals

If the independence hypothesis is rejected by a χ^2 test, one might, as in section 19.3, be interested in determining which cells contain observed numbers deviating extremely from the expected numbers. The **standardized residuals** are computed as

$$r_{ij} = \frac{O_{ij} - R_i S_j/n}{\sqrt{\big(R_i S_j/n\big)\big(1 - R_i/n\big)\big(1 - S_j/n\big)}} \ .$$

If the independence hypothesis were true, each r_{ij} would be normally distributed with expected value $\mu = 0$ and standard deviation $\sigma = 1$. Standardized residuals numerically greater than 2 are therefore signs of an extreme observed number.

20.3 Example (students' political orientation)

EXERCISE. At three Danish universities, 488 students were asked about their faculty and which party they would vote for if there were to be an election tomorrow. The result (in simplified form) was:

	A	B	C	F	O	V	Ø	R_i
Humanities	37	48	15	26	4	17	10	157
Natural Sci.	32	38	19	18	7	51	2	167
Social Sci.	32	24	15	7	12	69	5	164
S_j	101	110	49	51	23	137	17	488

Investigate whether there is independence between the students' political orientation and their faculty.

SOLUTION. We are dealing with a 3×7 table and perform a χ^2 test for independence. First, the expected numbers

$$E = \frac{R_i S_j}{488}$$

are computed and presented in a table:

	A	B	C	F	O	V	Ø
Humanities	32.5	35.4	15.8	16.4	7.4	44.1	5.5
Natural Sci.	34.6	37.6	16.8	17.5	7.9	46.9	5.8
Social Sci.	33.9	37.0	16.5	17.1	7.7	46.0	5.7

Now the statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

can be computed, since the observed numbers O are the numbers in the first table:

$$\chi^2 = \frac{(37-32.5)^2}{32.5} + \cdots + \frac{(5-5.7)^2}{5.7} = 60.9\,.$$

The statistic is to be compared with a χ^2 distribution with $df = (3-1)(7-1) = 12$ degrees of freedom. Table B.3 shows that the significance probability is well below 0.1%, and we therefore confidently reject the independence hypothesis.

Let us now compute the standardized residuals to see in which cells the observed numbers are extreme. We use the formula for r_{ij} in section 20.2 and get

	A	B	C	F	O	V	Ø
Humanities	1.1	2.9	–0.2	3.0	–1.6	–5.8	2.4
Natural Sci.	–0.6	0.1	0.7	0.2	–0.4	0.9	–2.0
Social Sci.	–0.5	–3.0	–0.5	–3.2	1.9	4.9	–0.4

We find that there are extreme observations in many cells.

20.4 χ^2 test for 2×2 tables

A contingency table with two rows and two columns is called a **2×2 table.** Let us write the observed numbers as follows:

$$\begin{array}{c|c} a & b \\ \hline c & d \end{array}$$

The statistic thus becomes

$$\chi^2 = \left(\frac{ad-bc}{N}\right)^2 \left(\frac{1}{E_{11}} + \frac{1}{E_{12}} + \frac{1}{E_{21}} + \frac{1}{E_{22}}\right)$$

where $N = a+b+c+d$ is the total number of observations, and E_{ij} is the expected number in the ij'th cell. The statistic χ^2 is to be compared with the χ^2 distribution with $df = (2-1)(2-1) = 1$ degree of freedom.

If we wish to perform a one-sided test of the independence hypothesis, the statistic

$$u = \left(\frac{ad-bc}{N}\right) \sqrt{\left(\frac{1}{E_{11}} + \frac{1}{E_{12}} + \frac{1}{E_{21}} + \frac{1}{E_{22}}\right)}$$

is used instead. Under the independence hypothesis, u will be standard normally distributed.

20.5 Fisher's exact test for 2×2 tables

Given a 2 × 2 table, nothing stands in the way of using the χ^2 test, but there is a better test in this situation called **Fisher's exact test.** Fisher's exact test does not use any normal approximation, and may therefore still be applied when the number of expected observations in one or more of the cells is smaller than five.

METHOD. Let there be given a 2×2 table:

$$\begin{array}{c|c} a & b \\ \hline c & d \end{array}$$

with row sums $R_1 = a+b$ and $R_2 = c+d$ and column sums $S_1 = a+c$ and $S_2 = b+d$ and total sum $N = R_1 + R_2 = S_1 + S_2 = a+b+c+d$. We test the independence hypothesis $\mathbf{H}_0$ against the alternative hypothesis $\mathbf{H}_1$ that the "diagonal probabilities" p_{11} and p_{22} are greater than what they would have been had there been independence. (This situation can always be arranged by switching the rows if necessary.) The conditional probability of obtaining exactly the 2×2 table above, *given that the row sums are R_1 and R_2, and that the column sums are S_1 and S_2*, is

$$P_{\text{conditional}} = \frac{R_1!\,R_2!\,S_1!\,S_2!}{N!\,a!\,b!\,c!\,d!}\ .$$

The significance probability in Fisher's exact test is the sum of $P_{\text{conditional}}$ taken on all 2×2 tables with the same row and column sums as in the given table, and which are at least as extreme as the given table:

$$P_{\text{Fisher}} = \sum_{i=0}^{\min\{b,c\}} \frac{R_1!\,R_2!\,S_1!\,S_2!}{N!\,(a+i)!\,(b-i)!\,(c-i)!\,(d+i)!}\ .$$

The independence hypothesis $\mathbf{H}_0$ is rejected if P_{Fisher} is smaller than 5% (or whatever significance level one has chosen).

ADDENDUM: If a two-sided test is performed, i.e. if one does not test against any specific alternative hypothesis, the significance probability becomes $2 \cdot P_{\text{Fisher}}$. It is then necessary that the 2×2 table is written in such a way that the observed numbers in the diagonal are greater than the expected numbers (this can always be obtained by switching the rows if necessary).

20.6 Example (Fisher's exact test)

In a medical experiment concerning alternative treatments, ten patients are randomly divided into two groups with five patients in each. The patients in the first group receive acupuncture, while the patients in the other group receive no treatment. It is then seen which patients are fit or ill at the end of the experiment. The result can be presented in a 2×2 table:

	fit	ill
acupuncture	4	1
no treatment	2	3

The significance probability in Fisher's exact test is computed as

$$P_{\text{Fisher}} = \sum_{i=0}^{1} \frac{5!\,5!\,6!\,4!}{10!\,(4+i)!\,(1-i)!\,(2-i)!\,(3+i)!} = 26\%\ .$$

With such a large significance probability, there is no evidence that acupuncture had any effect.

21 Distribution-free tests

In all tests considered so far, we have known something about the distribution from which the given samples originated. We knew, for example, that the distribution was a normal distribution even though we didn't know the expected value or the standard deviation.

Sometimes, though, one knows nothing at all about the underlying distribution. It then becomes necessary to use a **distribution-free test** (also known as a **non-parametric test**). The two examples considered in this chapter are due to Frank Wilcoxon (1892–1965).

21.1 Wilcoxon's test for one set of observations

Let there be given n independent observations $d_1, \ldots, d_n$ from an unknown distribution. We test the null hypothesis

$\mathbf{H}_0$: *The unknown distribution is symmetric around 0.*

Each observation d_i is given a **rank** which is one of the numbers $1, 2, \ldots, n$. The observation with the smallest numerical value is assigned rank 1, the observation with the second smallest numerical value is assigned rank 2, etc. Now define the statistics

$t_+ = \sum$(ranks corresponding to positive d_i),

$t_- = \sum$(ranks corresponding to negative d_i).

(One can check at this point whether $t_+ + t_- = n(n+1)/2$; if not, one has added the numbers incorrectly.) If $\mathbf{H}_0$ holds true, then t_+ and t_- should be more or less equal. When to reject $\mathbf{H}_0$ depends on which alternative hypothesis is tested against.

If we test $\mathbf{H}_0$ against the alternative hypothesis

$\mathbf{H}_1$: *The unknown distribution primarily gives positive observations,*

then $\mathbf{H}_0$ is rejected if t_- is extremely small. Choose a significance level α and consult Table B.8

under n and α. If t_- is smaller than or equal to the table value, $\mathbf{H}_0$ is rejected. If t_- is greater than the table value, $\mathbf{H}_0$ is accepted.

If we test $\mathbf{H}_0$ against the alternative hypothesis

$\mathbf{H}_1$: *The unknown distribution primarily gives negative observations,*

then $\mathbf{H}_0$ is rejected if t_+ is extremely small. Choose a significance level α and consult Table B.8 under n and α. If t_+ is smaller than or equal to the table value, $\mathbf{H}_0$ is rejected. If t_+ is greater than the table value, $\mathbf{H}_0$ is accepted.

If we don't test $\mathbf{H}_0$ against any particular alternative hypothesis, the null hypothesis is rejected if the minimum $t := \min\{t_+, t_-\}$ is extremely small. Choose a significance level α and consult Table B.8 under n and $\alpha/2$ (if, for example, we choose the significance level $\alpha = 5\%$, then we look up in the table under n and 0.025). If t is smaller than or equal to the table value, we reject $\mathbf{H}_0$. If t is greater than the table value, we accept $\mathbf{H}_0$.

The above test applies in particular when two sets of observations $x_1, \dots, x_n$ and $y_1, \dots, y_n$ are given and d_i is the difference between the "before values" x_i and the "after values" y_i, i.e. $d_i = x_i - y_i$. If there are only random, unsystematic differences between the before and after values, it follows that the d_i's are distributed symmetrically around 0.

21.2 Example

An experiment involving ten patients is carried out to determine whether physical exercise lowers blood pressure. At the beginning of the experiment, the patients' blood pressures are measured. These observations are denoted $x_1, \dots, x_{10}$. After a month of exercise, the blood pressures are measured again. These observations are denoted $y_1, \dots, y_{10}$. We now test the null hypothesis

$\mathbf{H}_0$: *Physical exercise has no influence on blood pressure. The ten differences $d_i = x_i - y_i$ are therefore distributed symmetrically around 0,*

against the alternative hypothesis

$\mathbf{H}_1$: *Physical exercise causes the blood pressure to decrease. The ten differences d_i are therefore primarily positive.*

We compute the ranks and t_+ and t_-:

Person	1	2	3	4	5	6	7	8	9	10
Before x_i	140	125	110	130	170	165	135	140	155	145
After y_i	137	137	102	104	172	125	140	110	140	126
Difference d_i	3	–12	8	26	–2	40	–5	30	15	19
Rank	2	5	4	8	1	10	3	9	6	7

$t_+ = 2 + 4 + 6 + 7 + 8 + 9 + 10 = 46,$

$t_- = 1 + 3 + 5 = 9.$

We shall reject $\mathbf{H}_0$ if $t_- = 9$ is extremely small. Table B.8 with significance level $\alpha = 5\%$ shows that "extremely small" means $\leqq 10$. Conclusion: *The test shows that the null hypothesis* $\mathbf{H}_0$ *must be rejected against the alternative hypothesis* $\mathbf{H}_1$ *at significance level* 5%.

21.3 The normal approximation to Wilcoxon's test for one set of observations

Table B.8 includes values up to $n = 50$. If the number of observations is greater, a normal distribution approximation can be applied. Indeed, if the null hypothesis is true, the statistic t_+ is approximately normally distributed with expected value

$$\mu = \frac{n(n+1)}{4}$$

and standard deviation

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} \, .$$

The significance probability is therefore found by comparison of the statistic

$$z = \frac{t_+ - \mu}{\sigma}$$

with Table B.2 of the standard normal distribution.

EXAMPLE. Let us use the normal approximation to find the significance probability in the previous example (even though n here is smaller than 50 and the approximation therefore is not highly precise). We get $\mu = 27.5$ and $\sigma = 9.81$. The statistic therefore becomes $z = 1.89$, which gives a significance probability of 2.9%. The conclusion is thus the same, namely that $\mathbf{H}_0$ is rejected at significance level 5%.

21.4 Wilcoxon's test for two sets of observations

Suppose we have two sets $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ of independent observations. We test the null hypothesis

$\mathbf{H}_0$: *The observations come from the same distribution.*

Each of the $n + m$ observations is assigned a **rank** which is one of the numbers $1, 2, \ldots, n + m$. The observation with the smallest numerical value is assigned rank 1, the observation with the second smallest numerical value is assigned rank 2, etc. Define the statistic

$t_x = \sum(\text{ranks of the } x_i\text{'s})$.

Whether $\mathbf{H}_0$ is rejected or not depends on which alternative hypothesis we test against.

If we test $\mathbf{H}_0$ against the alternative hypothesis

$\mathbf{H}_1$: *The x_i's are primarily smaller than the y_i's,*

then $\mathbf{H}_0$ is rejected if t_x is extremely small. Look up in Table B.9 under n and m. If t_x is smaller than or equal to the table value, then $\mathbf{H}_0$ is rejected at significance level $\alpha = 5\%$. If t_x is greater than the table value, then $\mathbf{H}_0$ is accepted at significance level $\alpha = 5\%$.

If we test $\mathbf{H}_0$ against the alternative hypothesis

$\mathbf{H}_1$: *The x_i's are primarily greater than the y_i's,*

then one has to switch the roles of x_i's and y_i's and continue as described above.

If we don't test $\mathbf{H}_0$ against any particular alternative hypothesis, then the null hypothesis is rejected if the minimum

$$t := \min\{t_x, n(n+m+1) - t_x\}$$

is extremely small. Look up in Table B.9 under n and m. If t is smaller than or equal to the table value, then $\mathbf{H}_0$ is rejected at significance level $\alpha = 10\%$. If t is greater than the table value, then $\mathbf{H}_0$ is accepted at significance level 10%.

21.5 The normal approximation to Wilcoxon's test for two sets of observations

Table B.9 applies for moderate values of n and m. If the number of observations is greater, one can use a normal distribution approximation. Indeed, if the null hypothesis holds true, the statistic

t_x is approximately normally distributed with expected value

$$\mu = \frac{n(n+m+1)}{2}$$

and standard deviation

$$\sigma = \sqrt{\frac{nm(n+m+1)}{12}}\ .$$

The significance probability is then found by comparing the statistic

$$z = \frac{t_x - \mu}{\sigma}$$

with Table B.2 of the standard normal distribution.

22 Linear regression

22.1 The model

Suppose we have a sample consisting of n pairs of observations

$$(x_1, y_1),\ (x_2, y_2), \ldots,\ (x_n, y_n)\ .$$

We propose the **model** that each y_i is an observation from a random variable

$$Y_i = \beta_0 + \beta_1 x_i + E_i$$

where the E_i's are independent normally distributed random variables with expected value 0 and common variance σ^2. Thus we can express each y_i as

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

where e_i is an observation from E_i. We call y_i the **response variable**, x_i the **declaring variable** and e_i the **remainder term**.

22.2 Estimation of the parameters β_0 and β_1

Let $\bar{x}$ be the mean value of the x_i's and $\bar{y}$ the mean value of the y_i's. Define the **sum of products of errors** as

$$SPE_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

and the **sum of squares of errors** as

$$SSE_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

The parameters β_0 and β_1 of the regression equation are now estimated as

$$\begin{cases} \hat{\beta}_1 = \dfrac{SPE_{xy}}{SSE_x} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{cases}$$

22.3 The distribution of the estimators

If the model's assumptions are met, the estimator $\hat{\beta}_0$ is normally distributed with expected value β_0 (the estimator thus is unbiased) and variance $\sigma^2(1/n + \bar{x}^2/SSE_x)$. In other words, it holds that

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SSE_x}\right)\right) .$$

Moreover, the estimator $\hat{\beta}_1$ is normally distributed with expected value β_1 (this estimator is therefore unbiased too) and variance σ^2/SSE_x. In other words, it holds that

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{SSE_x}\right) .$$

22.4 Predicted values $\hat{y}_i$ and residuals $\hat{e}_i$

From the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, the **predicted value** of y_i can be computed for each i as

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i .$$

The i'th **residual** $\hat{e}_i$ is the difference between the *actual value* y_i and the *predicted value* $\hat{y}_i$:

$$\hat{e}_i = y_i - \hat{y}_i .$$

The residual $\hat{e}_i$ is an estimate of the remainder term e_i.

22.5 Estimation of the variance σ^2

We introduce the **sum of squares of residuals** as

$$SSR = \sum_{i=1}^{n} \hat{e}_i^2 .$$

The variance σ^2 of the remainder terms is now estimated as

$$s^2 = \frac{SSR}{n-2} .$$

This estimator is unbiased (but different from the maximum likelihood estimator).

22.6 Confidence intervals for the parameters β_0 and β_1

After estimating the parameters β_0 and β_1, we can compute the confidence intervals with confidence level $1-\alpha$ around the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. These are

$$\begin{cases} \hat{\beta}_0 \pm t_{1-\alpha/2} s \sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{SSE_x}} \\ \hat{\beta}_1 \pm t_{1-\alpha/2} \dfrac{s}{\sqrt{SSE_x}} \end{cases}$$

The number $t_{1-\alpha/2}$ is determined by $F(u_{1-\alpha/2}) = 1 - \alpha/2$, where F is the distribution function of Student's t distribution with $n-1$ degrees of freedom (see also section 15.8).

22.7 The determination coefficient R^2

In order to investigate how well the model with the estimated parameters describes the actual observations, we compute the **determination coefficient**

$$R^2 = \frac{SSE_y - SSR}{SSE_y} \, .$$

R^2 lies in the interval $[0, 1]$ and measures the part of the variation of the y_i's which the model describes as a linear function of the x_i's.

Remember: The greater the determination coefficient R^2 is, the better the model describes the observations.

22.8 Predictions and prediction intervals

Let there be given a real number x_0. The function value

$$y_0 = \beta_0 + \beta_1 x_0$$

is then estimated, or **predicted**, as

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \, .$$

The confidence interval, or **prediction interval**, with confidence level $1-\alpha$ around the estimate $\hat{y}_0$ is

$$\hat{y}_0 \pm t_{1-\alpha/2} s \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{SSE_x}} \,.$$

The number $t_{1-\alpha/2}$ is determined by $F(u_{1-\alpha/2}) = 1-\alpha/2$, where F is the distribution function of Student's t distribution with $n-2$ degrees of freedom (see also section 15.8).

22.9 Overview of formulae

$S_x = \sum_{i=1}^n x_i$	The sum of the x_i's
$\bar{x} = S_x/n$	The mean value of the x_i's
$SS_x = \sum_{i=1}^n x_i^2$	The sum of the squares of the x_i's
$SSE_x = \sum_{i=1}^n (x_i-\bar{x})^2 = SS_x - S_x^2/n$	The sum of the squares of the errors
$s_x^2 = SSE_x/(n-1)$	Empirical variance of the x_i's
$SP_{xy} = \sum_{i=1}^n x_i y_i$	The sum of the products
$SPE_{xy} = \sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y}) = SP_{xy} - S_x S_y/n$	The sum of the products of the errors
$\hat{\beta}_1 = SPE_{xy}/SSE_x$	The estimate of β_1
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$	The estimate of β_0
$\hat{y}_i = \hat{be}_0 + \hat{\beta}_1 x_i$	Predicted value of y_i
$\hat{e}_i = y_i - \hat{y}_i$	The i'th residual
$SSR = \sum_{i=1}^n \hat{e}_i^2 = SSE_y - SPE_{xy}^2/SSE_x$	The sum of the squares of the residuals
$s^2 = SSR/(n-2)$	The estimate σ^2
$R^2 = 1 - SSR/SSE_y$	The determination coefficient

22.10 Example

EXERCISE. It is claimed that the temperature in the Andes Mountains decreases by six degrees per 1000 metres. The following temperatures were measured simultaneously at ten different localities in the same region:

Altitude x_i (metres)	Temperature y_i (degrees)
500	15
1000	14
1500	11
2000	6
2500	–1
3000	2
3500	0
4000	–4
4500	–8
5000	–14

We use a linear regression model

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

where the remainder terms e_i are independent normally distributed with expected value 0 and the same (unknown) variance σ^2.

1) Estimate the parameters β_0 and β_1.

2) Determine the confidence interval with confidence level 95% for β_1.

3) Can the hypothesis $\mathbf{H}_0 : \beta_1 = -0.006$ be accepted?

4) To how large degree can the difference of temperature be explained as a linear function of the altitude?

SOLUTION. First we perform the relevant computations:

$S_x = \sum_{i=1}^{10} x_i = 27500$ $\qquad$ $S_y = \sum_{i=1}^{10} y_i = 21$

$\bar{x} = S_x/10 = 2750$ $\qquad$ $\bar{y} = S_y/10 = 2.1$

$SS_x = \sum_{i=1}^{10} x_i^2 = 96250000$ $\qquad$ $SS_y = \sum_{i=1}^{10} y_i^2 = 859$

$SSE_x = SS_x - S_x^2/10 = 20625000$ $\qquad$ $SSE_y = SS_y - S_y^2/10 = 814.9$

$SP_{xy} = \sum_{i=1}^{10} x_i y_i = -68500$ $\qquad$ $SPE_{xy} = SP_{xy} - S_x S_y/10 = -126250$

$\hat{\beta}_1 = SPE_{xy}/SSE_x = -0.0061$ $\qquad$ $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 18.9$

$SSR = SSE_y - SPE_{xy}^2/SSE_x = 42.1$ $\qquad$ $s^2 = SSR/8 = 5.26$

$R^2 = 1 - SSR/SSE_y = 0.948$

1) It appears directly from the computations that the estimates of β_0 and β_1 are

$$\hat{\beta}_0 = 18.9 \ , \ \hat{\beta}_1 = -0.0061 \ .$$

2) Table B.4, under $df = 10 - 1 = 9$ degrees of freedom, shows that $t_{0.975} = 2.26$ (see also section 15.8). The confidence interval around $\hat{\beta}_1$ thus becomes

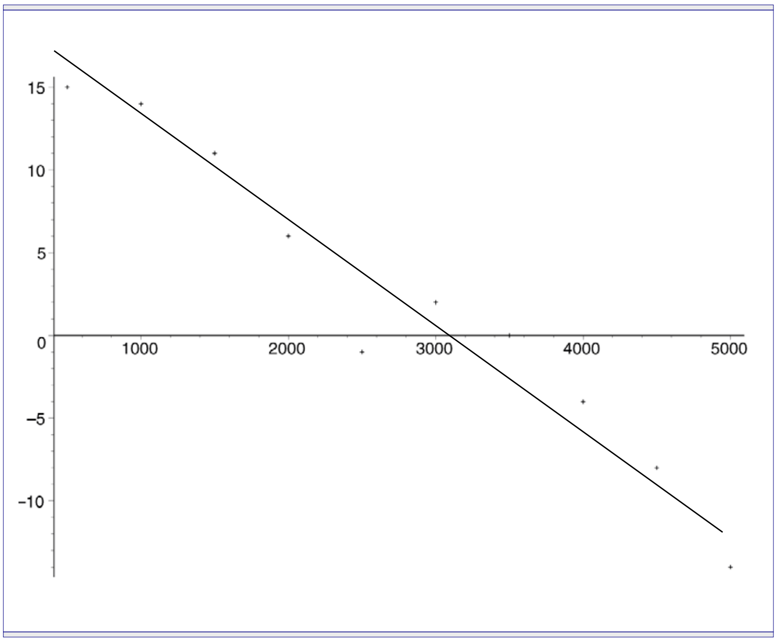
$$\left[-0.0061 - 2.26\frac{\sqrt{5.26}}{\sqrt{20625000}} \, , \, -0.0061 + 2.26\frac{\sqrt{5.26}}{\sqrt{20625000}} \right] = [-0.0072 \, , \, -0.0050] \, .$$

3) The hypothesis $\mathbf{H}_0 : \beta_1 = -0.006$ is accepted, since this value lies within the confidence interval.

4) The part of the temperature difference describable as a linear function of the altitude is nothing other than the determination coefficient

$$R^2 = 94.8\% \, .$$

The fact that R^2 is large (close to 100%) shows that the actual temperatures are quite close to those predicted. This also appears from the figure below, which shows that the actual temperatures are very close to the regression line:


15
10
5
0
−5
−10
1000
2000
3000
4000
5000

A Overview of discrete distributions

Distribution	Description	Values	Point probabilities	Mean value	Variance
Binomial distribution Bin(n,p)	Number of successes in n tries	$k=0,1,\dots,n$	$\binom{n}{k} p^k q^{n-k}$	np	npq
Poisson distribution Pois(λ)	Number of spontaneous events in a time interval	$k=0,1,\dots$	$\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ
Geometrical distribution Geo(p)	Number of failures before success	$k=0,1,\dots$	$q^k p$	q/p	q/p^2
Hyper-geometrical distribution HG(n,r,N)	Number of red balls among n balls	$k=0,\dots$ $\dots,\min\{n,r\}$	$\frac{\binom{r}{k}\binom{s}{n-k}}{\binom{N}{n}}$	nr/N	$\frac{nrs(N-n)}{N^2(N-1)}$
Negative binomial distribution NB(n,p)	Number of failures before the n'th success	$k=0,1,\dots$	$\binom{n+k-1}{n-1}\cdot p^n\cdot q^k$	nq/p	nq/p^2
Multi-nomial-distribution Mult$(n,\dots)$	Number of sample points of each type	$(k_1,\dots,k_r)$ where $\sum k_i = n$	$\binom{n}{k_1\cdots k_r}\cdot\prod p_i^{k_i}$	—	—

B Tables

B.1 How to read the tables

Table B.2 gives values of the distribution function

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \, \exp\left(-\frac{1}{2}t^2\right) dt$$

of the standard normal distribution.

Table B.3 gives values of x for which the distribution function $F = F_{\chi^2}$ of the χ^2 distribution with df degrees of freedom takes the values $F(x) = 0.500$, $F(x) = 0.600$, etc.

Table B.4 gives values of x for which the distribution function $F = F_{\mathrm{Student}}$ of Student's t distribution with df degrees of freedom takes the values $F(x) = 0.600$, $F(x) = 0.700$, etc.

Table B.5, Table B.6 and **Table B.7** give values of x for which the distribution function $F = F_{\mathrm{Fisher}}$ of Fisher's F distribution with n degrees of freedom in the numerator (top line) and m degrees of freedom in the denominator (leftmost column) takes the values $F(x) = 0.90$, $F(x) = 0.95$, and $F(x) = 0.99$, respectively.

Table B.8 and **Table B.9** give critical values for Wilcoxon's tests for one and two sets of observations. See Chapter 21 for further details.

B.2 The standard normal distribution

u	Φ(u)	Φ(−u)	u	Φ(u)	Φ(−u)	u	Φ(u)	Φ(−u)
0.00	0.5000	0.5000	0.36	0.6406	0.3594	0.72	0.7642	0.2358
0.01	0.5040	0.4960	0.37	0.6443	0.3557	0.73	0.7673	0.2327
0.02	0.5080	0.4920	0.38	0.6480	0.3520	0.74	0.7704	0.2296
0.03	0.5120	0.4880	0.39	0.6517	0.3483	0.75	0.7734	0.2266
0.04	0.5160	0.4840	0.40	0.6554	0.3446	0.76	0.7764	0.2236
0.05	0.5199	0.4801	0.41	0.6591	0.3409	0.77	0.7794	0.2206
0.06	0.5239	0.4761	0.42	0.6628	0.3372	0.78	0.7823	0.2177
0.07	0.5279	0.4721	0.43	0.6664	0.3336	0.79	0.7852	0.2148
0.08	0.5319	0.4681	0.44	0.6700	0.3300	0.80	0.7881	0.2119
0.09	0.5359	0.4641	0.45	0.6736	0.3264	0.81	0.7910	0.2090
0.10	0.5398	0.4602	0.46	0.6772	0.3228	0.82	0.7939	0.2061
0.11	0.5438	0.4562	0.47	0.6808	0.3192	0.83	0.7967	0.2033
0.12	0.5478	0.4522	0.48	0.6844	0.3156	0.84	0.7995	0.2005
0.13	0.5517	0.4483	0.49	0.6879	0.3121	0.85	0.8023	0.1977
0.14	0.5557	0.4443	0.50	0.6915	0.3085	0.86	0.8051	0.1949
0.15	0.5596	0.4404	0.51	0.6950	0.3050	0.87	0.8078	0.1922
0.16	0.5636	0.4364	0.52	0.6985	0.3015	0.88	0.8106	0.1894
0.17	0.5675	0.4325	0.53	0.7019	0.2981	0.89	0.8133	0.1867
0.18	0.5714	0.4286	0.54	0.7054	0.2946	0.90	0.8159	0.1841
0.19	0.5753	0.4247	0.55	0.7088	0.2912	0.91	0.8186	0.1814
0.20	0.5793	0.4207	0.56	0.7123	0.2877	0.92	0.8212	0.1788
0.21	0.5832	0.4168	0.57	0.7157	0.2843	0.93	0.8238	0.1762
0.22	0.5871	0.4129	0.58	0.7190	0.2810	0.94	0.8264	0.1736
0.23	0.5910	0.4090	0.59	0.7224	0.2776	0.95	0.8289	0.1711
0.24	0.5948	0.4052	0.60	0.7257	0.2743	0.96	0.8315	0.1685
0.25	0.5987	0.4013	0.61	0.7291	0.2709	0.97	0.8340	0.1660
0.26	0.6026	0.3974	0.62	0.7324	0.2676	0.98	0.8365	0.1635
0.27	0.6064	0.3936	0.63	0.7357	0.2643	0.99	0.8389	0.1611
0.28	0.6103	0.3897	0.64	0.7389	0.2611	1.00	0.8413	0.1587
0.29	0.6141	0.3859	0.65	0.7422	0.2578	1.01	0.8438	0.1562
0.30	0.6179	0.3821	0.66	0.7454	0.2546	1.02	0.8461	0.1539
0.31	0.6217	0.3783	0.67	0.7486	0.2514	1.03	0.8485	0.1515
0.32	0.6255	0.3745	0.68	0.7517	0.2483	1.04	0.8508	0.1492
0.33	0.6293	0.3707	0.69	0.7549	0.2451	1.05	0.8531	0.1469
0.34	0.6331	0.3669	0.70	0.7580	0.2420	1.06	0.8554	0.1446
0.35	0.6368	0.3632	0.71	0.7611	0.2389	1.07	0.8577	0.1423

$\mathbf{u}$	$\mathbf{\Phi(u)}$	$\mathbf{\Phi(-u)}$
1.08	0.8599	0.1401
1.09	0.8621	0.1379
1.10	0.8643	0.1357
1.11	0.8665	0.1335
1.12	0.8686	0.1314
1.13	0.8708	0.1292
1.14	0.8729	0.1271
1.15	0.8749	0.1251
1.16	0.8770	0.1230
1.17	0.8790	0.1210
1.18	0.8810	0.1190
1.19	0.8830	0.1170
1.20	0.8849	0.1151
1.21	0.8869	0.1131
1.22	0.8888	0.1112
1.23	0.8907	0.1093
1.24	0.8925	0.1075
1.25	0.8944	0.1056
1.26	0.8962	0.1038
1.27	0.8980	0.1020
1.28	0.8997	0.1003
1.29	0.9015	0.0985
1.30	0.9032	0.0968
1.31	0.9049	0.0951
1.32	0.9066	0.0934
1.33	0.9082	0.0918
1.34	0.9099	0.0901
1.35	0.9115	0.0885
1.36	0.9131	0.0869
1.37	0.9147	0.0853
1.38	0.9162	0.0838
1.39	0.9177	0.0823
1.40	0.9192	0.0808
1.41	0.9207	0.0793
1.42	0.9222	0.0778
1.43	0.9236	0.0764
1.44	0.9251	0.0749

$\mathbf{u}$	$\mathbf{\Phi(u)}$	$\mathbf{\Phi(-u)}$
1.45	0.9265	0.0735
1.46	0.9279	0.0721
1.47	0.9292	0.0708
1.48	0.9306	0.0694
1.49	0.9319	0.0681
1.50	0.9332	0.0668
1.51	0.9345	0.0655
1.52	0.9357	0.0643
1.53	0.9370	0.0630
1.54	0.9382	0.0618
1.55	0.9394	0.0606
1.56	0.9406	0.0594
1.57	0.9418	0.0582
1.58	0.9429	0.0571
1.59	0.9441	0.0559
1.60	0.9452	0.0548
1.61	0.9463	0.0537
1.62	0.9474	0.0526
1.63	0.9484	0.0516
1.64	0.9495	0.0505
1.65	0.9505	0.0495
1.66	0.9515	0.0485
1.67	0.9525	0.0475
1.68	0.9535	0.0465
1.69	0.9545	0.0455
1.70	0.9554	0.0446
1.71	0.9564	0.0436
1.72	0.9573	0.0427
1.73	0.9582	0.0418
1.74	0.9591	0.0409
1.75	0.9599	0.0401
1.76	0.9608	0.0392
1.77	0.9616	0.0384
1.78	0.9625	0.0375
1.79	0.9633	0.0367
1.80	0.9641	0.0359
1.81	0.9649	0.0351

$\mathbf{u}$	$\mathbf{\Phi(u)}$	$\mathbf{\Phi(-u)}$
1.82	0.9656	0.0344
1.83	0.9664	0.0336
1.84	0.9671	0.0329
1.85	0.9678	0.0322
1.86	0.9686	0.0314
1.87	0.9693	0.0307
1.88	0.9699	0.0301
1.89	0.9706	0.0294
1.90	0.9713	0.0287
1.91	0.9719	0.0281
1.92	0.9726	0.0274
1.93	0.9732	0.0268
1.94	0.9738	0.0262
1.95	0.9744	0.0256
1.96	0.9750	0.0250
1.97	0.9756	0.0244
1.98	0.9761	0.0239
1.99	0.9767	0.0233
2.00	0.9772	0.0228
2.01	0.9778	0.0222
2.02	0.9783	0.0217
2.03	0.9788	0.0212
2.04	0.9793	0.0207
2.05	0.9798	0.0202
2.06	0.9803	0.0197
2.07	0.9808	0.0192
2.08	0.9812	0.0188
2.09	0.9817	0.0183
2.10	0.9821	0.0179
2.11	0.9826	0.0174
2.12	0.9830	0.0170
2.13	0.9834	0.0166
2.14	0.9838	0.0162
2.15	0.9842	0.0158
2.16	0.9846	0.0154
2.17	0.9850	0.0150
2.18	0.9854	0.0146

$\mathbf{u}$	$\mathbf{\Phi(u)}$	$\mathbf{\Phi(-u)}$	$\mathbf{u}$	$\mathbf{\Phi(u)}$	$\mathbf{\Phi(-u)}$	$\mathbf{u}$	$\mathbf{\Phi(u)}$	$\mathbf{\Phi(-u)}$
2.19	0.9857	0.0143	2.56	0.9948	0.0052	2.93	0.9983	0.0017
2.20	0.9861	0.0139	2.57	0.9949	0.0051	2.94	0.9984	0.0016
2.21	0.9864	0.0136	2.58	0.9951	0.0049	2.95	0.9984	0.0016
2.22	0.9868	0.0132	2.59	0.9952	0.0048	2.96	0.9985	0.0015
2.23	0.9871	0.0129	2.60	0.9953	0.0047	2.97	0.9985	0.0015
2.24	0.9875	0.0125	2.61	0.9955	0.0045	2.98	0.9986	0.0014
2.25	0.9878	0.0122	2.62	0.9956	0.0044	2.99	0.9986	0.0014
2.26	0.9881	0.0119	2.63	0.9957	0.0043	3.00	0.9987	0.0013
2.27	0.9884	0.0116	2.64	0.9959	0.0041	3.10	0.9990	0.0010
2.28	0.9887	0.0113	2.65	0.9960	0.0040	3.20	0.9993	0.0007
2.29	0.9890	0.0110	2.66	0.9961	0.0039	3.30	0.9995	0.0005
2.30	0.9893	0.0107	2.67	0.9962	0.0038	3.40	0.9997	0.0003
2.31	0.9896	0.0104	2.68	0.9963	0.0037	3.50	0.9998	0.0002
2.32	0.9898	0.0102	2.69	0.9964	0.0036	3.60	0.9998	0.0002
2.33	0.9901	0.0099	2.70	0.9965	0.0035	3.70	0.9999	0.0001
2.34	0.9904	0.0096	2.71	0.9966	0.0034	3.80	0.9999	0.0001
2.35	0.9906	0.0094	2.72	0.9967	0.0033	3.90	1.0000	0.0000
2.36	0.9909	0.0091	2.73	0.9968	0.0032	4.00	1.0000	0.0000
2.37	0.9911	0.0089	2.74	0.9969	0.0031			
2.38	0.9913	0.0087	2.75	0.9970	0.0030			
2.39	0.9916	0.0084	2.76	0.9971	0.0029			
2.40	0.9918	0.0082	2.77	0.9972	0.0028			
2.41	0.9920	0.0080	2.78	0.9973	0.0027			
2.42	0.9922	0.0078	2.79	0.9974	0.0026			
2.43	0.9925	0.0075	2.80	0.9974	0.0026			
2.44	0.9927	0.0073	2.81	0.9975	0.0025			
2.45	0.9929	0.0071	2.82	0.9976	0.0024			
2.46	0.9931	0.0069	2.83	0.9977	0.0023			
2.47	0.9932	0.0068	2.84	0.9977	0.0023			
2.48	0.9934	0.0066	2.85	0.9978	0.0022			
2.49	0.9936	0.0064	2.86	0.9979	0.0021			
2.50	0.9938	0.0062	2.87	0.9979	0.0021			
2.51	0.9940	0.0060	2.88	0.9980	0.0020			
2.52	0.9941	0.0059	2.89	0.9981	0.0019			
2.53	0.9943	0.0057	2.90	0.9981	0.0019			
2.54	0.9945	0.0055	2.91	0.9982	0.0018			
2.55	0.9946	0.0054	2.92	0.9982	0.0018			

B.3 The χ^2 distribution (values x with $F_{\chi^2}(x) = 0.500$ etc.)

df	0.500	0.600	0.700	0.800	0.900	0.950	0.975	0.990	0.995	0.999
1	0.45	0.71	1.07	1.64	2.71	3.84	5.02	6.63	7.88	10.83
2	1.39	1.83	2.41	3.22	4.61	5.99	7.38	9.21	10.60	13.82
3	2.37	2.95	3.66	4.64	6.25	7.81	9.35	11.34	12.84	16.27
4	3.36	4.04	4.88	5.99	7.78	9.49	11.14	13.28	14.86	18.47
5	4.35	5.13	6.06	7.29	9.24	11.07	12.83	15.09	16.75	20.52
6	5.35	6.21	7.23	8.56	10.64	12.59	14.45	16.81	18.55	22.46
7	6.35	7.28	8.38	9.80	12.02	14.07	16.01	18.48	20.28	24.32
8	7.34	8.35	9.52	11.03	13.36	15.51	17.53	20.09	21.95	26.12
9	8.34	9.41	10.66	12.24	14.68	16.92	19.02	21.67	23.59	27.88
10	9.34	10.47	11.78	13.44	15.99	18.31	20.48	23.21	25.19	29.59
11	10.34	11.53	12.90	14.63	17.28	19.68	21.92	24.72	26.76	31.26
12	11.34	12.58	14.01	15.81	18.55	21.03	23.34	26.22	28.30	32.91
13	12.34	13.64	15.12	16.98	19.81	22.36	24.74	27.69	29.82	34.53
14	13.34	14.69	16.22	18.15	21.06	23.68	26.12	29.14	31.32	36.12
15	14.34	15.73	17.32	19.31	22.31	25.00	27.49	30.58	32.80	37.70
16	15.34	16.78	18.42	20.47	23.54	26.30	28.85	32.00	34.27	39.25
17	16.34	17.82	19.51	21.61	24.77	27.59	30.19	33.41	35.72	40.79
18	17.34	18.87	20.60	22.76	25.99	28.87	31.53	34.81	37.16	42.31
19	18.34	19.91	21.69	23.90	27.20	30.14	32.85	36.19	38.58	43.82
20	19.34	20.95	22.77	25.04	28.41	31.41	34.17	37.57	40.00	45.31
21	20.34	21.99	23.86	26.17	29.62	32.67	35.48	38.93	41.40	46.80
22	21.34	23.03	24.94	27.30	30.81	33.92	36.78	40.29	42.80	48.27
23	22.34	24.07	26.02	28.43	32.01	35.17	38.08	41.64	44.18	49.73
24	23.34	25.11	27.10	29.55	33.20	36.42	39.36	42.98	45.56	51.18
25	24.34	26.14	28.17	30.68	34.38	37.65	40.65	44.31	46.93	52.62
26	25.34	27.18	29.25	31.79	35.56	38.89	41.92	45.64	48.29	54.05
27	26.34	28.21	30.32	32.91	36.74	40.11	43.19	46.96	49.64	55.48
28	27.34	29.25	31.39	34.03	37.92	41.34	44.46	48.28	50.99	56.89
29	28.34	30.28	32.46	35.14	39.09	42.56	45.72	49.59	52.34	58.30
30	29.34	31.32	33.53	36.25	40.26	43.77	46.98	50.89	53.67	59.70
31	30.34	32.35	34.60	37.36	41.42	44.99	48.23	52.19	55.00	61.10
32	31.34	33.38	35.66	38.47	42.58	46.19	49.48	53.49	56.33	62.49
33	32.34	34.41	36.73	39.57	43.75	47.40	50.73	54.78	57.65	63.87
34	33.34	35.44	37.80	40.68	44.90	48.60	51.97	56.06	58.96	65.25
35	34.34	36.47	38.86	41.78	46.06	49.80	53.20	57.34	60.27	66.62

df	0.500	0.600	0.700	0.800	0.900	0.950	0.975	0.990	0.995	0.999
36	35.34	37.50	39.92	42.88	47.21	51.00	54.44	58.62	61.58	67.99
37	36.34	38.53	40.98	43.98	48.36	52.19	55.67	59.89	62.88	69.35
38	37.34	39.56	42.05	45.08	49.51	53.38	56.90	61.16	64.18	70.70
39	38.34	40.59	43.11	46.17	50.66	54.57	58.12	62.43	65.48	72.05
40	39.34	41.62	44.16	47.27	51.81	55.76	59.34	63.69	66.77	73.40
41	40.34	42.65	45.22	48.36	52.95	56.94	60.56	64.95	68.05	74.74
42	41.34	43.68	46.28	49.46	54.09	58.12	61.78	66.21	69.34	76.08
43	42.34	44.71	47.34	50.55	55.23	59.30	62.99	67.46	70.62	77.42
44	43.34	45.73	48.40	51.64	56.37	60.48	64.20	68.71	71.89	78.75
45	44.34	46.76	49.45	52.73	57.51	61.66	65.41	69.96	73.17	80.08
46	45.34	47.79	50.51	53.82	58.64	62.83	66.62	71.20	74.44	81.40
47	46.34	48.81	51.56	54.91	59.77	64.00	67.82	72.44	75.70	82.72
48	47.34	49.84	52.62	55.99	60.91	65.17	69.02	73.68	76.97	84.04
49	48.33	50.87	53.67	57.08	62.04	66.34	70.22	74.92	78.23	85.35
50	49.33	51.89	54.72	58.16	63.17	67.50	71.42	76.15	79.49	86.66
51	50.33	52.92	55.78	59.25	64.30	68.67	72.62	77.39	80.75	87.97
52	51.33	53.94	56.83	60.33	65.42	69.83	73.81	78.62	82.00	89.27
53	52.33	54.97	57.88	61.41	66.55	70.99	75.00	79.84	83.25	90.57
54	53.33	55.99	58.93	62.50	67.67	72.15	76.19	81.07	84.50	91.87
55	54.33	57.02	59.98	63.58	68.80	73.31	77.38	82.29	85.75	93.17
56	55.33	58.04	61.03	64.66	69.92	74.47	78.57	83.51	86.99	94.46
57	56.33	59.06	62.08	65.74	71.04	75.62	79.75	84.73	88.24	95.75
58	57.33	60.09	63.13	66.82	72.16	76.78	80.94	85.95	89.48	97.04
59	58.33	61.11	64.18	67.89	73.28	77.93	82.12	87.17	90.72	98.32
60	59.33	62.13	65.23	68.97	74.40	79.08	83.30	88.38	91.95	99.61
61	60.33	63.16	66.27	70.05	75.51	80.23	84.48	89.59	93.19	100.89
62	61.33	64.18	67.32	71.13	76.63	81.38	85.65	90.80	94.42	102.17
63	62.33	65.20	68.37	72.20	77.75	82.53	86.83	92.01	95.65	103.44
64	63.33	66.23	69.42	73.28	78.86	83.68	88.00	93.22	96.88	104.72
65	64.33	67.25	70.46	74.35	79.97	84.82	89.18	94.42	98.11	105.99
66	65.33	68.27	71.51	75.42	81.09	85.96	90.35	95.63	99.33	107.26
67	66.33	69.29	72.55	76.50	82.20	87.11	91.52	96.83	100.55	108.53
68	67.33	70.32	73.60	77.57	83.31	88.25	92.69	98.03	101.78	109.79
69	68.33	71.34	74.64	78.64	84.42	89.39	93.86	99.23	103.00	111.06
70	69.33	72.36	75.69	79.71	85.53	90.53	95.02	100.43	104.21	112.32

B.4 Student's t distribution (values x with $F_{\mathrm{Student}}(x) = 0.600$ etc.)

df	0.600	0.700	0.800	0.900	0.950	0.975	0.990	0.995	0.999
1	0.32	0.73	1.38	3.08	6.31	12.71	31.82	63.66	318.31
2	0.29	0.62	1.06	1.89	2.92	4.30	6.96	9.92	22.33
3	0.28	0.58	0.98	1.64	2.35	3.18	4.54	5.84	10.2
4	0.27	0.57	0.94	1.53	2.13	2.78	3.75	4.60	7.17
5	0.27	0.56	0.92	1.48	2.02	2.57	3.36	4.03	5.89
6	0.26	0.55	0.91	1.44	1.94	2.45	3.14	3.71	5.21
7	0.26	0.55	0.90	1.41	1.89	2.36	3.00	3.50	4.79
8	0.26	0.55	0.89	1.40	1.86	2.31	2.90	3.36	4.50
9	0.26	0.54	0.88	1.38	1.83	2.26	2.82	3.25	4.30
10	0.26	0.54	0.88	1.37	1.81	2.23	2.76	3.17	4.14
11	0.26	0.54	0.88	1.36	1.80	2.20	2.72	3.11	4.02
12	0.26	0.54	0.87	1.36	1.78	2.18	2.68	3.05	3.93
13	0.26	0.54	0.87	1.35	1.77	2.16	2.65	3.01	3.85
14	0.26	0.54	0.87	1.35	1.76	2.14	2.62	2.98	3.79
15	0.26	0.54	0.87	1.34	1.75	2.13	2.60	2.95	3.73
16	0.26	0.54	0.86	1.34	1.75	2.12	2.58	2.92	3.69
17	0.26	0.53	0.86	1.33	1.74	2.11	2.57	2.90	3.65
18	0.26	0.53	0.86	1.33	1.73	2.10	2.55	2.88	3.61
19	0.26	0.53	0.86	1.33	1.73	2.09	2.54	2.86	3.58
20	0.26	0.53	0.86	1.33	1.72	2.09	2.53	2.85	3.55
21	0.26	0.53	0.86	1.32	1.72	2.08	2.52	2.83	3.53
22	0.26	0.53	0.86	1.32	1.72	2.07	2.51	2.82	3.50
23	0.26	0.53	0.86	1.32	1.71	2.07	2.50	2.81	3.48
24	0.26	0.53	0.86	1.32	1.71	2.06	2.49	2.80	3.47
25	0.26	0.53	0.86	1.32	1.71	2.06	2.49	2.79	3.45
26	0.26	0.53	0.86	1.31	1.71	2.06	2.48	2.78	3.43
27	0.26	0.53	0.86	1.31	1.70	2.05	2.47	2.77	3.42
28	0.26	0.53	0.85	1.31	1.70	2.05	2.47	2.76	3.41
29	0.26	0.53	0.85	1.31	1.70	2.05	2.46	2.76	3.40
30	0.26	0.53	0.85	1.31	1.70	2.04	2.46	2.75	3.39
35	0.26	0.53	0.85	1.31	1.69	2.03	2.44	2.72	3.34
40	0.26	0.53	0.85	1.30	1.68	2.02	2.42	2.70	3.31
50	0.25	0.53	0.85	1.30	1.68	2.01	2.40	2.68	3.26
100	0.25	0.53	0.85	1.29	1.66	1.98	2.36	2.63	3.17
∞	0.25	0.52	0.84	1.28	1.64	1.96	2.33	2.58	3.09

B.5 Fisher's F distribution (values x with $F_{\text{Fisher}}(x) = 0.90$)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.02	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	3.01	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	3.00	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.98	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.97	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.96	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.95	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.94	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.93	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.92	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.92	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.91	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.90	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
31	2.90	2.48	2.27	2.14	2.04	1.97	1.92	1.88	1.84	1.81
32	2.89	2.48	2.26	2.13	2.04	1.97	1.91	1.87	0.84	1.81
33	2.89	2.47	2.26	2.12	2.03	1.96	1.91	1.86	1.83	1.80
34	2.88	2.47	2.25	2.12	2.02	1.96	1.90	1.86	1.82	1.79
35	2.88	2.46	2.25	2.11	2.02	1.95	1.90	1.85	1.82	1.79

B.6 Fisher's F distribution (values x with $F_{\text{Fisher}}(x) = 0.95$)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.43	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.41	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.38	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.35	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.33	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.31	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.29	3.40	3.01	2.78	2.62	2.51	2.42	2.36	4.62	2.25
25	4.27	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.25	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.24	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.22	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.21	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.20	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
31	4.18	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15
32	4.17	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14
33	4.16	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13
34	4.15	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12
35	4.15	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11

B.7 Fisher's F distribution (values x with $F_{\text{Fisher}}(x) = 0.99$)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.30	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.22	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.13	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.05	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.98	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.91	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.85	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.80	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.75	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.71	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.67	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.63	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.59	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
31	7.56	5.36	4.48	3.99	3.67	3.45	3.28	3.15	3.04	2.96
32	7.53	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93
33	7.50	5.31	4.44	3.95	3.63	3.41	3.24	3.11	3.00	2.91
34	7.47	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89
35	7.45	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88

B.8 Wilcoxon's test for one set of observations

n	0.005	0.010	0.025	0.050
5	–	–	–	0
6	–	–	0	2
7	–	0	2	3
8	0	1	3	5
9	1	3	5	8
10	3	5	8	10
11	5	7	10	13
12	7	9	13	17
13	9	12	17	21
14	12	15	21	25
15	15	19	25	30
16	19	23	29	35
17	23	27	34	41
18	27	32	40	47
19	32	37	46	53
20	37	43	52	60
21	42	49	58	67
22	48	55	65	75
23	54	62	73	83
24	61	69	81	91
25	68	76	89	100
26	75	84	98	110
27	83	92	107	119
28	91	101	116	130
29	100	110	126	140
30	109	120	137	151
31	118	130	147	163
32	128	140	159	175
33	138	151	170	187
34	148	162	182	200
35	159	173	195	213
36	171	185	208	227
37	182	198	221	241
38	194	211	235	256
39	207	224	249	271
40	220	238	264	286
41	233	252	279	302
42	247	266	294	319
43	261	281	310	336
44	276	296	327	353
45	291	312	343	371
46	307	328	361	389
47	322	345	378	407
48	339	362	396	426
49	355	379	415	446
50	373	397	434	466

B.9 Wilcoxon's test for two sets of observations, $\alpha = 5\%$

	m = 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n = 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	2	2	3	3	3	4	4	4	4	5	5	6	6
3	5	5	6	6	7	8	8	9	10	10	11	11	12	13	13
4	9	9	10	11	12	13	14	15	16	17	18	19	20	21	22
5	14	15	16	17	19	20	21	23	24	26	27	28	30	31	33
6	20	21	23	24	26	28	29	31	33	35	37	38	40	42	44
7	27	28	30	32	34	36	39	41	43	45	47	49	52	54	56
8	35	37	39	41	44	46	49	51	54	56	59	62	64	67	69
9	44	46	49	51	54	57	60	63	66	69	72	75	78	81	84
10	54	56	59	62	66	69	72	75	79	82	86	89	92	96	99
11	65	67	71	74	78	82	85	89	93	97	100	104	108	112	116
12	77	80	83	87	91	95	99	104	108	112	116	120	125	129	133
13	90	93	97	101	106	110	115	119	124	128	133	138	142	147	152
14	104	108	112	116	121	126	131	136	141	146	151	156	161	166	171
15	119	123	127	132	138	143	148	153	159	164	170	175	181	186	192
16	135	139	144	150	155	161	166	172	178	184	190	196	201	207	213
17	152	156	162	168	173	179	186	192	198	204	210	217	223	230	236
18	170	175	180	187	193	199	206	212	219	226	232	239	246	253	259
19	190	194	200	207	213	220	227	234	241	248	255	262	270	277	284
20	210	214	221	228	235	242	249	257	264	272	279	287	294	302	310
21	231	236	242	250	257	265	272	280	288	296	304	312	320	328	336
22	253	258	265	273	281	289	297	305	313	321	330	338	347	355	364
23	276	281	289	297	305	313	322	330	339	348	357	366	374	383	392
24	300	306	313	322	330	339	348	357	366	375	385	394	403	413	422
25	325	331	339	348	357	366	375	385	394	404	414	423	433	443	453

C Explanation of symbols

A, B, C	events
Ω	sample space
P	probability function, significance probability
$P(A\|B)$	conditional probability of A given B
$\cap$, $\cup$	intersection, union
$\wedge$, $\vee$	and, or
$A \subseteq \Omega$	A is a subset of Ω
$\omega \in \Omega$	ω belongs to Ω
$\complement A$	complement of the set A
$A \backslash B$	difference of the sets A and B ("A minus B")
$f : \Omega \to \mathbb{R}$	f is a map from Ω into $\mathbb{R}$
$:=$	equals by definition
$\|x\|$	absolute value of x (e.g. $\|-2\| = 2$)
$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	the set of natural, integral, real numbers
$\emptyset$	the empty set
$[0, \infty[$	the interval $\{x \in \mathbb{R} \mid x \geq 0\}$
X, Y	random variables
$E(X)$	the expected value of X
$\text{var}(X)$	the variance of X
$\text{Cov}(X, Y)$	the covariance of X and Y
μ	expected value
σ^2	variance
σ	standard deviation
Bin	binomial distribution
Pois	Poisson distribution
Geo	geometrical distribution
HG	hypergeometrical distribution
Mult	multinomial distribution
NB	negative binomial distribution
Exp	exponential distribution
N	normal distribution
s^2	empirical variance
s	empirical standard deviation
$F(x)$	distribution function
$f(x)$	density function
$\Phi(x)$	distribution function of standard normal distribution
$\varphi(x)$	density function of standard normal distribution
n	number of observations or tries
λ	intensity (in a Poisson process)
R^2	determination coefficient

ρ	correlation coefficient
$\bar{x}, \bar{y}$	mean value
df	number of degrees of freedom

Index