Essentials of Statistics: Exercises

David Brink

David Brink

Statistics – Exercises

Statistics – Exercises

ISBN 978-87-7681-409-0

Contents

1 Preface

This collection of *Problems with Solutions* is a companion to my book *Statistics*. All references here are to this compendium.

2 Problems for Chapter 2: Basic concepts of probability theory

Problem 1
A *poker hand* consists of five cards chosen randomly from an ordinary pack of 52 cards. How many different possible hands N are there?

Problem 2
What is the probability of having the poker hand *royal flush*, i.e. Ace, King, Queen, Jack, 10, all of the same suit?

Problem 3
What is the probability of having the poker hand *straight flush*, i.e. five cards in sequence, all of the same suit?

Problem 4
What is the probability of having the poker hand *four of a kind*, i.e. four cards of the same value (four aces, four 7s, etc.)?

Problem 5
What is the probability of having the poker hand *full house*, i.e. three of a kind plus two of a kind?

Problem 6
What is the probability of having the poker hand *flush*, i.e. five cards of the same suit?

Problem 7
What is the probability of having the poker hand *straight*, i.e. five cards in sequence?

Problem 8
What is the probability of having the poker hand *three of a kind*?

Problem 9
What is the probability of having the poker hand *two pair*?

Problem 10
What is the probability of having the poker hand *one pair*?

Problem 11
A red and a black die are thrown. What is the probability P of having at least ten? What is the conditional probability Q of having at least ten, given that the black die shows five? What is the conditional probability R of having at least ten, given that at least one of the dice shows five?

Problem 12
How many subsets with three elements are there of a set with ten elements? How many subsets

with seven elements are there of a set with ten elements?

Problem 13

In how many ways can a set with 30 elements be divided into three subsets with five, ten and fifteen elements, respectively?

3 Problems for Chapter 3: Random variables

Problem 14
Consider a random variable X with point probabilities $P(X = k) = 1/6$ for $k = 1, 2, 3, 4, 5, 6$. Draw the graph of X's distribution function $F : \mathbb{R} \to \mathbb{R}$.

Problem 15
Consider a random variable Y with density function $f(x) = 1$ for x in the interval $[0, 1]$. Draw the graph of Y's distribution function $F : \mathbb{R} \to \mathbb{R}$.

Problem 16
A red and a black die are thrown. Let the random variable X be the sum of the dice, and let the random variable Y be the difference (red minus black). Determine the point probabilities of X and Y. Are X and Y independent?

Problem 17
A continuous random variable X has density

$$f(x) = \begin{cases} e^{-x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Determine the distribution function F. What is $P(X > 1)$?

4 Problems for Chapter 4: Expected value and variance

Problem 18

A red and a black die are thrown, and X denotes the sum of the two dice. What is X's expected value, variance, and standard deviation? What fraction of the probability mass lies within one standard deviation of the expected value?

Problem 19

A red and a black die are thrown. Let the random variable X be the sum of the two dice, and let the random variable Y be the difference (red minus black). Calculate the covariance of X and Y. How does this agree with the result of Problem 16, where we showed that X and Y are independent?

5 Problems for Chapter 5: The Law of Large Numbers

Problem 20
Let X be a random variable with expected value μ and standard deviation σ. What does Chebyshev's Inequality say about the probability $P(|X - \mu| \geq n\sigma)$? For which n is Chebyshev's Inequality interesting?

Problem 21
A coin is tossed n times and the number k of heads is counted. Calculate for $n = 10$, 25, 50, 100, 250, 500, 1000, 2500, 5000, 10000 the probability P_n that k/n lies between 0.45 and 0.55. Determine if Chebyshev's Inequality is satisfied. What does the *Law of Large Numbers* say about P_n? Approximate P_n by means of the *Central Limit Theorem*.

Problem 22
Let X be normally distributed with standard deviation σ. Determine $P(|X - \mu| \geq 2\sigma)$. Compare with Chebyshev's Inequality.

Problem 23
Let X be exponentially distributed with intensity λ. Determine the expected value μ, the standard deviation σ, and the probability $P(|X - \mu| \geq 2\sigma)$. Compare with Chebyshev's Inequality.

Problem 24
Let X be binomially distributed with parameters $n = 10$ and $p = 1/2$. Determine the expected value μ, the standard deviation σ, and the probability $P(|X - \mu| \geq 2\sigma)$. Compare with Chebyshev's Inequality.

Problem 25
Let X be Poisson distributed with intensity $\lambda = 10$. Determine the expected value μ, the standard deviation σ, and the probability $P(|X - \mu| \geq 2\sigma)$. Compare with Chebyshev's Inequality.

Problem 26
Let X be geometrically distributed with probability parameter $p = 1/2$. Determine the expected value μ, the standard deviation σ, and the probability $P(|X - \mu| \geq 2\sigma)$. Compare with Chebyshev's Inequality.

6 Problems for Chapter 6: Descriptive statistics

Problem 27
Ten observations x_i are given:

$$4,\ 7,\ 2,\ 9,\ 12,\ 2,\ 20,\ 10,\ 5,\ 9$$

Determine the median, upper, and lower quartile and the inter-quartile range.

Problem 28
Four observations x_i are given:

$$2,\ 5,\ 10,\ 11$$

Determine the mean, empirical variance, and empirical standard deviation.

7 Problems for Chapter 7: Statistical hypothesis testing

Problem 29

In order to test whether a certain coin is fair, it is tossed ten times and the number k of heads is counted. Let p be the "head probability". We wish to test the null hypothesis

$$\mathbf{H}_0 : p = \frac{1}{2} \text{ (the coin is fair)}$$

against the alternative hypothesis

$$\mathbf{H}_1 : p > \frac{1}{2} \text{ (the coin is biased)}$$

We fix a significance level of 5%. What is the significance probability P if the number of heads is $k = 8$? Which values of k lead to acceptance and rejection, respectively, of $\mathbf{H}_0$? What is the risk of an error of type I? What is the strength of the test and the risk of an error of type II if the true value of p is 0.75?

8 Problems for Chapter 8: The binomial distribution

Problem 30
What is the probability P_1 of having at least six heads when tossing a coin ten times?

Problem 31
What is the probability P_2 of having at least 60 heads when tossing a coin 100 times?

Problem 32
What is the probability P_3 of having at least 600 heads when tossing a coin 1000 times?

9 Problems for Chapter 9: The Poisson distribution

Problem 33

In a certain shop, an average of ten customers enter per hour. What is the probability P that at most eight customers enter during a given hour?

Problem 34

What is the probability Q that at most 80 customers enter the shop from the previous problem during a day of 10 hours?

Problem 35

At the 2006 FIFA World Championship, a total of 64 games were played. The number of goals per game was distributed as follows:

8	games with	0	goals
13	games with	1	goal
18	games with	2	goals
11	games with	3	goals
10	games with	4	goals
2	games with	5	goals
2	games with	6	goals

Determine whether the number of goals per game may be assumed to be Poisson distributed.

10 Problems for Chapter 10: The geometrical distribution

Problem 36
A die is thrown until one gets a 6. Let V be the number of throws used. What is the expected value of V? What is the variance of V?

Problem 37
Assume W is geometrically distributed with probability parameter p. What is $P(W < n)$?

Problem 38
In order to test whether a given die is fair, it is thrown until a 6 appears, and the number n of throws is counted. How great should n be before we can reject the null hypothesis

$$\mathbf{H}_0 : \text{ the die is fair}$$

against the alternative hypothesis

$$\mathbf{H}_1 : \text{ the probability of having a 6 is less than 1/6}$$

at significance level 5%?

11 Problems for Chapter 11: The hypergeometrical distribution

Problem 39

At a lotto game, seven balls are drawn randomly from an urn containing 37 balls numbered from 0 to 36. Calculate the probability P of having exactly k balls with an even number for $k = 0, 1, \ldots, 7$.

Problem 40

Determine the same probabilities as in the previous problem, this time using the normal approximation.

12 Problems for Chapter 12: The multinomial distribution

Problem 41

A die is thrown six times. What is the probability of having two 4s, two 5s, and two 6s?

Problem 42

A die is thrown six times. What is the probability of having all six different numbers of pips?

13 Problems for Chapter 13: The negative binomial distribution

Problem 43

At the 2006 FIFA World Championship, a total of 64 games were played. The number of goals per game is given in Problem 35. Investigate whether the number of goals per game may be assumed to be negatively binomially distributed.

14 Problems for Chapter 14: The exponential distribution

Problem 44

A device contains two electrical components, A and B. The lifespans of A and B are both exponentially distributed with expected lifespans of five years and ten years, respectively. The device works as long as both components work. What is the expected lifespan of the device?

Problem 45

A device contains two electrical components, A and B. The lifespans of A and B are both exponentially distributed with expected lifespans of five years. The device works as long as at least one of the components works. What is the expected lifespan of the device?

15 Problems for Chapter 15: The normal distribution

Problem 46
Let X be a normally distributed random variable with expected value $\mu = 3$ and variance $\sigma^2 = 4$. What is $P(X \geq 6)$?

Problem 47
Let X be a normally distributed random variable with expected value $\mu = 5$. Assume $P(X \leq 0) = 10\%$. What is the variance of X?

Problem 48
A normally distributed random variable X satisfies $P(X \leq 0) = 0.40$ and $P(X \geq 10) = 0.10$. What is the expected value μ and the standard deviation σ?

Problem 49
Consider independent random variables $X \sim N(1, 3)$ and $Y \sim N(2, 4)$. What is $P(X + Y \leq 5)$?

16 Problems for Chapter 16: Distributions connected to the normal distribution

Problem 50
Let Q be χ^2 distributed with $df = 70$ degrees of freedom. What is the expected value and the variance? What is $P(Q < 100)$?

Problem 51
Let $X_1, \ldots, X_{10}$ be independent standard normally distributed random variables. What is $P(X_1^2 + \cdots + X_{10}^2 \leq 15)$?

Problem 52
Let T be t distributed with $df = 4$ degrees of freedom. What is the expected value and the variance? What is $P(T < 2)$?

Problem 53
Let V be F distributed with five degrees of freedom in the numerator and seven degrees of freedom in the denominator. Determine x such that $P(V < x) = 90\%$.

17 Problems for Chapter 17: Tests in the normal distribution

Problem 54
Suppose we have a sample $x_1, \ldots, x_{10}$ of 10 independent observations from a normal distribution with variance $\sigma^2 = 3$ and unknown expected value μ. Assume that the samle has mean $\bar{x} = 0.7$. Test the null hypothesis

$$\mathbf{H}_0 : \ \mu = 0$$

Problem 55
How great should the size of the sample in the previous problem have been in order to be able to reject $\mathbf{H}_0$?

Problem 56
Suppose we have a sample consisting of four observations

$$2, \ 5, \ 10, \ 11$$

from a normal distribution with unknown expected value μ and unknown variance σ^2. Test the null hypothesis

$$\mathbf{H}_0 : \mu = 0$$

against the alternative hypothesis

$$\mathbf{H}_1 : \mu > 0$$

Problem 57
Suppose we have a sample consisting of four observations

$$2, \ 5, \ 10, \ 11$$

from a normal distribution with unknown expected value μ and unknown variance σ^2. Test test null hypothesis

$$\mathbf{H}_0 : \sigma^2 = 10$$

against the alternative hypothesis

$$\mathbf{H}_1 : \sigma^2 > 10$$

Problem 58
Suppose we have a sample consisting of four observations

$$2, \ 5, \ 10, \ 11$$

from a normal distribution with unknown expected value μ_1 and unknown variance σ_1^2. Moreover, let there be given a sample

$$8, \ 12, \ 15, \ 17$$

from another independent normal distribution with unknown expected value μ_2 and unknown variance σ_2^2. The observations of the second sample are somewhat greater than the observations of the first sample, but the question is whether this difference is significant. Test the null hypothesis

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

against the alternative hypothesis

$$\mathbf{H}_1 : \mu_1 < \mu_2$$

18 Problems for Chapter 18: Analysis of variance (ANOVA)

Problem 59

Consider three samples:

sample 1: 2, 5, 10, 11

sample 2: 7, 11, 14, 16

sample 3: 3, 8, 9, 12

It is assumed that the samples come from independent normal distributions with common variance. Let μ_i be the expected value of the i'th normal distribution. Test the null hypothesis

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

using analysis of variance (ANOVA).

19 Problems for Chapter 19: The chi-squared test

Problem 60

In 1998, Danish newspaper subscriptions were distributed as follows (simplified):

	Share
Berlingske Tidende	14%
Politiken	17%
Jyllands-Posten	20%
Information	5%
B.T.	25%
Ekstra Bladet	19%

In a 2008 market analysis, 100 randomly chosen persons were asked about their subscriptions. The result was:

	Number
Berlingske Tidende	18
Politiken	13
Jyllands-Posten	22
Information	2
B.T.	16
Ekstra Bladet	29

Determine using a χ^2-test if the shares of the various newspapers have changed significantly since 1998.

Problem 61

Continuing the previous problem, we wish to determine which newspaper shares have significantly increased or decreased.

Problem 62

Over one week in spring 2005, the number of cars crossing the bridge between Denmark and Sweden was counted. The result was:

	Number
Monday	13804
Tuesday	13930
Wednesday	13863
Thursday	14023
Friday	14345
Saturday	14944
Sunday	15044

Is there a significant difference between these numbers?

20 Problems for Chapter 20: Contingency tables

Problem 63

In an opinion poll, randomly chosen Danes and Swedes were asked about their opinion ("pro" or "contra") about euthanasia. The result was:

	pro	contra
Danes	70	22
Swedes	85	46

Investigate whether there is a significant difference between Denmark and Sweden in opinions about euthanasia.

Problem 64

A new medicine is tested in an experiment involving 40 patients. During the experiment, the medicine is given to 20 randomly chosen patients, and the remaining 20 patients are given a placebo treatment. After the treatment, it is seen which patients are still ill. The result was:

	fit	ill
medicine	8	12
placebo	2	18

Investigate whether the medicine has had a significant positive effect.

Problem 65

In a sociological investigation, three men and three women are asked if they watch football regularly. All the men say yes, while all the women say no. Is this difference statistically significant?

21 Problems for Chapter 21: Distribution-free tests

Problem 66

In a biological experiment, ten plants are treated with a certain pesticide. Before the treatment, the numbers of plant lice x_i on each plant are counted. One week after the treatment, the numbers of plant lice y_i are counted again. The result was:

Plant no.	x_i	y_i
1	41	27
2	51	59
3	66	76
4	68	65
5	46	36
6	69	54
7	47	49
8	44	51
9	60	55
10	44	35

Use Wilcoxon's distribution-free test to determine whether the pesticide had a significant (positive) effect.

Problem 67

The experiment of the previous problem is repeated, this time with 100 plants. The statistics now become

$$t_+ = 3045 \text{ and } t_- = 2005$$

Is there a significant effect now?

Problem 68

A car factory counts the number of defects in randomly chosen cars from two different production lines.

Car no.	Line 1	Line 2
1	170	79
2	197	126
3	50	189
4	151	137
5	94	167
6	46	188
7	173	54
8	26	155
9	118	82
10	171	218
11	70	242
12	146	
13	55	
14	132	

Use a distribution-free test to determine if there are significantly more defects on one of the two production lines.

Problem 69

The sample from the previous problem is enlarged such that there are now $n = 100$ observations from Line 1 and $m = 50$ observations from Line 2. The statistic is found to be

$$t_x = 7002$$

Is there now a significant difference between the numbers of defects on the two lines?

22 Solutions

Solution of Problem 1
As stated in section 2.5 of the *Compendium*, this number can be computed as a binomial coefficient:

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

Solution of Problem 2
We know the number of possible poker hands N from Problem 1. Of these, only four hands are royal flush. Therefore, the probability becomes

$$P = \frac{4}{N} = \frac{4}{2598960} = \frac{1}{649740} \approx 0.00015\%$$

Solution of Problem 3
We know the number of possible poker hands N from Problem 1. We have to calculate the number n of hands with straight flush. There are four possibilities for the suit. There are ten possibilities for the value of the highest card (from 5 to ace). This gives

$$4 \cdot 10 = 40$$

possibilities. However, we have to subtract the number of hands with royal flush (Problem 2) from this number. In total we get

$$n = 40 - 4 = 36$$

hands with straight flush. The probability of a straight flush thus becomes

$$P = \frac{n}{N} = \frac{36}{2598960} \approx \frac{1}{72193} \approx 0.0014\%$$

Solution of Problem 4
We know the number of possible poker hands N from Problem 1. We have to calculate the number n of hands with four of a kind. There are 13 possible values (ace, king, queen, etc.) of the four cards, and moreover 48 possibilities for the fifth card. In total this is

$$n = 13 \cdot 48 = 624$$

hands with four of a kind. The probability of four of a kind thus becomes

$$P = \frac{n}{N} = \frac{624}{2598960} = \frac{1}{4165} \approx 0.024\%$$

Solution of Problem 5
We know the number of possible poker hands N from Problem 1. We have to calculate the number n of hands with "full house". There are 13 possibilities for the value of the group of three cards.

Moreover, there are 12 possibilities for the value of the pair of cards. Finally, one has to take into account that if we have, say, three aces and two kings, then the three aces can be chosen in

$$\binom{4}{3} = 4$$

ways, and the two kings can be chosen in

$$\binom{4}{2} = 6$$

ways. In total there are

$$n = 13 \cdot 12 \cdot 4 \cdot 6 = 3744$$

hands with full house. The probability of full house thus becomes

$$P = \frac{n}{N} = \frac{3744}{2598960} \approx \frac{1}{694} \approx 0.14\%$$

Solution of Problem 6

We have to calculate the number n of hands with a flush. There are four possibilities for the suit (spades, hearts, diamonds, clubs). Out of 13 cards of the same suit, one can choose five cards in

$$\binom{13}{5} = 1287$$

ways. This gives

$$4 \cdot 1287 = 5148$$

possibilities. However, we have to subtract the number of hands with a straight flush (Problem 3) and the number of hands with a royal flush (Problem 2) from this number. In total we get

$$n = 5148 - 36 - 4 = 5108$$

hands with a flush. The probability of a flush thus becomes

$$P = \frac{n}{N} = \frac{5108}{2598960} \approx \frac{1}{509} \approx 0.20\%$$

Solution of Problem 7

We determine the number n of hands with a straight. For the value of the highest card, there are ten possibilities (from 5 to ace). For each of the five cards, there are four possibilities for the suit (spades, hearts, diamonds, clubs). This gives

$$10 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 10240$$

possibilities. However, we have to subtract the number of hands with a straight flush (Problem 3) and the number of hands with a royal flush (Problem 2) from this number. In total we get

$$n = 10240 - 36 - 4 = 10200$$

hands with a straight. The probability of a straight thus becomes

$$P = \frac{n}{N} = \frac{10200}{2598960} \approx \frac{1}{255} \approx 0.39\%$$

Solution of Problem 8

We have to calculate the number n of hands with three of a kind. There are 13 possibilities for the value of the three cards. If we have, say, three aces, then these aces can be chosen in

$$\binom{4}{3} = 4$$

different ways. Moreover, there are

$$\binom{12}{2} = 66$$

possibilities for the values of the remaining two cards, and four possibilities for the suit of each of these. In total there are

$$n = 13 \cdot 4 \cdot 66 \cdot 4 \cdot 4 = 54912$$

hands with three of a kind. The probability of three of a kind thus becomes

$$P = \frac{n}{N} = \frac{54912}{2598960} \approx \frac{1}{47} \approx 2.1\%$$

Solution of Problem 9
We have to calculate the number n of hands with two pair. There are

$$\binom{13}{2} = 78$$

possibilities for the values of the two pair. For both of these two values, a pair can be chosen in

$$\binom{4}{2} = 6$$

different ways. There are 44 possibilities for the fifth card. In total there are

$$n = 78 \cdot 6 \cdot 6 \cdot 44 = 123552$$

hands with two pair. The probability of two pair thus becomes

$$P = \frac{n}{N} = \frac{123552}{2598960} \approx \frac{1}{21} \approx 4.8\%$$

Solution of Problem 10
We have to calculate the number n of hands with one pair. There are 13 possibilities for the value of the pair. One pair of given value can be chosen in

$$\binom{4}{2} = 6$$

different ways. For the values of the remaining three cards there are

$$\binom{12}{3} = 220$$

possibilities (the remaining three cards must have values different from each other and different from the pair). Finally there are four possibilities for the suit of each of the remaining three cards. In total this gives

$$n = 13 \cdot 6 \cdot 220 \cdot 4 \cdot 4 \cdot 4 = 1098240$$

hands with "one pair". The probability of one pair thus becomes

$$P = \frac{n}{N} = \frac{1098240}{2598960} \approx 42\%$$

Solution of Problem 11
There is a total of 36 different sample points (i, j). Of these, there are six sample points with sum at least ten, namely $(4, 6),\ (5, 5),\ (5, 6),\ (6, 4),\ (6, 5)$ and $(6, 6)$. The probability thus becomes

$$P = \frac{6}{36} = \frac{1}{6} \approx 17\%$$

There are six possible sample points $(i, 5)$ where the black die shows 5. Of these, there are two sample points where the sum is at least ten, namely $(5, 5)$ and $(6, 5)$. The conditional probability thus becomes

$$Q = \frac{2}{6} = \frac{1}{3} \approx 33\%$$

There are 11 possible sample points where at least one of the dice shows 5. Of these, there are three sample points where the sum is at least ten, namely $(5, 5)$, $(5, 6)$ and $(6, 5)$. The conditional probability thus becomes

$$R = \frac{3}{11} \approx 27\%$$

Solution of Problem 12

The number of subsets with three elements equals the binomial coefficient

$$\binom{10}{3} = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$$

The number of subsets with seven elements is the same:

$$\binom{10}{7} = \binom{10}{3} = 120$$

Solution of Problem 13
The answer is the multinomial coefficient

$$\binom{30}{5\ 10\ 15} = \frac{30!}{5!10!15!} = 465817912560$$

Solution of Problem 14
The distribution function F is a step function given by

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ 1/6 & \text{for } 1 \leq x < 2 \\ 2/6 & \text{for } 2 \leq x < 3 \\ 3/6 & \text{for } 3 \leq x < 4 \\ 4/6 & \text{for } 4 \leq x < 5 \\ 5/6 & \text{for } 5 \leq x < 6 \\ 1 & \text{for } x \geq 6 \end{cases}$$

Solution of Problem 15
The distribution function F is the continuous function given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$

Solution of Problem 16
X takes values in the set $\{2, 3, 4, \ldots, 12\}$. The point probabilities are

$$P(X = k) = \begin{cases} 1/36 & \text{for } k = 2 \\ 2/36 & \text{for } k = 3 \\ 3/36 & \text{for } k = 4 \\ 4/36 & \text{for } k = 5 \\ 5/36 & \text{for } k = 6 \\ 6/36 & \text{for } k = 7 \\ 5/36 & \text{for } k = 8 \\ 4/36 & \text{for } k = 9 \\ 3/36 & \text{for } k = 10 \\ 2/36 & \text{for } k = 11 \\ 1/36 & \text{for } k = 12 \end{cases}$$

Y takes values in the set $\{-5, -4, \ldots, 4, 5\}$. The point probabilities are

$$P(Y=k) = \begin{cases} 1/36 & \text{for } k=-5 \\ 2/36 & \text{for } k=-4 \\ 3/36 & \text{for } k=-3 \\ 4/36 & \text{for } k=-2 \\ 5/36 & \text{for } k=-1 \\ 6/36 & \text{for } k=0 \\ 5/36 & \text{for } k=1 \\ 4/36 & \text{for } k=2 \\ 3/36 & \text{for } k=3 \\ 2/36 & \text{for } k=4 \\ 1/36 & \text{for } k=5 \end{cases}$$

If we know, say, that X takes the value 12, then we can conclude that Y takes the value 0 (since both dice in that case show 6). Consequently, X and Y are *not* independent.

Solution of Problem 17
The distribution function is found by integrating the density:

$$F(x) = \int_{-\infty}^{x} e^{-t} dt = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-x} & \text{for } x > 0 \end{cases}$$

Moreover:

$$P(X > 1) = 1 - F(1) = e^{-1}$$

Solution of Problem 18
We know the point probabilities from Problem 16. Since the point probabilities are symmetrical around 7, we get at once

$$E(X) = 7$$

Let us compute the variance using the formula $\text{var}(X) = E(X^2) - E(X)^2$ from section 4.5. We get

$$E(X^2) = \sum_{k=2}^{12} P(X=k) \cdot k^2 = \frac{1974}{36} = 54.8$$

and thus

$$\text{var}(X) = E(X^2) - E(X)^2 = 54.8 - 7^2 = 5.8$$

The standard deviation is the square root of the variance:

$$\sigma = \sqrt{\text{var}(X)} = \sqrt{5.8} = 2.4$$

Within 1 standard deviation around the the expected value, that is in the interval from $7 - 2.4$ to $7 + 2.4$, we have

$$\sum_{k=5}^{9} P(X = k) = \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} = \frac{24}{36}$$

i.e. two thirds of the probability mass (as predicted in section 4.4).

Solution of Problem 19

Let us find the covariance using the formula

$$\mathrm{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

from section 4.6. From the point probabilities it appears that

$$E(X) = 7 \quad \text{and} \quad E(Y) = 0$$

It is somewhat tedious to calculate $E(X \cdot Y)$. There is a total of 36 possible sample points (r, s) when the red and the black die are thrown. For any given sample point (r, s), we know that X takes the value $r + s$, whereas Y takes the value $r - s$. The product $X \cdot Y$ thus takes the value $(r + s)(r - s) = r^2 - s^2$. The expected value of $X \cdot Y$ is the mean of these 36 values, i.e.

$$E(X \cdot Y) = \frac{1}{36} \sum_{r=1}^{6} \sum_{s=1}^{6} (r^2 - s^2) = 0$$

We therefore find the covariance

$$\text{Cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0 - 7 \cdot 0 = 0$$

The covariance of *independent* random variables is always 0 (section 4.6). We have shown in 16 that X and Y are *not* independent. This problem therefore shows that we *cannot* conclude conversely that the variables are independent when the covariance is 0.

Solution of Problem 20
Chebyshev's Inequality says

$$P(|X - \mu| \geq n\sigma) \leq \frac{1}{n^2}$$

In words this means that the probability that X takes values more than n standard deviations away from its expected value is small when n is large. Of course this is only interesting for $n > 1$ since every probability a priori is at most 1.

Solution of Problem 21
The number k is obviously binomially distributed with parameters n and $p = 1/2$. It is seen that k/n lies in the interval $[0.45\ ;\ 0.55]$ if and only if k lies in the interval $[0.45n\ ;\ 0.55n]$. Vi may therefore calculate P_n using the formula for the point probabilities of the binomial distribution. The results are seen here:

n	P_n
10	0.25
25	0.38
50	0.52
100	0.73
250	0.886
500	0.978
1000	0.9986
2500	0.99999949
5000	0.9999999999987
10000	0.99999999999999999999987

Since k has expected value $np = n/2$ and standard deviation $\sqrt{npq} = \sqrt{n/4} = \sqrt{n}/2$, it follows that k/n will have expected value $\mu = 1/2$ and standard deviation $\sigma = 1/2\sqrt{n}$. Note that the the standard deviation converges to 0 as n goes to infinity.

Chebyshev's Inequality gives

$$P_n = P(|k/n - 1/2| \leq 0.05) \geq 1 - \frac{\sigma^2}{(0.05)^2} = 1 - \frac{100}{n}$$

which is only interesting for $n > 100$. Here a table shows the right-hand side of Chebyshev's

Inequality:

n	$1 - 100/n$
10	-9
25	-3
50	-1
100	0
250	0.6
500	0.8
1000	0.9
2500	0.96
5000	0.98
10000	0.99

By comparison with the first table, it is seen that Chebyshev's Inequality really is satisfied.

The Law of Large Numbers says that P_n converges to 100% as n goes to infinity. We note that this is a direct consequence of Chebyshev's Inequality and also appears very clearly from both of the above tables.

The Central Limit Theorem tells that the distribution of k/n approaches a normal distribution when n goes to infinity. Thus P_n can be approximated by

$$Q_n := \Phi\left(\frac{\sqrt{n}}{10}\right) - \Phi\left(-\frac{\sqrt{n}}{10}\right) = 1 - 2 \cdot \Phi\left(-\frac{\sqrt{n}}{10}\right)$$

since 0.05 equals $\sqrt{n}/10$ times the standard deviation σ. Here a table shows how well Q_n approximates P_n:

n	Q_n
10	0.25
25	0.38
50	0.52
100	0.68
250	0.886
500	0.975
1000	0.9984
2500	0.99999943
5000	0.9999999999985
10000	0.999999999999999999999985

Solution of Problem 22

The probability is

$$P(|X - \mu| \geq 2\sigma) = 2 \cdot \Phi(-2) = 2.6\%$$

In contrast, Chebyshev's Inequality only gives the weaker statement

$$P(|X - \mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = 25\%$$

Solution of Problem 23

The expected value of X is $1/\lambda$ and the standard deviation is $\sigma = \sqrt{1/\lambda^2} = 1/\lambda$. The probability that X takes a value more than two standard deviations from μ is

$$P(X \geq 3/\lambda) = 1 - F(3/\lambda) = \exp(-3) = 5.0\%$$

where we have used the distribution function of the exponential distribution, $F(x) = 1 - \exp(-\lambda x)$. In contrast, Chebyshev's Inequality only gives the weaker statement

$$P(|X - \mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = 25\%$$

Solution of Problem 24

The expected value of X is $\mu = np = 5$. The standard deviation of X is

$$\sigma = \sqrt{npq} = \sqrt{2.5} = 1.6$$

The probability that X takes a value more than two standard deviations from μ is

$$P(|X - \mu| \geq 4) = 2.1\%$$

In contrast, Chebyshev's Inequality only gives the weaker statement

$$P(|X-\mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = 25\%$$

Solution of Problem 25

The expected value of X is $\mu = \lambda$ and the standard deviation is $\sigma = \sqrt{\lambda} = 3.2$. The probability that X takes a value more than two standard deviations from μ is

$$P(X < 4) + P(X > 16) = 3.7\%$$

In contrast, Chebyshev's Inequality only gives the weaker statement

$$P(|X-\mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = 25\%$$

Solution of Problem 26

The expected value of X is $q/p = 1$ and the standard deviation is $\sigma = \sqrt{q/p^2} = 1.4$. The probability that X takes a value more than two standard deviations from μ is

$$P(X \geq 4) = \left(\frac{1}{2}\right)^4 = 6.3\%$$

In contrast, Chebyshev's Inequality only gives the weaker statement

$$P(|X-\mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = 25\%$$

Solution of Problem 27

The observations are ordered according to size:

$$2,\ 2,\ 4,\ 5,\ 7,\ 9,\ 9,\ 10,\ 12,\ 20$$

The median is the mean of the two "middle" observations, i.e.

$$x(0.5) = \frac{7+9}{2} = 8$$

Similarly the lower and upper quartiles are

$$x(0.25) = \frac{2+4}{2} = 3\ ,\ x(0.75) = \frac{10+12}{2} = 11$$

The inter-quartile range thus becomes $11 - 3 = 8$.

Solution of Problem 28

The mean is

$$\bar{x} = \frac{2+5+10+11}{4} = 7$$

The empirical variance is

$$s^2 = \frac{(2-7)^2 + (5-7)^2 + (10-7)^2 + (11-7)^2}{4-1} = 18$$

The empirical standard deviation thus becomes

$$s = \sqrt{18} \approx 4.24$$

Solution of Problem 29

If the coin is fair, then k originates from a Bin$(10, \frac{1}{2})$ distribution. A Bin$(10, \frac{1}{2})$ distributed random variable X has the following point probabilities:

k	$\cdots$	5	6	7	8	9	10
$P(X=k)$	$\cdots$	24.6%	20.5%	11.7%	4.4%	1.0%	0.1%

The test's significance probability P is the probability of having k or more "heads" when tossing a fair coin ten times, i.e.

$$P = P(X \geq k)$$

For $k = 8$ we get $P = 5.5\%$ and $\mathbf{H}_0$ must be accepted. For $k = 9$ we get $P = 1.1\%$. A test at significance level 5% should therefore reject $\mathbf{H}_0$ if $k \geq 9$, and accept $\mathbf{H}_0$ if $k \leq 8$. The risk of committing an error of type I is thus 1.1%.

If the true value of p is 0.75, then the strength of the test is

$$P(Y \geq 9) = 24.4\%$$

and the risk of an error of type II is

$$P(Y \leq 8) = 75.6\%$$

Here Y is a Bin$(10; 0.75)$ distributed random variable.

Solution of Problem 30

The number X of heads is obviously Bin$(10, \frac{1}{2})$ distributed. In order to calculate P_1 we simply use the point probabilities since this is doable in this case, and moreover we are at the borderline with respect to when the normal approximation may be used. We get

$$\begin{aligned} P_1 &= P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10) \\ &= 0.205 + 0.117 + 0.044 + 0.010 + 0.001 \\ &= 37.7\% \end{aligned}$$

Solution of Problem 31

The number X of heads is obviously Bin$(100, \frac{1}{2})$ distributed. It is now advantageous to use the normal approximation to the binomial distribution, and we get

$$P_2 = P(X \geq 60) \approx 1 - \Phi\left(\frac{9.5}{\sqrt{25}}\right) = 1 - \Phi(1.9)$$

Table C.2 shows

$$P_2 = 1 - 0.971 = 2.9\%$$

Solution of Problem 32

The number X of heads is Bin$(1000, \frac{1}{2})$ distributed. We use the normal approximation again:

$$P_3 = P(X \geq 600) \approx 1 - \Phi\left(\frac{99.5}{\sqrt{250}}\right) = 1 - \Phi(6.29)$$

Table C.2 shows that P_3 is (much) less than 0.01%.

Solution of Problem 33

The number X of customers during an hour may be assumed to be Poisson distributed with intensity $\lambda = 10$, i.e. $X \sim \text{Pois}(10)$. The sought-after probability

$$P = P(X \leq 8) = \sum_{i=0}^{8} P(X = i)$$

is calculated using point probabilities of the Poisson distribution (section 9.3):

$$P = 33.2\%$$

Solution of Problem 34

The number Y of customers during an entire day is the sum of ten independent $\text{Pois}(10)$ distributed random variables:

$$Y = X_1 + \cdots + X_{10}$$

Thus Y is itself Poisson distributed with intensity 100, cf. section 9.5, and thus $Y \sim \text{Pois}(100)$ (one could also argue more directly to show this). Our probability

$$Q = P(Y \leq 80)$$

is calculated using the normal approximation to the Poisson distribution:

$$Q \approx \Phi\left(\frac{-19.5}{\sqrt{100}}\right) = \Phi(-1.95) = 2.6\%$$

where we looked up Φ in Table C.2.

Solution of Problem 35

If O_i denotes the number of games with i goals, then the observations can be presented in a table as follows:

i	O_i
0	8
1	13
2	18
3	11
4	10
5	2
6	2

We have to investigate whether these observations could originate from a $\text{Pois}(\lambda)$ distribution. The intensity λ is estimated as $144/64 = 2.25$ since a total of 144 goals were scored in the 64 games. The point probabilities in a $\text{Pois}(2.25)$ distribution are

i	p_i
0	0.105
1	0.237
2	0.267
3	0.200
4	0.113
5	0.051
6	0.019
≥ 7	0.008

The expected numbers thus become (since $E_i = 64p_i$)

i	E_i
0	6.7
1	15.2
2	17.1
3	12.8
4	7.2
≥ 5	4.9

Note that we have merged some categories here in order to get $E_i \geq 3$. Now we compute the statistic:

$$\begin{aligned}\chi^2 &= \frac{(8-6.7)^2}{6.7} + \frac{(13-15.2)^2}{15.2} + \frac{(18-17.1)^2}{17.1} + \\ &\quad \frac{(11-12.8)^2}{12.8} + \frac{(10-7.2)^2}{7.2} + \frac{(4-4.9)^2}{4.9} \\ &= 2.1\end{aligned}$$

Since there are six categories and we have estimated one parameter from the data, the statistic must be compared with the χ^2 distribution with $df = 6 - 1 - 1 = 4$ degrees of freedom. Table C.3 gives a significance probability of more than 50%.

Conclusion: *It is reasonable to claim that the number of goals per game is Poisson distributed.*

Solution of Problem 36

The expected wait until success is the reciprocal probability of success:

$$E(V) = \frac{1}{1/6} = 6$$

The number of failures before success, $W = V - 1$, is Geo$(1/6)$ distributed. W and V have the same variance:

$$\text{var}(V) = \text{var}(W) = q/p^2 = 30$$

Solution of Problem 37

We get

$$P(W < n) = 1 - P(W \geq n) = 1 - q^n$$

where as always $q = 1 - p$.

Solution of Problem 38

The significance probability, i.e. the probability of having to use at least n throws given $\mathbf{H}_0$, is

$$P = \left(\frac{5}{6}\right)^{n-1}$$

P is less than 5% for $n \geq 18$. Therefore, we have to reject $\mathbf{H}_0$ if n is at least 18.

Solution of Problem 39

The number Y of even numbers is hypergeometrically distributed with parameters $n = 7$, $r = 19$, $s = 18$ and $N = 37$, i.e. $Y \sim \mathrm{HG}(7, 19, 37)$. The probability P is computed using the formula for the point probabilities:

$$P = P(Y = k) = \frac{\binom{19}{k} \cdot \binom{18}{7-k}}{\binom{37}{7}}$$

We get (in percentages):

k	0	1	2	3	4	5	6	7
$P(Y = k)$	0.3	3.4	14.2	28.8	30.7	17.3	4.7	0.5

Solution of Problem 40

A random variable $Y \sim \mathrm{HG}(7, 19, 37)$ has expected value

$$E(Y) = nr/N = 7 \cdot 19/37 = 3.595$$

and standard deviation

$$\begin{aligned} \sqrt{\mathrm{var}(Y)} &= \sqrt{nrs(N-n)/(N^2(N-1))} \\ &= \sqrt{7 \cdot 19 \cdot 18 \cdot (37-7)/(37^2(37-1))} \\ &= 1.207. \end{aligned}$$

The normal approximation now gives

$$P = P(Y = k) \approx \varphi\left(\frac{k - 3.595}{1.207}\right) \cdot \frac{1}{1.207}$$

where φ is the density of the standard normal distribution (section 15.4). Using the different possible values for k, we get (in percentages):

k	0	1	2	3	4	5	6	7
$P(Y = k)$	0.4	3.3	13.8	29.3	31.2	16.8	4.5	0.6

If we compare with the previous problem, we see that the normal approximation in this case is acceptable without being magnificent. The condition that "n be small compared to both r and s" (section 11.6), in our case "7 is small compared to both 19 and 18", can thus be said to be just fulfilled here.

Solution of Problem 41

The formula for the point probabilities in the multinomial distribution gives

$$P = \binom{6}{0\,0\,0\,2\,2\,2} \cdot \left(\frac{1}{6}\right)^6 \approx 0.19\%$$

Solution of Problem 42

It is most easy to compute ad hoc:

$$P = 1 \cdot \frac{5}{6} \cdot \frac{4}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{1}{6} \approx 1.5\%$$

Solution of Problem 43

The average number of goals per game is

$$\bar{k} = 2.250$$

and the empirical variance is

$$s^2 = 2.254$$

The parameters of the negative binomial distribution are estimated as

$$\hat{n} = \frac{\bar{k}^2}{s^2 - \bar{k}} = 1276$$

and

$$\hat{p} = \frac{\bar{k}}{s^2} = 0.998$$

Now we can calculate the expected values:

i	E_i
0	6.8
1	15.2
2	17.1
3	12.8
4	7.2
≥ 5	5.0

These expected values are almost identical with those from Problem 35. Since the Poisson distribution is a more natural choice and also requires one parameter less, there is no reason to explain the observations using a negative binomial distribution.

Solution of Problem 44
A's lifespan is the wait until an event (death) that occurs spontaneously with intensity $\lambda_A = 1/5$. Similarly, B's lifespan is the wait until an event that occurs spontaneously with intensity $\lambda_B = 1/10$. The lifespan of the device is thus the wait until an event that occurs spontaneously with intensity $\lambda = \lambda_A + \lambda_B = 3/10$. The expected lifespan of the device thus becomes $\lambda^{-1} \approx 3.3$ years.

Solution of Problem 45
The expected wait until the first component dies is seen to be 2.5 years as in the previous problem. After this, the expected wait until the second component dies is seen to be five years. The expected lifespan of the device thus becomes 7.5 years.

Solution of Problem 46
The standard deviation is $\sigma = 2$. We get

$$P(X \geq 6) = \Phi\left(-\frac{6-\mu}{\sigma}\right) = \Phi(-1.5) = 6.68\%$$

by looking up in Table C.2.

Solution of Problem 47
The information $\Phi(u) = P(X \leq 0) = 0.10$ implies $u = -1.28$ (Table C.2). This yields $0 = 5 - 1.28 \cdot \sigma$ and thereby $\sigma = 5/1.28 = 3.91$. The variance thus becomes $\sigma^2 = 15.3$.

Solution of Problem 48
The information $P(X \leq 0) = 0.40$ implies

$$\Phi\left(\frac{0-\mu}{\sigma}\right) = 0.40 \text{ and thus } \frac{0-\mu}{\sigma} = -0.25$$

The information $P(X \geq 10) = 0.10$ implies

$$\Phi\left(-\frac{10-\mu}{\sigma}\right) = 0.10 \text{ and thus } \frac{10-\mu}{\sigma} = 1.28$$

These two equations together yield $\mu = 1.6$ and $\sigma = 6.5$.

Solution of Problem 49
$X+Y$ is itself normally distributed with expected value $\mu = 1+2 = 3$ and variance $\sigma^2 = 3+4 = 7$ (section 15.10). Therefore

$$P(X+Y \leq 5) = \Phi\left(\frac{5-3}{\sqrt{7}}\right) = \Phi(0.76) = 78\%$$

Solution of Problem 50
The expected value is 70, and the variance 140. Table C.3 shows that $P(Q < 100)$ is a bit below 99%.

Solution of Problem 51
The sum is χ^2 distributed with $df = 10$ degrees of freedom. Table C.3 shows that the desired probability lies somewhere between 80% and 90%.

Solution of Problem 52
The expected value is 0, and the variance 2. Table C.4 shows that $P(T < 2)$ is between 90% and 95%.

Solution of Problem 53
Table C.5 shows $x = 2.88$.

Solution of Problem 54
We calculate the statistic

$$u = \frac{\sqrt{10} \cdot 0.7}{\sqrt{3}} = 1.28$$

Since we test $\mathbf{H}_0$ against all possible alternative hypotheses, the significance probability becomes

$$P = 2 \cdot \Phi(-u) = 20\%$$

If we, as commonly, test at significance level 5%, then $\mathbf{H}_0$ *cannot* be rejected.

Solution of Problem 55
The statistic is

$$u = \frac{\sqrt{n} \cdot 0.7}{\sqrt{3}}$$

If the significance probability

$$P = 2 \cdot \Phi(-u)$$

is to be less then 5%, then u must be greater than 1.96. This implies

$$n > \left(\frac{1.96 \cdot \sqrt{3}}{0.7} \right)^2 \approx 23$$

Solution of Problem 56
The obvious thing to do is to use *Student's t-test* in this situation. We have earlier (Problem 28) computed the mean $\bar{x} = 7$ and the empirical standard deviation $s = 4.24$ of the sample. The statistic now becomes

$$t = \frac{\sqrt{4}(\bar{x} - 0)}{s} = 3.30$$

The significance probability thus becomes

$$1 - F_{\text{Student}}(3.30) < 1 - 0.975 = 2.5\%$$

where F_{Student} is the distribution function for Student's t distribution with $df = 4-1 = 3$ degrees of freedom. It is seen in Table C.4 that $F_{\text{Student}}(3.30)$ is a bit above 0.975. Since the significance probability thereby is less than 5%, we reject $\mathbf{H}_0$ after testing against $\mathbf{H}_1$.

Solution of Problem 57
We have earlier (Problem 28) computed the mean $\bar{x} = 7$ and the empirical variance $s^2 = 18$ of the sample. The relevant statistic (section 17.3) is

$$q = \frac{(4-1)s^2}{10} = 5.4$$

The significance probability thus becomes

$$1 - F_{\chi^2}(5.4) \approx 1 - 0.85 = 15\%$$

where F_{χ^2} is the distribution function of the χ^2 distribution with $df = 4 - 1 = 3$ degrees of freedom. It is seen in Table C.3 that $F_{\chi^2}(5.4)$ is more or less in the middle between 0.8 and 0.9. Since the significance probability thereby is greater than 5%, $\mathbf{H}_0$ *cannot* be rejected – even though

the empirical variance 18 is markedly greater than 10; the explanation lies in the very small size of the sample.

Solution of Problem 58

If the variances σ_1^2 and σ_2^2 are different, then the problem is unsolvable (this is the so-called *Fisher-Behrens problem* mentioned in the Compendium). Therefore, we first have to test the auxiliary hypothesis

$$\mathbf{H}_0^* : \sigma_1 = \sigma_2$$

against the alternative

$$\mathbf{H}_1^* : \sigma_1 \neq \sigma_2$$

Only if $\mathbf{H}_0^*$ is accepted, the original problem can be solved.

The first sample has mean $\bar{x} = 7$ and empirical variance $s_1^2 = 18$. The second sample has mean $\bar{y} = 13$ and empirical variance $s_2^2 = 15.3$. We calculate the statistic (section 17.7 and 17.9)

$$v = \frac{s_1^2}{s_2^2} = 1.17$$

as well as

$$v^* = \max\left\{v, \frac{1}{v}\right\} = 1.17$$

The significance probability is

$$P = 2 \cdot (1 - F_{\text{Fisher}}(v^*))$$

where F_{Fisher} is the distribution function of Fisher's F distribution with $4 - 1 = 3$ degrees of freedom in both numerator and denominator. Table C.5 gives $F_{\text{Fisher}}(5.39) = 90\%$ which implies $F_{\text{Fisher}}(1.17) < 90\%$. We thus get

$$P > 2 \cdot (1 - 90\%) = 20\%$$

With a significance probability as large as that, we *accept* the auxiliary hypothesis $\mathbf{H}_0^*$.

Now we are ready to test $\mathbf{H}_0$ against $\mathbf{H}_1$. The procedure is described in section 17.8. The "pooled" variance is computed as

$$s_{\text{pool}}^2 = \frac{3s_1^2 + 3s_2^2}{6} = 16.7$$

The statistic thus becomes

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(1/4 + 1/4)s_{\text{pool}}^2}} = \frac{7 - 13}{\sqrt{(1/4 + 1/4)16.7}} = -2.08$$

The significance probability is now seen to be

$$P = 1 - F_{\text{Student}}(-t) = 1 - F_{\text{Student}}(2.08) < 5\%$$

where we have looked up $F_{\text{Student}}(2.08)$ in Table C.4 under $df = 4 + 4 - 2 = 6$. Since P is *less* than 5%, we *reject* $\mathbf{H}_0$.

Solution of Problem 59

We have $k = 3$ samples with $n_i = 4$ observations each, in total $n = 12$ observations. The means of the samples are

$$\bar{x}_1 = 7\ ,\ \bar{x}_2 = 12\ ,\ \bar{x}_3 = 8$$

and the empirical variances are

$$s_1^2 = 18\ ,\ s_2^2 = 15.3\ ,\ s_3^2 = 14$$

The grand mean is

$$\bar{x} = \frac{7 + 12 + 8}{3} = 9$$

Next we estimate the variance **within** the samples

$$s_I^2 = \frac{1}{n-k}\sum_{j=1}^{3}(4-1)s_j^2 = \frac{18 + 15.3 + 14}{3} = 15.4$$

and the variance **between** the samples

$$s_M^2 = \frac{1}{k-1}\sum_{j=1}^{3} 4(\bar{x}_j - \bar{x})^2 = 2((7-9)^2 + (12-9)^2 + (8-9)^2) = 28$$

Finally the statistic can be computed:

$$v = \frac{s_M^2}{s_I^2} = \frac{28}{15.4} = 1.81$$

The significance probability is

$$P = 1 - F_{\text{Fisher}}(v) = 1 - F_{\text{Fisher}}(1.81)$$

where F_{Fisher} is the distribution function of Fisher's F distribution with $k - 1 = 2$ degrees of freedom in the numerator and $n - k = 9$ degrees of freedom in the denominator. Table C.5 shows $F_{\text{Fisher}}(1.81) < 90\%$ and thus $P > 10\%$. Consequently, we accept $\mathbf{H}_0$.

Let us finally sum up the calculations in the usual ANOVA table:

sample number	**1**	**2**	**3**
	2	7	3
	5	11	8
	10	14	9
	11	16	12
Mean $\bar{x}_j$	7	12	8
Empirical variance s_j^2	18	15.3	14

$\bar{x} = 9$	(grand mean)
$s_I^2 = (s_1^2 + s_2^2 + s_3^2)/3 = 15.4$	(variance within samples)
$s_M^2 = 2\sum(\bar{x}_j - \bar{x})^2 = 28$	(variance between samples)
$v = s_M^2/s_I^2 = 1.81$	(statistic)

Solution of Problem 60

The *observed* numbers O_i are the numbers of the second table. The *expected* numbers E_i are easy to find:

	E_i
Berlingske Tidende	14
Politiken	17
Jyllands-Posten	20
Information	5
B.T.	25
Ekstra Bladet	19

The statistic thus becomes

$$\chi^2 = \frac{(18-14)^2}{14} + \frac{(13-17)^2}{17} + \frac{(22-20)^2}{20} + \frac{(2-5)^2}{5} + \frac{(16-25)^2}{25} + \frac{(29-19)^2}{19} = 12.6$$

In order to find the significance probability P of the test, we use the χ^2 distribution with $df = 6-1 = 5$ (the degrees of freedom are one less than the number of categories). The result is (by looking up in Table C.3):

$$P = 1 - F_{\chi^2}(12.6) \approx 4\%$$

Since the significance probability is less than 5%, we conclude that the ratios have changed significantly.

Solution of Problem 61
We compute the *standardized residuals:*

	r_i
Berlingske Tidende	1.2
Politiken	–1.1
Jyllands-Posten	0.5
Information	–1.4
B.T.	–2.1
Ekstra Bladet	2.5

Since standardized residuals numerically greater than 2 are a sign of an extreme observed number, we conclude that *B.T.* has decreased markedly, whereas *Ekstra Bladet* has increased markedly.

Solution of Problem 62
A total of $N = 99953$ cars were observed which gives a mean of $\lambda = 14279$ cars per day. If the seven observed numbers are from the same distribution (e.g. a Pois(14279) distribution), the expected numbers will be

	E_i
Monday	14279
Tuesday	14279
Wednesday	14279
Thursday	14279
Friday	14279
Saturday	14279
Sunday	14279

The statistic thus becomes

$$\chi^2 = \sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i} = 113.3$$

This must be compared with the χ^2 distribution with $df = 7-1-1 = 5$ degrees of freedom (since we have estimated one parameter λ from the observations). We get a significance probability

$$P = 1 - F_{\chi^2}(113.3)$$

far below 0.1%, and can clearly reject the hypothesis that the numbers were from the same distribution.

Solution of Problem 63
What we have is a contingency table with two rows and two columns, i.e. a 2×2 table. In total, there are $N = 223$ observations. If there is independence between rows and columns, the expected

numbers will be

	pro	contra
Danes	63.9	28.1
Swedes	91.1	39.9

since the expected number in, say, the upper left cell is

$$E_{11} = \frac{R_1 S_1}{N} = \frac{92 \cdot 155}{223} = 63.9$$

The statistic thus becomes

$$\chi^2 = \left(\frac{70 \cdot 46 - 22 \cdot 85}{223}\right)^2 \left(\frac{1}{63.9} + \frac{1}{28.1} + \frac{1}{91.1} + \frac{1}{39.9}\right) = 3.20$$

The significance probability of the test thus becomes

$$P = 1 - F_{\chi^2}(3.20) \approx 20\%$$

where we have used the χ^2 distribution with $df = 1$ degree of freedom. Since P is greater than 5%, we *cannot* detect any regional dependence in the opinion poll.

Solution of Problem 64

It is reasonable to perform a one-sided test, i.e. to test the null hypothesis

$$\mathbf{H}_0 : \text{no effect of medicine}$$

against the alternative hypothesis

$$\mathbf{H}_1 : \text{positive effect of medicine}$$

Given $\mathbf{H}_0$ the expected numbers become

	fit	ill
medicine	5	15
placebo	5	15

We now calculate the one-sided statistic

$$u = \left(\frac{8 \cdot 18 - 2 \cdot 12}{40}\right) \sqrt{\left(\frac{1}{5} + \frac{1}{15} + \frac{1}{5} + \frac{1}{15}\right)} = 2.19$$

Given $\mathbf{H}_0$, u will be standard normally distributed, whereas a positive value of u will be expected given $\mathbf{H}_1$. The significance probability thus becomes

$$1 - \Phi(2.19) = 1.4\%$$

and thereby we reject $\mathbf{H}_0$ in favour of $\mathbf{H}_1$.

Solution of Problem 65
Let us use a one-sided test, i.e. test the null hypothesis

$$\mathbf{H}_0 : \text{ women watch as much football as men}$$

against the alternative hypothesis

$$\mathbf{H}_1 : \text{ men watch more football than women}$$

If we present the observations in a 2×2 table, it looks like this:

	yes	no
men	3	0
women	0	3

Given $\mathbf{H}_0$, the expected numbers are

	yes	no
men	1.5	1.5
women	1.5	1.5

Since these are less than 5, we *cannot* use a χ^2-test. Instead we use Fisher's exact test. The significance probability thus becomes

$$P_{\text{Fisher}} = \frac{3!3!3!3!}{6!3!0!0!3!} = 5\% \text{ (exact!)}$$

Therefore, we can reject $\mathbf{H}_0$ at significance level 5%.

Solution of Problem 66

We have to use Wilcoxon's test for one set of observations. First we calculate the differences $d_i = x_i - y_i$ and assign a *rank* to each of them:

Plant no.	x_i	y_i	d_i	rank
1	41	27	14	9
2	51	59	–8	5
3	66	76	–10	7.5
4	68	65	3	2
5	46	36	10	7.5
6	69	54	15	10
7	47	49	–2	1
8	44	51	–7	4
9	60	55	5	3
10	44	35	9	6

Note: since no. 3 and no. 5 have the same numerical value, they were both given their average rank.

We now calculate the two statistics:

$$t_+ = 9 + 2 + 7.5 + 10 + 3 + 6 = 37.5 \text{ and } t_- = 5 + 7.5 + 1 + 4 = 17.5$$

At this point we can check that $t_+ + t_- = 55 = 10 \cdot (10 + 1)/2$, which shows that we have computed correctly.

If the pesticide has had an effect, the d_i's should be mainly positive. So we have to investigate whether t_- is "extremely" small. Table C.8 shows that the significance probability is considerably more than 5%. Therefore, we cannot prove any significant effect of the pesticide.

Solution of Problem 67

We can use the normal approximation and seek the standard normally distributed statistic

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

Since

$$\mu = \frac{100 \cdot 101}{4} = 2525 \text{ and } \sigma = \sqrt{\frac{100 \cdot 101 \cdot 201}{24}} = 290.8$$

we get

$$z = 1.79$$

The significance probability thereby becomes

$$1 - \Phi(1.79) \approx 4\%$$

and we can conclude that the pesticide *has* had a significant effect.

Solution of Problem 68

We have to use Wilcoxon's test for two sets of observations. So we have $n = 14$ observations in the first set and $m = 11$ observations in the second set. Each observation is given a rank between 1 and $n + m = 25$:

Car no.	Line 1	Line 2
1	18	7
2	23	11
3	3	22
4	15	13
5	9	17
6	2	21
7	20	4
8	1	16
9	10	8
10	19	24
11	6	25
12	14	
13	5	
14	12	

The statistic t_x is the sum of the n ranks corresponding to Line 1:

$$t_x = 18 + 23 + 3 + 15 + 9 + 2 + 20 + 1 + 10 + 19 + 6 + 14 + 5 + 12 = 157$$

Since we are not testing against any particular alternative hypothesis, we have to consider the minimum

$$t := \min\{t_x, n(n+m+1) - t_x\} = \min\{157, 207\} = 157$$

We look up in Table C.9 under $n = 14$ and $m = 11$ and find the number 151. Since t is *not* less than the table value, we *cannot* prove any significant difference between the number of defects at significance level $\alpha = 10\%$.

Solution of Problem 69
We may use the normal approximation to Wilcoxon's test for two sets of observations. We know that t_x is normally distributed with expected value

$$\mu = \frac{n(n+m+1)}{2} = 7550$$

and standard deviation

$$\sigma = \sqrt{\frac{nm(n+m+1)}{12}} = \sqrt{62917} = 251$$

So we have to look at the standard normally distributed statistic

$$z = \frac{t_x - \mu}{\sigma} = -2.18$$

This gives a significance probability of

$$\Phi(-2.18) = 1.5\%$$

We see that there were significantly fewer defects in Line 1.