

Uvod v statistiko

Janez Stare

Medicinska fakulteta, Ljubljana

Ljubljana, 2018

OBVEZNOSTI

1. domače naloge
2. izpit

Neobvezno, a zelo pohvalno :)

Najdete neko statistično neumnost v časopisu, na spletu, na TV ...

Statistika je povsod okrog nas

Statistika je povsod okrog nas

Šport - koliko zadetkov, kotov, strelov, ...

Statistika je povsod okrog nas

Šport - koliko zadetkov, kotov, strelov, ...

Promet - nesreč, alkohola, ...

Statistika je povsod okrog nas

Šport - koliko zadetkov, kotov, strelov, ...

Promet - nesreč, alkohola, ...

Politika - kdo bi koga, kdo je koga

Statistika je povsod okrog nas

Šport - koliko zadetkov, kotov, strelov, ...

Promet - nesreč, alkohola, ...

Politika - kdo bi koga, kdo je koga (mislim volil)

Statistika je povsod okrog nas

Šport - koliko zadetkov, kotov, strelov, ...

Promet - nesreč, alkohola, ...

Politika - kdo bi koga, kdo je koga (mislim volil)

Demografija, ekonomija, zdravstvo ... - rojstva, inflacija, obolenost ...

Statistika je povsod okrog nas

Šport - koliko zadetkov, kotov, strelov, ...

Promet - nesreč, alkohola, ...

Politika - kdo bi koga, kdo je koga (mislim volil)

Demografija, ekonomija, zdravstvo ... - rojstva, inflacija, obolenost ...

Statistika je osnovno orodje raziskovanja na (skoraj) vsakem področju znanosti (vam pride na misel kakšna izjema?)

Statistika je povsod okrog nas

Šport - koliko zadetkov, kotov, strelov, ...

Promet - nesreč, alkohola, ...

Politika - kdo bi koga, kdo je koga (mislim volil)

Demografija, ekonomija, zdravstvo ... - rojstva, inflacija, obolenost ...

Statistika je osnovno orodje raziskovanja na (skoraj) vsakem področju znanosti (vam pride na misel kakšna izjema?)

...

Šale na račun statistike

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

- ▶ Obstajajo laži, velike laži in statistika.

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

- ▶ Obstajajo laži, velike laži in statistika.
- ▶ S statistiko je mogoče vse dokazati.

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

- ▶ Obstajajo laži, velike laži in statistika.
- ▶ S statistiko je mogoče vse dokazati.
- ▶ Trije od štirih ...

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

- ▶ Obstajajo laži, velike laži in statistika.
- ▶ S statistiko je mogoče vse dokazati.
- ▶ Trije od štirih ...
- ▶ Povprečni človek ima ...

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

- ▶ Obstajajo laži, velike laži in statistika.
- ▶ S statistiko je mogoče vse dokazati.
- ▶ Trije od štirih ...
- ▶ Povprečni človek ima ...
- ▶ Glava v peči ...

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

- ▶ Obstajajo laži, velike laži in statistika.
- ▶ S statistiko je mogoče vse dokazati.
- ▶ Trije od štirih ...
- ▶ Povprečni človek ima ...
- ▶ Glava v peči ...
- ▶ Kot ljudje, če jih dovolj dolgo ...

Šale na račun statistike

Največ, kar večina ljudi ve o statistiki, je kakšna šala.

- ▶ Obstajajo laži, velike laži in statistika.
- ▶ S statistiko je mogoče vse dokazati.
- ▶ Trije od štirih ...
- ▶ Povprečni človek ima ...
- ▶ Glava v peči ...
- ▶ Kot ljudje, če jih dovolj dolgo ...
- ▶ Meni najljubša definicija statistika:
Statistik je človek, ki ne verjame, da je Kolumb odkril Ameriko, ker tega ni bilo v načrtu raziskave.

Po drugi strani pa ...

It's amazing how authoritative you can sound just by quoting some statistics ...

In prav gotovo

Without data it is anyone's opinion . . .

(In God we trust. All others must bring data).

Kaj statistika NI

Kaj statistika NI

Matematik, fizik in statistik so šli lovit jelena.

Kaj statistika NI

Matematik, fizik in statistik so šli lovit jelena.

Enega opazijo in matematik strelja prvi - zgreši za meter v levo.

Kaj statistika NI

Matematik, fizik in statistik so šli lovit jelena.

Enega opazijo in matematik strelja prvi - zgreši za meter v levo.

Potem poskusi fizik - meter v desno.

Kaj statistika NI

Matematik, fizik in statistik so šli lovit jelena.

Enega opazijo in matematik strelja prvi - zgreši za meter v levo.

Potem poskusi fizik - meter v desno.

Statistik pa ves srečen zavpije:

Kaj statistika NI

Matematik, fizik in statistik so šli lovit jelena.

Enega opazijo in matematik strelja prvi - zgreši za meter v levo.

Potem poskusi fizik - meter v desno.

Statistik pa ves srečen zavpije:

“PA SMO GA!”

Kaj statistika NI

Matematik, fizik in statistik so šli lovit jelena.

Enega opazijo in matematik strelja prvi - zgreši za meter v levo.

Potem poskusi fizik - meter v desno.

Statistik pa ves srečen zavpije:

“PA SMO GA!”

Statistika NI veda o povprečjih!!

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Španija ima več časa žogo in več strelov na gol.

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Španija ima več časa žogo in več strelov na gol.

Komentator: "**Vse te statistike popolnoma nič ne pomenijo!**"

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Španija ima več časa žogo in več strelov na gol.

Komentator: "**Vse te statistike popolnoma nič ne pomenijo!**"

Drug dan, druga tekma, isti komentator. Rezultat je 0:0, vendar ena ekipa dosti več napada.

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Španija ima več časa žogo in več strelov na gol.

Komentator: "**Vse te statistike popolnoma nič ne pomenijo!**"

Drug dan, druga tekma, isti komentator. Rezultat je 0:0, vendar ena ekipa dosti več napada.

Komentator: "**Gol visi v zraku!**"

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Španija ima več časa žogo in več strelov na gol.

Komentator: "**Vse te statistike popolnoma nič ne pomenijo!**"

Drug dan, druga tekma, isti komentator. Rezultat je 0:0, vendar ena ekipa dosti več napada.

Komentator: "**Gol visi v zraku!**"

Zakaj to reče? Na osnovi česa?

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Španija ima več časa žogo in več strelov na gol.

Komentator: "**Vse te statistike popolnoma nič ne pomenijo!**"

Drug dan, druga tekma, isti komentator. Rezultat je 0:0, vendar ena ekipa dosti več napada.

Komentator: "**Gol visi v zraku!**"

Zakaj to reče? Na osnovi česa?

In mimogrede: Španija je postala svetovni prvak

Kaj statistika JE

Primer (nogomet): kako TV komentator razume statistiko

Tekma Švica : Španija (1:0)

Španija ima več časa žogo in več strelov na gol.

Komentator: "**Vse te statistike popolnoma nič ne pomenijo!**"

Drug dan, druga tekma, isti komentator. Rezultat je 0:0, vendar ena ekipa dosti več napada.

Komentator: "**Gol visi v zraku!**"

Zakaj to reče? Na osnovi česa?

In mimogrede: Španija je postala svetovni prvak

Bistvo: **trditve** v statistiki so **verjetnostne!!**

Tole je pa res ...



"Drugače povedano, statistika dokazuje, da statistiki nijamo vedno prav."

Seveda se verjetnost lahko slabo razume ...

Seveda se verjetnost lahko slabo razume ...

Študent statistike je vedno pospešil pred križiščem, peljal hitro skozi in potem spet upočasnil. Nekega sopotnika je to iritiralo in je vprašal, zakaj to počne. Študent je odgovoril:

Seveda se verjetnost lahko slabo razume ...

Študent statistike je vedno pospešil pred križičem, peljal hitro skozi in potem spet upočasnil. Nekega sopotnika je to iritiralo in je vprašal, zakaj to počne. Študent je odgovoril:

“Statistično gledano so nesreče mnogo bolj verjetne v križičih, zato skušam tam prebiti čim manj časa.”

Doctors are often unable to explain exactly why one person gets cancer and another doesn't. The researchers speculate that "a diet rich in refined cereals and poor in vegetables may have an unfavorable role on RCC [renal cell carcinoma]."

Study Links Bread, Kidney Cancer Risk

Those Without Kidney Cancer Ate More Vegetables And Less Bread

<http://www.cbsnews.com/stories/2006/10/20/health/webmd/main2111478.shtml>

[Video](#)[U.S.](#)[World](#)[Politics](#)[SciTech](#)[Health](#)[WebMD](#)[Healthy Living](#)[Entertainment](#)[Business](#)[CBS Investigates](#)[Sports](#)[Strange](#)[Travel](#)[Opinion](#)[Blogs](#)[In-Depth Photos](#)[Puzzles & Toons](#)[Mobile Services](#)[E-Mail Services](#)[RSS Feeds](#)[Podcasts](#)[Get Widgets](#)SEARCH CBS News

GO • Tips

Home > Health > WebMD

Study Links Bread, Kidney Cancer Risk

These Without Kidney Cancer Ate More Vegetables And Less Bread

Oct. 20, 2006

[EMAIL STORY](#)[PRINT STORY](#)

SHARE

TEXT SIZE: A A A



(CBS/AP)

Double-click any word ([what's this?](#))

(WebMD) An Italian study shows that people with renal cell carcinoma, the most common type of kidney cancer, may eat more bread and fewer vegetables than those without kidney cancer.

But the study, published online in the International Journal of Cancer, doesn't claim bread causes kidney cancer.

The researchers included Francesca Bravi, M.D., of the Istituto di Ricerche Farmacologiche "Mario Negri" in Milan.

Between 1992 and 2004, Bravi's team interviewed 1,000 patients with renal cell carcinoma at Italian hospitals. They also interviewed 1,534 patients without kidney cancer. Patients completed surveys about their diets during the previous two years. The questions covered 78 foods and beverages.

The findings show renal cell carcinoma patients were more likely than those without kidney cancer to have the highest intake of bread, and, to a lesser extent, pasta and rice. People without renal cell carcinoma were more likely to eat the greatest amount of vegetables, poultry, and processed meats.

The researchers found no association between renal cell carcinoma and coffee, tea, soups, eggs, red meat, fish, cheese, potatoes, fruit, desserts, or sugars.

RELATED



INTERACTIVE

Diet And Nutrition

Are you eating right? See the government's guidelines, calculate your body mass index and quiz yourself on healthy food choices.

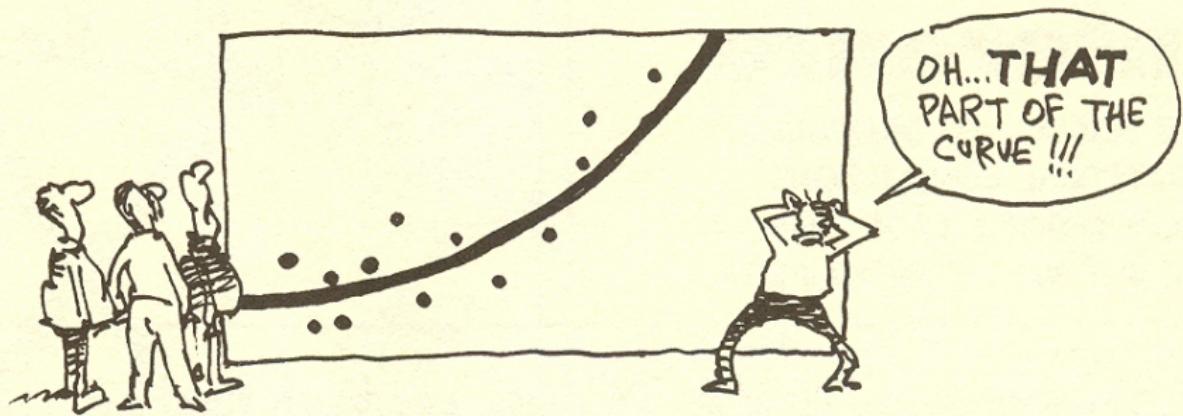


INTERACTIVE

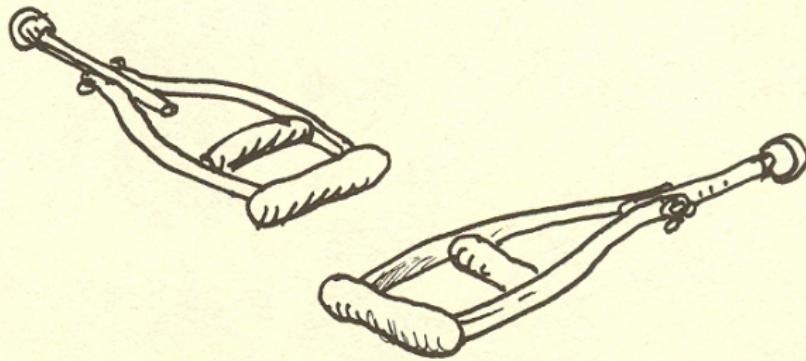
Food Pyramid

The government's latest guidelines for healthy eating get personal.

FOR EXAMPLE, IN 1986, THE SPACE SHUTTLE CHALLENGER EXPLODED, KILLING SEVEN ASTRONAUTS. THE DECISION TO LAUNCH IN 29-DEGREE WEATHER HAD BEEN MADE WITHOUT DOING A SIMPLE ANALYSIS OF PERFORMANCE DATA AT LOW TEMPERATURE.



A MORE POSITIVE EXAMPLE IS THE SALK POLIO VACCINE. IN 1954, VACCINE TRIALS WERE PERFORMED ON SOME 400,000 CHILDREN, WITH STRICT CONTROLS TO ELIMINATE BIASED RESULTS. GOOD STATISTICAL ANALYSIS OF THE RESULTS FIRMLY ESTABLISHED THE VACCINE'S EFFECTIVENESS, AND TODAY POLIO IS ALMOST UNKNOWN.



Top 11 contributions to medicine over the millennium (NEJM, 2000)

- Elucidation of Human Anatomy and Physiology
- Discovery of Cells and Their Substructures
- Elucidation of the Chemistry of Life
- Application of Statistics to Medicine
- Development of Anaesthesia
- Discovery of the Relation of Microbes to Disease
- Elucidation of Inheritance and Genetics
- Knowledge of the Immune System
- Development of Body Imaging
- Discovery of Antimicrobial Agents
- Development of Molecular Pharmacotherapy

Top 11 contributions to medicine over the millennium (NEJM, 2000)

Elucidation of Human Anatomy and Physiology

Discovery of Cells and Their Substructures

Elucidation of the Chemistry of Life

Application of Statistics to Medicine

Development of Anaesthesia

Discovery of the Relation of Microbes to Disease

Elucidation of Inheritance and Genetics

Knowledge of the Immune System

Development of Body Imaging

Discovery of Antimicrobial Agents

Development of Molecular Pharmacotherapy



The New England Journal of Medicine

Established in 1812 as THE NEW ENGLAND JOURNAL OF MEDICINE AND SURGERY

VOLUME 342

JANUARY 6, 2000

NUMBER 1

ORIGINAL ARTICLES

- The Relation between Blood Pressure and Mortality Due to Coronary Heart Disease among Men in Different Parts of the World 1
 P.C.W. VAN DEN HOOGEN AND OTHERS

- Noninvasive Diagnosis by Doppler Ultrasoundography of Fetal Anemia Due to Maternal Red-Cell Alloimmunization 9
 G. MARI

- Group B Streptococcal Disease in the Era of Intrapartum Antibiotic Prophylaxis 15
 S.J. SCHREAG AND OTHERS

- Brief Report: Paraneoplastic Cerebellar Ataxia Due to Autoantibodies against a Glutamate Receptor 21
 P.S. SMITH AND OTHERS

- IMAGES IN CLINICAL MEDICINE**
Histoplasma capsulatum in a Peripheral-Blood Smear 28
 M. EDELMAN AND J. MCKITTRICK

- REVIEW ARTICLE**
 Primary Care: Avoiding Pitfalls in the Diagnosis of Subarachnoid Hemorrhage 29
 J.A. EDLOW AND L.R. GAFLAN

CLINICAL PROBLEM-SOLVING

- Inpatient Inpatient Care 37
 M. GULATI, S. SAINT, AND L.M. TIERNEY, JR.

EDITORIALS

- Looking Back on the Millennium in Medicine 43
 THE EDITORS

- Blood Pressure and the Risk of Cardiovascular Disease 50
 S. MACMAHON

- Noninvasive Testing for Fetal Anemia 52
 G.R. SAADE

- INFORMATION FOR AUTHORS** 54

CORRESPONDENCE

- Oral Antibiotics for Febrile Patients with Neutropenia Due to Cancer
 Chemotherapy 55
 Clinical Effect of Grass-Pollen Immunotherapy 58
 West Nile Viral Encephalitis in an HIV-Positive Woman in New York 59
 Molecular Diagnosis of Familial Mediterranean Fever 60
 Empysematosus Pyelonephritis 60
 Anthrax 61

- BOOK REVIEWS** 63

- BOOKS RECEIVED** 65

- NOTICES** 66

CENTRALNA MEDICINSKA KNJIZINICA

E145-146
New Engl J Med

1220 342 1

145-146
New Engl J Med
1220 342 1
CORR 8

© 2000 by the Massachusetts Medical Society
 published by the Massachusetts Medical Society and printed in England
 (Archver) Ltd., West Portway, Andover, SP10 3SE, U.K.
 as a newspaper at the Post Office. ISSN 0028-4793

Editorials

LOOKING BACK ON THE MILLENNIUM IN MEDICINE

THE second millennium is over. The editors of the *Journal* first thought to ignore this passage. After all, the changing of the millennium would no doubt be the subject of incessant media attention. Why should we add to it? Yet, looking back, it is hard not to be moved by the astounding course of medical history over the past thousand years. No one alive in the year 1000 could possibly have imagined what was in store. Furthermore, medicine is one of the few spheres of human activity in which the purposes are unambiguously altruistic — in itself, a remarkable achievement.

We therefore decided to yield to the temptation to comment on the end of the second millennium by choosing the most important medical developments of the past thousand years and reviewing them briefly. None of the developments we selected was an isolated discovery or event; instead, each was a series of notable steps — some huge, some smaller — along a path that led to a crucial body of knowledge in a particular area. That is the usual way medical science progresses. For example, Vesalius took giant steps toward elucidating human anatomy, but he was not alone, and what was important was the totality of the work in that area.

We deliberately restricted ourselves to developments that changed the face of clinical medicine, not preventive medicine or public health or health care delivery or medical ethics. Yet there is obviously overlap. Understanding the relation of microbes to disease, for example, inevitably affected not only clinical medicine, but also preventive medicine and public health. Indeed, it is hard to think of a more important advance in all three areas than immunization. Medical ethics has become increasingly important as the power of clinical medicine grows, but we arbitrarily decided not to include that topic here.

Except for some early work by the ancient Greeks, much of it wrong, there were few advances in clinical medicine until the Renaissance. In the 1400 years between Galen and Vesalius, medicine was stagnant, dominated by the belief that illness reflected an imbalance in the four humors of the body — blood, phlegm, yellow bile, and black bile. Life was nasty, brutish, and short, and medical care did not help. There are many reasons little progress was made until the Renaissance, but one of them was surely that the only fit pursuit for scholars in those centuries was considered to be knowledge of God, not of man. Only with the flowering of humanism that characterized the Renaissance did that change, and it changed very rapidly.

Readers will note that the developments we discuss

were the work largely of white men in Europe and North America. For a variety of reasons, that is the way it was. In the new millennium, it will be different. That is one prediction we make with confidence. The other is that the pace of change will continue to accelerate, as it did in the second millennium. Beyond this, it would be foolhardy to speculate about what the new millennium holds, just as it would have been impossible for anyone in the year 1000 to dream of everything that was to come.

Here, then, we present our choices for the most important medical developments of the past millennium. In what may be our only claim to distinction in the process, we arbitrarily chose 11, not 10. Obviously, many more could have been selected. We present them not in order of importance, but in rough chronological order according to the first noteworthy step taken in a given area.

ELUCIDATION OF HUMAN ANATOMY AND PHYSIOLOGY

The emergence of a comprehensive understanding of the structure and function of the organ systems of the human body stands — without question — as one of the most influential advances of the past millennium. Although the contributions of the Greek physician Galen early in the first millennium A.D. were extraordinarily important to anatomy and physiology, Galen also introduced numerous errors that were not corrected until the Renaissance. Perhaps the greatest anatomist of the Renaissance, if not of all time, was Andreas Vesalius (1514–1564), born in Brussels but educated in France and Italy. Vesalius's anatomical treatise, *De Humani corporis fabrica libri septem* ("Seven Books on the Structure of the Human Body"), published in 1543, is regarded as one of the most important works in medicine. The extraordinary illustrations in the *Fabrics* (not actually drawn by the great anatomist but by an unknown artist) set a new standard for the understanding of human anatomy.

Less than 100 years after Vesalius, William Harvey (1578–1657), an English physician and physiologist, established that the blood circulates within a closed system, with the heart serving as a pump. He showed that the pulse results from the filling of arteries with blood after cardiac contraction and that the right ventricle pumps blood to the pulmonary circulation and the left ventricle pumps blood to the systemic circulation. The importance of Harvey's work, published in 1628 in *Exercitatio anatomica de motu cordis et sanguinis in animalibus* ("On the Motion of the Heart and Blood in Animals"), cannot be overstated. The physiologic principles that he established led to an un-

From the standpoint of medical practice, the growth of knowledge about the inorganic composition of body fluids is probably just as important to the amassing of knowledge about the organic chemistry of cells. The relation of sodium to edema or dehydration, the importance of potassium in the losses incurred in diarrhea, the distribution of water in the body, and the implications of the disturbances in acid-base balance that accompany vomiting, circulatory shock, uremia, or uncontrolled diabetes — all were clarified during the past hundred years and have become part of the basic knowledge required by doctors in every specialty for the delivery of good medical care.

APPLICATION OF STATISTICS TO MEDICINE

A natural starting point for a history of biostatistical thought in the past millennium is the work of Leonardo Fibonacci (c. 1170—after 1240), an Italian mathematician of the Middle Ages. By introducing Indian and Arabic mathematics and numbering to Europe in 1202, he freed Western thought from the limitations of the Roman-numeral system. This advance laid the foundation for modern computation and bookkeeping. Probability theory emerged only in the 16th and 17th centuries, when Pierre de Fermat (1601–1665) and Blaise Pascal (1623–1662) developed basic probabilistic calculations to analyze games of chance. Ideas of relative frequency were first applied to mortality statistics in 17th-century London at the time of the plague. John Graunt (1620–1674) introduced the notion of inference from a sample to an underlying population and described calculations of life expectancy that launched the insurance industry in the 17th and 18th centuries.

The German mathematician Karl Friedrich Gauss (1777–1855) played a central part in the development of modern statistical reasoning. His method of least-squares analysis, developed around 1794, underlies much of modern regression analysis. Thomas Bayes (1702–1761), the 18th-century English theologian and mathematician, was the first to show how probability can be used in inductive reasoning.

One of the earliest clinical trials took place in 1747, when James Lind treated 12 scurvy-prone ship passengers with cider, an elixir of vitriol, vinegar, sea water, oranges and lemons, or an electuary recommended by the ship's surgeon. The success of the citrus-containing treatment eventually led the British Admiralty to mandate the provision of lime juice to all sailors, thereby eliminating scurvy from the navy. The origin of modern epidemiology is often traced to 1854, when John Snow demonstrated the transmission of cholera from contaminated water by analyzing disease rates among citizens served by the Broad Street Pump in London's Golden Square. He arrested the further spread of the disease by removing the pump handle from the polluted well.

Biostatistical reasoning developed rapidly in Great

Britain in the late 19th and early 20th centuries. Sir Ronald Fisher (1890–1962), the most important figure in modern statistics, developed the analysis of variance and multivariate analysis. He also introduced the principle of randomization as a method for avoiding bias in experimental studies. In the United States, Jerzy Neyman, a Russian immigrant, developed the theories of estimation and testing that shaped contemporary biostatistical practice.

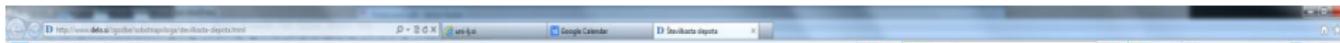
A landmark of quantitative observational research as a tool for exploring the determinants of disease was Sir Richard Doll's study of smoking among British physicians. Randomized clinical trials emerged in England in the 1950s and were adopted by the National Institutes of Health in the United States in the early 1960s; these followed an explosion of clinical trials of treatment for cancer, heart disease, diabetes, and other diseases. Biostatistical methods expanded rapidly during this period. Sir David Cox's 1972 paper on proportional-hazards regression ignited the fields of survival analysis and semiparametric inference (using partial specification of the probability distribution of the outcomes under investigation). Rapid improvements in computer support were essential to the growing role of empirical investigation and statistical inference.

DEVELOPMENT OF ANESTHESIA

Archaeological evidence makes it clear that surgery was practiced in the form of trephination of the skull well before recorded history. Some who suffered through the procedure even survived it. Written records from ancient Greece, Egypt, and China refer to the use of opium, cannabis, and mandragora (mandrake) to produce anesthesia, analgesia, and amnesia. It is clear, however, that for most of recorded history surgical procedures were crude, quick, and agonizing. Surgery was a fearsome treatment of last resort, rarely used. The development of anesthesia was the essential prelude to modern surgery.

The European scientific establishment laid the groundwork for the development of surgical anesthesia. In 1799, Sir Humphry Davy, the superintendent of the Pneumatic Institution in Clifton, England, recognized the analgesic properties of nitrous oxide when he inhaled it, during the course of his work, while he had a toothache. He coined the term "laughing gas" but carried the work no further.

Ether had been known to chemists since the 18th century, and chloroform was discovered in 1831, but the medical applications of inhaled agents to relieve the pain of surgery came about only after Horace Wells, a Connecticut dentist, used nitrous oxide to anesthetize 15 patients during December 1844. Flush with success, Wells persuaded his former partner, William Morton, to arrange a public exhibition of nitrous oxide anesthesia for a dental extraction at the Massachusetts General Hospital. The demonstration, in January 1845, was a disaster. The patient cried out in pain,



Prijava omogoča neomejen dostop do vseh vsebin.

Delo Slovenije novice Delo in delo, Tisk, Projekti, Zaposleni, Naučnilica, Vesti, Opis, in Tiskovina

DELO

Ljubljana, 16.02.2013

Novice Gospodarstvo Šport Kultura Družba Mnenja D zgodbe Tudi ti

zgodbe > Tiskovina prejela Revolka spletka

Ljubljana J. Kucić, Slobotna priloga
objavljeno: nov. 17.02.2013, 06:00

Številkasta slepota

G. H. Wells je dejal, da bo statistična pomenost postal najmanj tako pomembna za aktivno državljanstvo kot branje in pišanje.

Sklipna nemška ginekologinja, radiofizičarka, onkologinja in drugi zdravniki, ki sodelujejo pri izpodjem odkrivanja raka na doju, je morala rešiti preprost stresnjeni matematični vezeninski račun.

Izvedeli so, da se bolezni razvijejo pri približno 0,8 odstotku Nemik, starih do 45 do 50 let. Če ena ženska raka, ga bo preventivni manjman potrošen tudi pri sedmici odstotku žensk, ki raka nima. Zato se morali zdravniki izračunati, kolikšna je verjetnost, da se je pri pacienti s pozitivnim izsledkom v resniku razvila bolezni.

V poslanku, ki ga je vodil nemški pravnik in raziskovalnik, Gerd Gigerenzer, je sodelovalo 48 izbranih zdravnikov z različnimi petanjami in klinične prakse ter jih je ospal v knjigo *Reassessing With Risk*. Njihovi odgovori so izrazno različni. Vendar pa je vredno pojasniti, da je verjetnost, da je pacient 50-odstotno ozoren udruževanje je verjetnost ocenjeno na manj kot deset odstotkov. V pogovoru so združeno verjetnost raka ocenila na 70 odstotkov, pravilen rezultat pa sta izsledovali in ebiti oskrbi en sam deset odstotkov s postavljenim izvenom.

Sklip je zamenjal za zagodbilo zelo pogosto napako. Zagodbila je v vsakem zaporedju medicinskega testa, ki preverja zaznava raka pri bozni pacientih (nisi odstotek), zato niso uporabljali drugih dveh potiskov. Koliko ženski je dejansko bolnih in kakšno je verjetnost pozitivnega testa pri zdravi ženski (pri preventivnih pregledih je velika verjetnost pozitivne rezultat). Če se preventivni test rezultira pozitivno, pa je verjetnost, da bo bolezen skupaj, je samo zmanjšan zelenček (0,5 odstotku). Kot je združen ženskam pa je test potiskal tudi pozitiven rezultat, za katerega so celo izbrani zdravniki prepričani, da je skupaj zagotovo pozitiven.

Gigerenzer je v poslanku izpostavljal, da je zadružen nacionalnih zavodovarjev strategija, ki ne uporablja klasificacijske napovedi. Preventivni pregledi, ki jih spodbujajo zdravniki, medju in zvezne mestnosti, so zelo konzistenti za uporabo spoznavanja bolnikov in tako izboljšajo operativnost zdravstva, če zato izpoljujejo istake očevnosti brezvega in koristi. Vendar je v levičarski potencialni poskuški pokazati, da se zdravniki bolj izvlečajo v interesne teme, kot so napredek v zdravstvu, zdravstveni sistem, zdravstveni akademiji, katerih je vstopanje manjšinsko, presegajočo za vse in zasegljivi klicniki testa. Zato so napredek ostre in uvedoma preprečeni zato, ker napredujejo preventivni ukrepi, pozivajo veliko preprečevati stanje, naprejitev

REVOLT ALTERNATIVE

IDEJE BOLJŠI JUTRI

revolt@delo.si

Kaj je bilo predsednik vlade povedati?

Na ta vprašanje je Pretresnik odgovoril:

O postance in Pretresnik: zvezna

Dejavnja in razenje in univerzitetni retorčni

Na to vprašanje je Pretresnik odgovoril:

O dejanju Dr. Božeta Žiberna in zveznem in preventivnem delu na UPZ 250 (pa je tam) dejanja igrala D. Žiberna.

Ljubljana je najboljše mesto proti boležni

Na to vprašanje je Pretresnik odgovoril:

Kazino je v Ljubljani, vendar ne živijo vse ljudje proti kazinu. Torej ne.

DANGER



Lenart J. Kučić, Sobotna priloga

sobotnapRILOGA
ned, 17.02.2013, 06:00

Številkasta slepota

G. H. Wells je dejal, da bo statistična pismenost postala najmanj tako pomembna za aktivno državljanstvo kot branje in pisanje.

Citat iz članka

Goldacre je pojasnil, da medicinske fakultete študentov ne opremijo z znanjem, ki je potrebno za kritično branje strokovne literature: rezultatov kliničnih in populacijskih raziskav, ki ugotavljajo učinkovitost določenega zdravila ali zdravstvenega priporočila.

Data

Data

- ▶ Information gathering is at the heart of all sciences.

Data

- ▶ Information gathering is at the heart of all sciences.
- ▶ There are different information-gathering techniques. These provide **observations** on the units of interest.

Data

- ▶ Information gathering is at the heart of all sciences.
- ▶ There are different information-gathering techniques. These provide **observations** on the units of interest.
- ▶ Examples of such techniques in social sciences are: questionnaire surveys, telephone surveys, planned experiments, direct observation of behaviour in natural settings.

Data

- ▶ Information gathering is at the heart of all sciences.
- ▶ There are different information-gathering techniques. These provide **observations** on the units of interest.
- ▶ Examples of such techniques in social sciences are: questionnaire surveys, telephone surveys, planned experiments, direct observation of behaviour in natural settings.
- ▶ Additionally, information gathered for other purposes may be used (police records, census, ...)

Data

- ▶ Information gathering is at the heart of all sciences.
- ▶ There are different information-gathering techniques. These provide **observations** on the units of interest.
- ▶ Examples of such techniques in social sciences are: questionnaire surveys, telephone surveys, planned experiments, direct observation of behaviour in natural settings.
- ▶ Additionally, information gathered for other purposes may be used (police records, census, ...)
- ▶ Such observations are called **data**. Data then consist of measurements on the characteristics of interest. Examples are: party affiliation, annual income, marital status, opinion on some issue.

What is Statistics?

What is Statistics?

- ▶ **Statistics** is the science - and sometimes art - of collecting and processing data, summarizing information, estimating descriptive constants (parameters), discovering empirical laws, testing hypotheses and making inferences, and designing experiments in such a way that valid inferences can be drawn from empirical evidence.

What is Statistics?

- ▶ **Statistics** is the science - and sometimes art - of collecting and processing data, summarizing information, estimating descriptive constants (parameters), discovering empirical laws, testing hypotheses and making inferences, and designing experiments in such a way that valid inferences can be drawn from empirical evidence.
- ▶ **Inferential statistics** is a branch of statistics that consists of generalizing from samples to populations, performing hypothesis testing, determining relationships among variables, and making predictions.

What is Statistics?

- ▶ **Statistics** is the science - and sometimes art - of collecting and processing data, summarizing information, estimating descriptive constants (parameters), discovering empirical laws, testing hypotheses and making inferences, and designing experiments in such a way that valid inferences can be drawn from empirical evidence.
- ▶ **Inferential statistics** is a branch of statistics that consists of generalizing from samples to populations, performing hypothesis testing, determining relationships among variables, and making predictions.
- ▶ Sometimes Statistics is divided into **descriptive** and **inferential statistics**, as if statistics only starts when the data are already collected. This view is false, and one should remember that the design of experiments, and data collecting and processing are all very important statistical methods. Improperly collected data lead to worthless conclusions, no matter how good statistical analysis.

Example 1: Opinion of the Tea Party supporters on the 14th amendment of the US constitution

Question: Do voters of the Tea Party have a different opinion on the 14th amendment than the general population?

Assumption: we know the general population opinion.

Task: Design research to answer the above question!

Preferable solution: Select a random sample of Tea Party supporters, and ask them about their opinion on the 14th amendment.

Example 2: Opinion on the 14th amendment of the US constitution

Question: Do voters of different political parties have different opinions on the 14th amendment?

Task: Design research to answer the above question!

Preferable solution: Select a random sample of US citizens, and ask them about their opinion on the 14th amendment and about their voting preference.

Results: 14th amendment

The 14th amendment of the US Constitution guarantees American citizenship to any person born in the United States. Do you think this is a good or bad law?

	Total	Party ID			Tea Party Support	Political Philosophy		
		Rep. %	Dem. %	Ind. %		Cons. %	Mod. %	Lib. %
	%							
Good (NET)	66	58	76	66	56	53	69	84
Very good	40	27	52	39	28	29	39	60
Somewhat good	27	31	23	27	28	24	30	23
Bad (NET)	27	38	18	25	40	41	24	12
Somewhat bad	17	22	12	14	23	21	17	8
Very bad	10	15	6	11	17	19	7	4
Not at all sure	6	4	6	9	4	7	7	4

Pregnant foreigners and birthing trips

Some pregnant foreigners arrange trips to the United States, specifically timed so that they give birth during their stay, making any child born an automatic US citizen. Do you think Constitution should be changed to no longer allow for this?

	Total	Party ID			Tea Party Support	Political Philosophy		
		Rep. %	Dem. %	Ind. %		Cons. %	Mod. %	Lib. %
Should (NET)	67	79	54	70	81	75	67	52
Definitely should	45	62	31	44	63	60	39	33
Probably should	22	17	23	26	19	15	28	19
Should not (NET)	23	15	32	21	14	18	22	35
Probably should not	10	9	13	10	6	7	11	15
Definitely should not	13	7	19	11	8	11	11	20
Not at all sure	10	5	14	9	4	7	11	13

Primer: Starost matere in porodna teža otroka

Vprašanje:

Ali starost matere vpliva na porodno težo otroka?

Naloga:

Načrtuj raziskavo, ki bo odgovorila na zastavljeno vprašanje!

Možna rešitev:

Oglejmo si populacijo otrok, rojenih ob roku, katerih matere so bile ob porodu stare 35 let. Porodno težo teh otrok lahko primerjamo s porodno težo vseh slovenskih ob roku rojenih otrok.

Recimo, da je znano, da je povprečna porodna teža otrok v Sloveniji 3348 gramov.

Vprašanje sedaj postavimo takole: Ali je povprečna porodna teža otrok, ki so jih rodile 35-letne matere, različna od 3348 gramov?

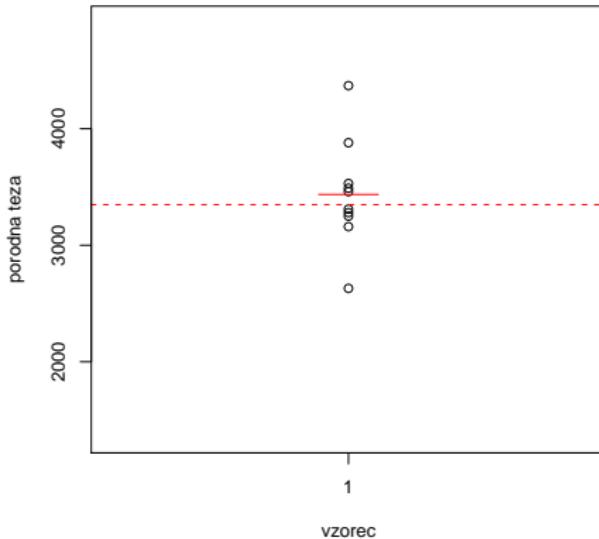
Opomba: Možne so boljše študije!

Populacija, ki nas zanima, so torej novorojenčki, ki so jih rodile 35-letne Slovenke. V kolikor nimamo registra vseh porodov (v Sloveniji ga imamo!), je seveda nemogoče izmeriti porodne teže otrok vseh 35-letnic v danem obdobju. Zato izberemo le vzorec, recimo 10 takšnih novorojenčkov. Dobimo naslednje vrednosti (v gramih):

3310, 3880, 3460, 3490, 3160, 3250, 2630, 4370, 3530, 3280.

Povprečje teh vrednosti je 3436, povprečje vseh slovenskih otrok (v resnici gre za povprečje nekaj čez 6000 porodov v ljubljanski regiji v določenem letu) pa je 3348 gramov.

Na naslednji sliki so posamezne vrednosti v vzorcu, njihovo povprečje (kratka neprekinjena črta) in povprečje populacije.



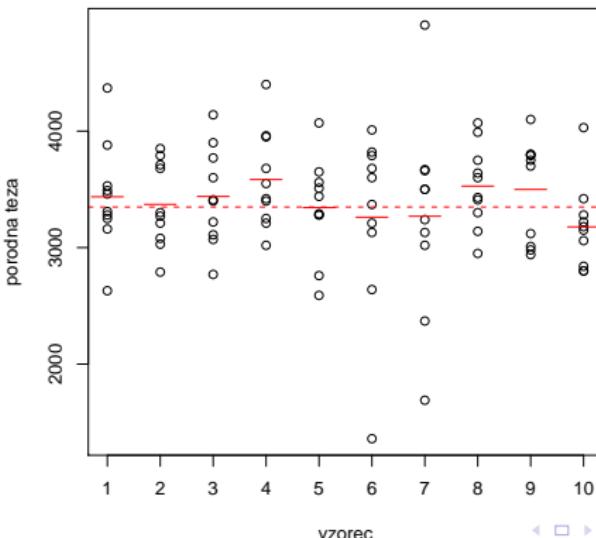
Kljub temu, da so vse matere enako stare, se porodne teže otrok razlikujejo, kar predvsem pripisujemo biološki variabilnosti. Za nas pomembno vprašanje pa je tote:

Ali je opažena razlika med povprečjema vzorca in populacije prava (sistematicna), ali pa gre le za naključno variabilnost?

Recimo, da je še 9 raziskovalcev naredilo isto, torej izmerilo porodno težo desetih otrok, ki so se rodili 35-letnim materam.

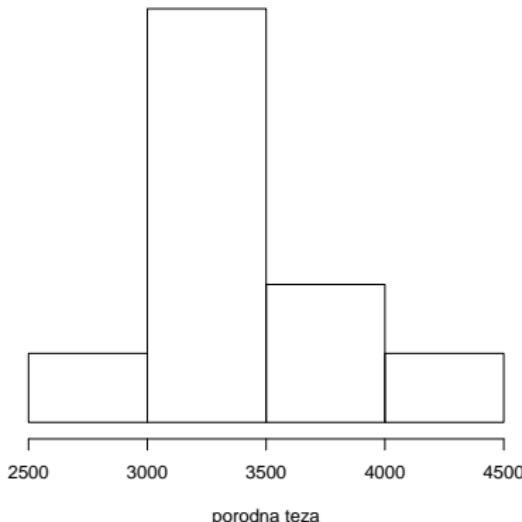
Na spodnji sliki je teh 9 vzorcev dodano prvemu vzorcu.

Vzorčna povprečja se očitno razlikujejo. Vendar, vsa se zdijo dovolj blizu populacijskemu povprečju. Smiselno je sklepati, da je povprečna porodna teža otrok 35-letnih mater enaka populacijski vrednosti.

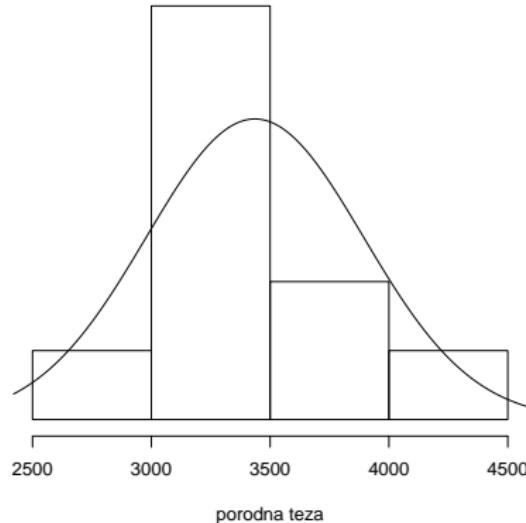


V praksi večinoma ne bomo mogli primerjati svojih rezultatov z rezultati drugih. Da bi odgovorili na vprašanje o naključni razliki, bomo potrebovali teoretični model porazdelitve porodnih tež.

Grafu spodaj rečemo **histogram**, prikazuje pa frekvenco (ali pa relativno frekvenco) vrednosti spremenljivke v določenih intervalih. Povedano drugače, histogram prikazuje porazdelitev vrednosti spremenljivke.



Pri samo desetih vrednostih je slika seveda precej groba. Če pa bi imeli mnogo več podatkov, bi bili intervali lahko zelo majhni in si lahko predstavljamo, da bi postali robovi histograma precej bolj gladki. Zaenkrat pač privzemimo nek model (o tem več pozneje) porazdelitve porodnih tež, ki ga ilustrira naslednja slika z vrisano teoretično porazdelitvijo.



Teoretično porazdelitev lahko uporabimo (kako? o tem spet kasneje) za to, da izračunamo verjetnost za vsaj tako veliko razliko, kot smo jo opazili. Ugotovljena razlika je bila $3436 - 3348 = 88$ gramov in teorija pravi, da bi vsaj tako veliko razliko dobili v 63% primerov, torej da je verjetnost vsaj tako velike razlike 0,63. Tej verjetnosti pravimo stopnja tveganja oz. kar ***p-vrednost***.

Sklep: Podatki ne nasprotujejo trditvi, da je povprečna porodna teža otrok 35-letnih mater enaka povprečju v populaciji. Naša raziskava torej govori o tem, da porodna teža otroka ni odvisna od starosti matere.

Uvodni primer ilustrira dejstvo, da so trditve v statistiki verjetnostne. Preden se torej lotimo statističnih metod, se nekoliko pomudimo pri osnovah verjetnosti.

OSNOVE VERJETNOSTI

Primer: V porodnišnici se je v določenem letu rodilo 1756 otrok, od tega 901 deček.

Frekvenca dečkov je 901, **relativna frekvenca** dečkov pa $901/1756 = 0,513$.

Opazimo:

- ▶ Relativna frekvenca leži med 0 in 1.
- ▶ Relativno frekvenco izračunamo iz podatkov, je torej ugotovljena (izmerjena, opažena) vrednost.

Osnovna pojma verjetnosti sta **poskus** in **dogodek**. O poskusu govorimo, kadar se neka množica dejstev vedno pojavi hkrati. Dogodek pa je pojav, ki se v poskusu lahko zgodi, a to ni nujno. V gornjem primeru je poskus porod, dogodek pa, da se rodi deček.

'Definicija' verjetnosti

Verjetnost je tista vrednost, pri kateri se stabilizira relativna frekvenca dogodka v velikem številu poskusov.

Gornja definicija ni matematično korektna, a bo za našo rabo zadoščala.

Spodnja tabela prikazuje število porodov v Sloveniji v letih 1991 do 1995. Vidimo, da se relativne frekvence dečkov gibljejo okrog 0,51, kar bi bilo nekako smiselno vzeti za verjetnost rojstva dečka.

Leto	1991	1992	1993	1994	1995
Porodov skupaj	21583	19982	19793	19463	18980
Dečki	11116	10333	10188	9899	9741
Deklice	10467	9649	9605	9564	9239
Relativna frekvenca	0,515	0,517	0,514	0,508	0,513

Notacija:

Dogodke bomo označevali z velikimi tiskanimi črkami, na primer

Pri metu kocke $A = \{\text{izid je sodo število}\}$

Ob rojstvu $B = \{\text{novorojenček je deček}\}$

Gestacijska starost $C = \{\text{tednov nosečnosti} \geq 37\}$

Verjetnost dogodka A bomo označevali s $P(A)$. Seveda za vsak dogodek A velja

$$0 \leq P(A) \leq 1.$$

Še nekaj definicij

AB ali $A \cap B$ (A krat B) je dogodek, ki se zgodi, kadar se zgodita A in B . Temu dogodku rečemo **produkt** dogodkov.

$A \cup B$ (A ali B) je dogodek, ki se zgodi, kadar se zgodi A ali B . Govorimo o **vsoti** dogodkov.

$A - B$ **razlika** se zgodi, če se zgodi A in ne zgodi B .

\bar{A} (**ne** A) je dogodku A nasproten dogodek.

Če se vedno, ko se zgodi A , zgodi tudi B , pravimo, da je A **način** dogodka B in pišemo $A \subset B$. Rečemo tudi, da je A **vsebovan v B** .

Dogodek, ki se vedno zgodi, imenujemo **gotov** dogodek in ga ponavadi zaznamujemo s črko G . Njemu nasproten je dogodek, ki se nikoli ne zgodi in mu rečemo **nemogoč** dogodek ter ga označimo z N .

Dogodka A in B sta **nezdružljiva**, če je njun produkt nemogoč dogodek. Torej kadar je $A \cap B = N$.

RAČUNANJE Z VERJETNOSTMI

- a. $P(A) + P(\bar{A}) = 1$
- b. $P(A \cap B) + P(\bar{A} \cap B) = P(B)$
- c. $C \subset B \Rightarrow P(C) \leq P(B)$
- d. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Primer:

$$A = \{\text{gestacijska starost} \geq 40\}$$

$$\bar{A} = \{\text{gestacijska starost} < 40\}$$

$$B = \{\text{novorojenček je deklica}\}$$

Formula b. potem pravi:

$$P(\text{novorojenček je deklica}) =$$

$$P(\text{novorojenček je deklica rojena v 40. tednu ali kasneje}) +$$

$$P(\text{novorojenček je deklica rojena pred 40. tednom})$$

in torej izraža preprosto dejstvo, da je

število novorojenih deklic =

število novorojenih deklic rojenih v ali po 40. tednu +

število novorojenih deklic rojenih pred 40. tednom.

Naredimo korak naprej. Naj bo:

$$A_1 = \{\text{gestacijska starost} \leq 35\}$$

$$A_2 = \{\text{gestacijska starost} = 36, 37, 38\}$$

$$A_3 = \{\text{gestacijska starost} = 39, 40, 41\}$$

$$A_4 = \{\text{gestacijska starost} \geq 42\}$$

$$B = \{\text{novorojenček je deklica}\}$$

Potem je

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + P(A_4 \cap B),$$

Zgornja formula velja zato, ker so dogodki A_1, A_2, A_3 in A_4 **paroma nezdružljivi** in **izčrpni** (pomeni, da skupaj predstavljajo vse mogoče dogodke v poskusu). Tako velja v splošnem:

Če so A_1, A_2, \dots, A_m paroma nezdružljivi in izčrpni dogodki, potem je

$$P(B) = \sum_{i=1}^m P(A_i \cap B)$$

Pogojna verjetnost

Primer: Za bolnika z rakom prostate bo verjetnost, da preživi 2 leti odvisna od tega, koliko je bila bolezen razširjena v času diagnoze. Ponavadi govorimo o štirih stadijih bolezni, označimo jih z I, II, III in IV.

Relativna frekvenca bolnikov, ki so živi po dveh letih in so bili v stadiju I potem **ni**

$$\frac{\text{število bolnikov, ki preživijo 2 leti}}{\text{število vseh bolnikov}}$$

ampak

$$\frac{\text{število bolnikov v \textbf{stadiju I}, ki preživijo 2 leti}}{\text{število vseh bolnikov v \textbf{stadiju I}}}$$

Delimo števec in imenovalec s skupnim številom bolnikov

$$\frac{\text{število bolnikov v \textbf{stadiju I}, ki preživijo 2 leti}}{\text{število vseh bolnikov v \textbf{stadiju I}}} / \frac{\text{število vseh bolnikov}}{\text{število vseh bolnikov v \textbf{stadiju I}}}$$

Z oznakami bomo zgornje izraze lažje prebrali. Naj bo:

A dogodek, da bolnik preživi 2 leti in

B dogodek, da je bolnik v stadiju I.

Berimo relativne frekvence kot verjetnosti in označimo s $P(A|B)$ **pogojno** verjetnost dogodka A glede na dogodek B , torej verjetnost, da se zgodi A , če se zgodi B . Potem zgornje relativne frekvence preberemo kot

$$P(A|B) = \frac{P(AB)}{P(B)}$$

kar vzemimo tudi za **definicijo pogojne verjetnosti**.

Primer: Delež dečkov med novorojenčki

V spodnji tabeli imamo relativne frekvence (v odstotkih) dečkov med novorojenčki. Vidimo, da je skupna relativna frekvenca dečkov 50,8% (verjetnost rojstva dečka torej 0,508), da pa se te spreminja glede na gestacijsko starost. Pogojna verjetnost rojstva dečka pri gestacijski starosti 35 tednov ali manj je na primer 0,532!

teden	št. dečkov	št. deklic	% dečkov
≤ 35	148	130	53,2
36	64	70	47,8
37	170	173	49,6
38	398	372	51,7
39	838	791	51,4
40	1163	1175	49,7
41	431	393	52,3
42	20	20	50,0
skupaj	3232	3124	50,8

Pogojna verjetnost je običajna verjetnost, definirana na novi, manjši množici dogodkov. Torej zanjo veljajo ista pravila za računanje, na primer

$$P(A|B) + P(\bar{A}|B) = 1.$$

Bayesova formula

Iz definicije pogojne verjetnosti preberemo, da je

$$P(AB) = P(B|A)P(A) = P(A|B)P(B)$$

in odtod dobimo **Bayesovo formulo**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

ter še

$$P(B) = P(AB) + P(\bar{A}B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}).$$

Neodvisnost

Dogodek A je neodvisen od dogodka B , če verjetnost dogodka A ni odvisna od verjetnosti dogodka B . Bolj formalno: A je neodvisen od B , kadar je

$$P(A) = P(A|B).$$

Neodvisnost pravzaprav pomeni, da B ne vsebuje nobene informacije o A .

Nekaj lastnosti

- Če je A neodvisen od B , je A neodvisen tudi od \bar{B} .
- Neodvisnost je simetrična relacija, zato lahko rečemo tudi "A in B sta neodvisna".
- Če sta A in B neodvisna, velja

$$P(A|B) = P(A|\bar{B}) = P(A) \quad \text{in}$$

$$P(B|A) = P(B|\bar{A}) = P(B).$$

- Če sta A in B neodvisna, velja

$$P(AB) = P(A)P(B).$$

Primer: Met kocke

Definirajmo dogodke

$$A = \{2,4,6\}$$

$$B = \{1,2,3,4\}$$

$$C = \{1,2,3\}$$

Seveda je

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

in tudi

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{2/6}{4/6} = \frac{1}{2}$$

in torej $P(A|B) = P(A)$. Se pravi, da sta A in B neodvisna.

Podobno bi ugotovili, da $P(A|C) \neq P(A)$ in s tem, da A in C nista neodvisna.

Primer: Rojstni dnevi

Vprašanje: Kolikšna je verjetnost, da imata med n ljudmi (med katerimi ni dvojčkov) vsaj dva na isti dan rojstni dan?

Primer: Rojstni dnevi

Vprašanje: Kolikšna je verjetnost, da imata med n ljudmi (med katerimi ni dvojčkov) vsaj dva na isti dan rojstni dan?

n	Verjetnost
20	0,41
23	0,51
30	0,71
40	0,89
50	0,97
60	0,99

SLUČAJNE SPREMENLJIVKE

Slučajna spremenljivka je merjena količina, katere vrednosti naključno variirajo. Največkrat rečemo kar spremenljivka. Označujemo jih ponavadi z velikimi tiskanimi črkami, njihove vrednosti pa z malimi tiskanimi črkami.

Nekaj primerov:

Poskus	Slučajna spremenljivka
Met kocke	Število pik
Met dveh kock	Vsota pik
Rojstvo	Teža novorojenčka

Neodvisne slučajne spremenljivke

Intuitivno: Dve slučajni spremenljivki sta neodvisni, če poznavanje vrednosti ene od njiju ne pove ničesar o vrednostih druge.

Nekaj primerov:

- Izida dveh zaporednih metov kocke sta neodvisna.
- Vrednosti krvnih pritiskov dveh različnih oseb so (ponavadi) neodvisne.
- Porazdelitvi višine in teže med ljudmi nista neodvisni.

Formalno: Slučajni spremenljivki X in Y sta neodvisni, če sta dogodka $\{X \leq x\}$ in $\{Y \leq y\}$ neodvisna za vsak x in y .

Neodvisnost oz. **pogojna neodvisnost** (spremenljivki sta neodvisni pri dani vrednosti tretje spremenljivke) sta v statistiki izjemno pomembni. Gre za idealizacijo ali poenostavitev procesov v naravi, ki jo s pridom izkoriščamo pri statističnem modeliranju.

Primera:

- Če je prognoza za bolnika z rakom **neodvisna** od histološke klasifikacije tumorja, lahko napovedujemo, ne da bi se ozirali na takšno klasifikacijo.
- Če je prognoza za bolnika z rakom **pogojno neodvisna** od histološke klasifikacije tumorja pri dani starosti bolnika, potem nam ni potrebno poznati histološke klasifikacije, če poznamo starost bolnika.

Kot bomo videli, bosta neodvisnost oz. pogojna neodvisnost pogosto predpostavki (hipotezi), ki ju bomo preverjali s statističnimi testi. Kadar bomo predpostavko o neodvisnosti zavrnili, bo pomembno opisati naravo odvisnosti.

VRSTE SPREMENLJIVK

1. **Opisne** (atributivne, kategorialne) - vrednosti spremenljivke le opišemo. Opisne spremenljivke ponavadi delimo na
 - ▶ **imenske** (nominalne) so tiste, katerih vrednosti ne moremo urediti
 - ▶ **vrstilne** (ordinalne) so tiste, katerih vrednosti lahko uredimo po velikosti (v tako imenovano ranžirno vrsto)
2. **Numerične** - vrednosti spremenljivke so števila, s katerimi lahko računamo. Imamo spet dve podskupini
 - ▶ **razmične** (intervalne) so tiste, ki jih lahko odštevamo, njihov kvocient pa nima pravega smisla, ker ne obstaja absolutna ničla (čeprav obstaja neka ničla). Na primer temperatura.
 - ▶ **razmernostne** (racionalne) imajo še absolutno ničlo in tako npr. kvocient 2 pomeni, da je ena vrednost dvakrat večja od druge.

Stevens's Classification

Nominal: hot or cold faucet



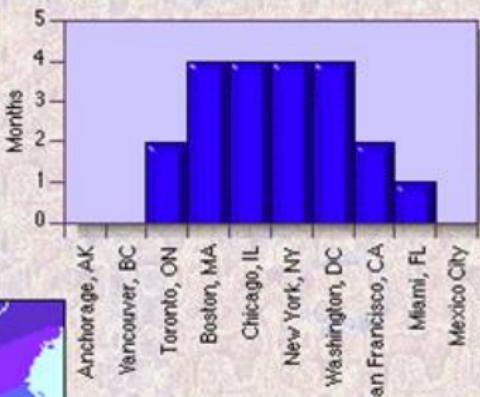
Interval: body temperature



Ordinal: isotherms



Ratio: typical number of months with one or more days with maximum temperatures of 100°F or more for ten North American cities



According to the psychometrician S. S. Stevens, a variable can also be categorized as:

Nominal: the same as qualitative.

Ordinal: categories with a natural quantitative sequence (order).

Interval: successive numerical values that are equally spaced; zero does not mean the absence of the characteristic being measured and hence can be chosen arbitrarily.

Ratio: the same as interval, but with a natural zero.

Gender Is a Qualitative Variable

Examples of **qualitative** data include:

sex (male, female);

5-year survival (survived five years, did not survive five years); and

presence of a particular symptom (present, absent).

In each example, the information of interest would be the **count** for each category.



▶ Play/Stop

■ Examples of Qualitative Data

Data Basics

Qualitative Data



Topics

Testing

Options

Blood Type Is a Qualitative Variable

Qualitative variables are those that cannot be expressed quantitatively, but rather can only be categorized.

For most qualitative variables, there are a small number of possible categories.



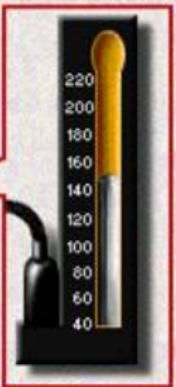
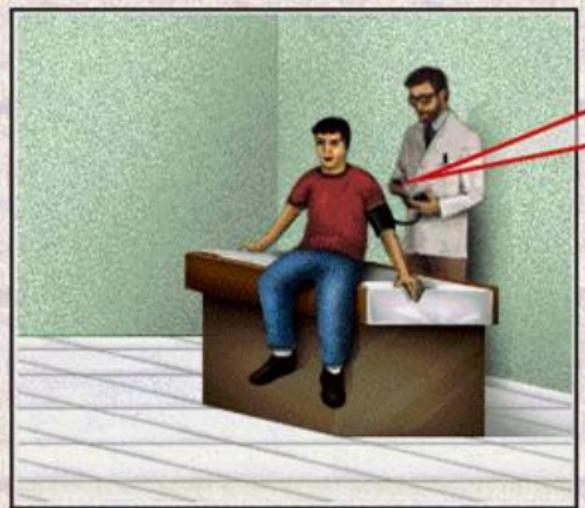
The ABO blood types (categories) are not quantitative.



Examples of Qualitative Data

[Data Basics](#)[Qualitative Data](#)[Topics](#)[Testing](#)[Options](#)

Blood Pressure Is a Quantitative Variable



Quantitative variables are those to which we can put a "measuring stick." These can be further classified as **continuous** (that is, the "stick" has all possible values in a given range) or **discrete** (in which only certain values are possible).

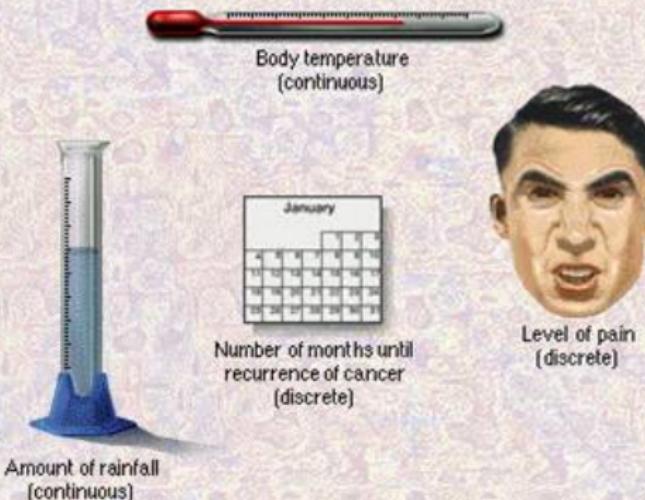


Examples of Quantitative Data

[Data Basics](#)[Quantitative Data](#)[Topics](#)[Testing](#)[Options](#)

Continuous Variables

Some examples of quantitative variables are:



- body temperature (**continuous**),
- amount of rainfall (**continuous**),
- number of months until recurrence of cancer (**discrete**), and
- level of pain, e.g., mild, moderate, or severe (**discrete**).

Examples of Quantitative Data

Country	Inflation	GDP per capita	People per MD	People per TV	Life expectancy
Kenya	0.1	1,377	9,851	78.4	59
Japan	0.2	21,328	566	1.6	80
Singapore	1.3	21,493	711	2.6	76
Fiji	1.5	5,220	2,080	30.8	72
Germany	1.8	20,165	333	1.8	76
France	2.0	19,774	333	2.5	77
Switzerland	2.0	24,483	585	2.5	78
Taiwan	2.0	13,235	928	3.1	75
Brunei	2.5	15,580	1,323	3.1	74
Canada	2.5	21,268	446	1.6	78
U.S.	2.5	25,900	419	1.2	77
Maldives	3.1	1,373	5,330	48.0	64
Malaysia	3.4	8,763	2,063	4.7	72
New Zealand	3.5	17,045	332	2.3	76
Bangladesh	3.7	1,290	12,500	170.5	56

Here, the observations and variables come from a larger data set published in *Asia Week*.

Observations: Each country is an observation. The table is part of a data set that includes 46 countries, mostly from Asia, Europe, and North America.

Variables: All the variables listed, except for country name, are quantitative. But we might decide to define a qualitative (or categorical) variable that isn't mentioned here—say, Region, with such categories as Asia, Europe, North America, Africa, and so on.

Primer avtoritativnosti: pokojninska reforma in neka statistika

Dobil sem tole elektronsko pošto:

To je treba razposlat, **da ne bo kak nepoučeni na referendumu slučajno volil ZA** pokojninsko reformo. Saj veste, da leta niso edina svinjarija, s katero nas v reformi čakajo, kajne.

Pokojninska reforma in neka statistika

Pokojninska reforma - kraja stoletja v višini 100.000 EUR po posamezniku! - Finance.si - Windows Internet Explorer



<http://www.finance.si/blog?ids=10&post=673>



Igor Juršič, Ruše



File Edit View Favorites Tools Help

Favorites

Najdi.si

Suggested Sites

Free Hotmail

Google

Web Slice Gallery



Pokojninska reforma - ...

Google Calendar



Page Safety Tools



Zdaj odprite trgovalni račun brokerjet

brokerjet Online Broker Banke Sparkasse

Prijava Registracija

PRVA STRAN

DANES

ČASOPIS

NAJ BRANO

BORZA

SKLADI

FORUM

TRGOVINA

BLOG

miniFIPO

RUBRIKE

NAROČAH

FINANCE

Blogi [Tekanje...]

Gostujeći blogi

PRVA STRAN · Vizitka



Blogi po izbiri uredništva
Finance.si objavljeno z dovoljenjem avtorjev

Arhiv

apr 2011 (2) arhiv

Rubrike

4 Kamenko Kesar

0 Krošek

1 Ljubljanska borza

0 Razvili trgi

0 Trgi v razvoju

Iskalnik

poisci

Pokojninska reforma - kraja stoletja v višini 100.000 EUR po posamezniku!

, 20.08.10 10:33:14

Ali - za kako butaste nas imajo?

Minister Svetlik nas poskuša prepričati, da je delo do 65 leta starosti nuja, saj drugače pokojninski sistem ne bo vzdržal. S svojim predlogom pa bo vsakemu, ki vplačuje v pokojninsko blagajno odtujil v povprečju skoraj 100.000 EUR.

Že pred časom sem dokazoval in dokazal, da tisti, ki prejema plačo v višini 1.000 EUR neto, odvede mesečno skupaj s podjetjem 370 EUR za pokojnino blagajno in bi samo ob polaganju na banko po 35 letih delovne dobe prejel 80% plače v obliki obresti, glavnica pa bi ostala nedotaknjena.

Ob tako glasni napovedi pokojninski reforme in negodovanju ministra Svetlikja, da je trenutno stanje nezvržljivo, sem se pojgral s statistiko in napovedimi ministra svetinja:

1. Povprečna starost umrilih v Sloveniji je 74,1 let (po podatkih statističnega urada iz leta 2008)
2. Upokojevali bi se pri 65 letih po približno 40 letih delovne dobe
3. Povprečna pokojnina (po podatkih SPIZ - dec. 2008) je znašala 602,90 €
4. Povprečna neto plača v RS (po podatkih SURS - dec. 2008) je znašala 938,66 €
5. Seževek prispevkov zaposlenega in delodajalca (na 938,66€) za SPIZ znaša 342,50 €

Sedaj pa nam ostane le izračun po zdravi pameti, pri čemer se zanašam na statistične podatke in sledi:

V 40 letih (torej do 65 leta starosti) vplačamo $342,5 \times 12 \times 40 = 164.160$ € za pokojnino. Pokojnini potem v povprečju prejemamo 9,1 leta torej $602,90 \times 12 \times 9,1 = 65.836,68$ €. Razlika med vplačanjem in dobljajem ob upoštevanju statističnih podatkov pa je $98.233,32$ €.

Vprašanje Svetliku in vsem, ki kimajo k pokojninski reformi torej glasi:

Kaj nameravate storiti z 98.233,32 €, ki jih vsak zaposlen v Sloveniji preplača v pokojninsko blagajno.

Dovolj je, da nismo imate za norca z drugimi »presvetlimi« prebiskali, zato Vas prosim, da nam ne odtujite še skoraj 100.000 EUR vsakemu ki (obvezno) vplačuje v pokojninski sistem!!!

Upravo, da sem napovedal 98.233,32 zplačevalcu, da je pokojninska reforma le dodatno začinj

Postanite (močan) MOŠKI, kakršnega si želi vsaka ŽENSKA



Facebook | Twitter



Zaslužimo si boljši vlado! [127]



Kje so danes graditelji slovenske države [87]



Baron Ivan Zidar, upnikov grobar [43]



Zares grozi, da lahko tudi zaradi AUKN Pahorja zapusti [42]



Velika noč in prvi maj bosta velik posel [4]

OGLAS

ELEMENTUM
INTERNSHIPS

**Nakup velikih palic
pomeni za
vas do 10 %
boljšo
nabavno
ceno**

Članki iz tiska
sreda, 20. apr / št. 077

Pokojninska reforma in neka statistika

Pokojninska reforma - kraja stoletja v višini 100.000 EUR po posamezniku!

, 20.08.10 10:33:14

Ali - za kako butaste nas imajo?

Minister Svetlik nas poskuša prepričati, da je delo do 65 leta starosti nuja, saj drugače pokojninski sistem ne bo vzdržal. S svojim predlogom pa bo vsakemu, ki vplačuje v pokojninsko blagajno odtujil v povprečju skoraj 100.000 EUR.

Že pred časom sem dokazoval in dokazal, da tisti, ki prejema plačo v višini 1.000 EUR neto, odvede mesečno skupaj s podjetjem 370 EUR za pokojninsko blagajno in bi samo ob polaganju na banko po 33 letih delovne dobe prejemal 80% plače v obliki obresti, glavnica pa bi ostala nedotaknjena.

Pokojninska reforma in neka statistika

Ob tako glasni napovedi pokojninske reforme in negodovanju ministra Svetlika, da je trenutno stanje nevzdržno, sem se poigral s statistiko in napovedmi ministra svetlika:

1. Povprečna starost umrlih v Sloveniji je 74,1 leto (po podatkih statističnega urada iz leta 2008)
2. Upokojevali bi se pri 65 letih po približno 40 letih delovne dobe
3. Povprečna pokojnina (po podatkih SPIZ - dec. 2008) je znašala 602,90 €
4. Povprečna neto plača v RS (po podatkih SURS - dec. 2008) je znašala 938,66 €
5. Seštevek prispevkov zaposlenega in delodajalca (na 938,66€) za SPIZ znaša 342,50 €

Pokojninska reforma in neka statistika

Sedaj pa nam ostane le izračun po zdravi pameti, pri čemer se zanašam na statistične podatke in sled

V 40 letih (torej do 65 leta starosti) vplačamo $342 \times 12 \times 40 = 164.160$ € za pokojnino.

Pokojnino potem v povprečju prejemamo 9,1 leto torej $602,90 \times 12 \times 9,1 = 65.836,68$ €

Razlika med vplačanim in dobljenim ob upoštevanju statističnih podatkov pa je $98.323,32$ €.

Vprašanje Svetliku in vsem, ki kimajo k pokojninski reformi torej glasi:

Kaj nameravate storiti z $98.323,32$ €, ki jih vsak zaposlen v Sloveniji preplača v pokojninsko blagajno.

Pokojninska reforma in neka statistika

Dovolj je, da nas imate za norca z drugimi »presvetlimi« prebliski, zato Vas prosim, da nam ne odtujite še skoraj 100.000 EUR vsakemu ki (obvezno) vplačuje v pokojninski sistem!!!

Upam, da sem navedel 98.323 razlogov, zakaj smatram, da je pokojninska reforma le dodaten način polnjenja državne blagajne iz katere se potem plačujejo neracionalni projekti gospodarskih botrov političnih veljakov in da ni prav nobenega opravičenega razloga zanjo, saj že sedaj vplačujemo preveč kar vidno bremeni tako podjetja, kot posameznike.

avtor: Igor Juršič, Ruše

In kaj je narobe s to statistiko?

In kaj je narobe s to statistiko?

Pričakovana vrednost preživetja je seveda odvisna od dosežene starosti!! (pogojna porazdelitev)

In kaj je narobe s to statistiko?

Pričakovana vrednost preživetja je seveda odvisna od dosežene starosti!! (pogojna porazdelitev)

Kdor pride do 65 let, lahko pričakuje precej več kot le še 9 let.

In kaj je narobe s to statistiko?

Pričakovana vrednost preživetja je seveda odvisna od dosežene starosti!! (pogojna porazdelitev)

Kdor pride do 65 let, lahko pričakuje precej več kot le še 9 let.

Leta 2008 je bilo to (za skupino od 65 do 69 let)

pričakovano trajanje življenja	
moški	16,28
ženske	20,16

Podobno napako delajo tisti, ki ugotavljajo, da

Podobno napako delajo tisti, ki ugotavljajo, da

- ▶ živijo Nobelovci (Oskarjevci, vrhunski dirigenti ...) dlje od povprečja.

Podobno napako delajo tisti, ki ugotavljajo, da

- ▶ živijo Nobelovci (Oskarjevci, vrhunski dirigenti ...) dlje od povprečja.
- ▶ da bolniki, ki so reagirali na zdravljenje živijo dlje od tistih, ki niso.

Tudi tole je seveda res ...

Si vedel, da
ženske živijo
dlje?

zlasti vdove



Ko smo že pri demografiji ... :)

Ko smo že pri demografiji ... :)

Mene podatek, da se populacija stara, ne zadane prav močno.

Ko smo že pri demografiji ... :)

Mene podatek, da se populacija stara, ne zadane prav močno.

Vedno je treba povedati še kako hitro (v V. Britaniji na primer za 2,5 let na desetletje).

Ko smo že pri demografiji ... :)

Mene podatek, da se populacija stara, ne zadane prav močno.

Vedno je treba povedati še kako hitro (v V. Britaniji na primer za 2,5 let na desetletje).

Posebej zadane pa me tole:

Ko smo že pri demografiji ... :)

Mene podatek, da se populacija stara, ne zadane prav močno.

Vedno je treba povedati še kako hitro (v V. Britaniji na primer za 2,5 let na desetletje).

Posebej zadane pa me tole:

Z delovno dobo 40 let naj bi zaslužil dovolj za **80 let življenja** (približno trenutno povprečje v razvitem svetu).

Podatki o samomorih na teletekstu RTV Slovenija

437 samomorov v 2011

Podatki o samomorih na teletekstu RTV Slovenija

437 samomorov v 2011

Od tega

poročeni	42,4%
samski	31,7%
ločeni	13,3%
ovdoveli	12,6%

Podatki o samomorih na teletekstu RTV Slovenija

437 samomorov v 2011

Od tega

poročeni	42,4%
samski	31,7%
ločeni	13,3%
ovdoveli	12,6%

Lahko bi dodali:

Podatki o samomorih na teletekstu RTV Slovenija

437 samomorov v 2011

Od tega

poročeni	42,4%
samski	31,7%
ločeni	13,3%
ovdoveli	12,6%

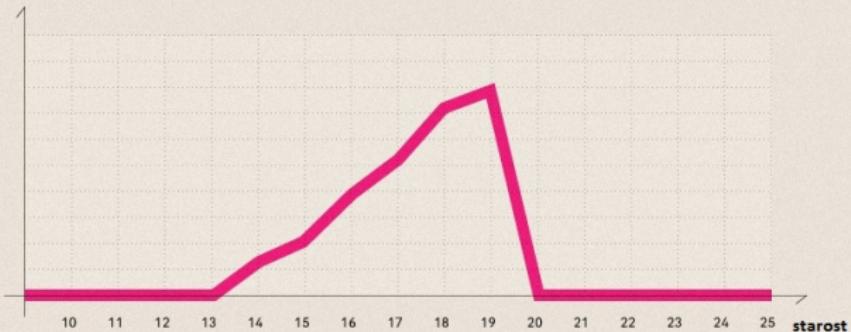
Lahko bi dodali:

In pred samomorom je bilo 100% še živih!

SURS je v resnici objavil ustrezeno tabelo

Zakonski stan	Umrli zaradi samomora (%)	Prebivalci Slovenije (%)
Skupaj	100,0	100,0
Samski	31,7	38,7
Poročeni	42,4	46,6
Razvezani	13,3	6,4
Ovdoveli	12,6	8,3

Statistika kaže, da število najstniških nosečnosti drastično pada po 20. letu



OPISOVANJE VARIABILNOSTI V POPULACIJI

Če možnih vrednosti ni veliko, preprosto navedemo verjetnost vsake vrednosti.

Primer: *Met kocke*

Možni izidi so 1, 2, 3, 4, 5, in 6, njihove verjetnosti pa

$$P(1) = P(2) = \dots = P(6) = 1/6.$$

Primer: *Krvna skupina*

Če imata starša oba krvno skupino AB , so verjetnosti krvne skupine pri otroku naslednje

$$P(A) = 1/4, \quad P(AB) = 1/2, \quad P(B) = 1/4.$$

V praksi seveda ni vedno tako enostavno.

Primer: Met dveh kock

Možnih je 36 izidov, vsak ima verjetnost $1/36$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Slučajna spremenljivka $Y = \text{"vsota pik"}$ lahko zavzame vrednosti $2, 3, 4, \dots, 12$.

Verjetnostno porazdelitev spremenljivke Y lahko izpeljemo iz verjetnostne porazdelitve parov pik. Dobimo

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Porazdelitvene funkcije

Za zvezne spremenljivke, kot so teža, krvni pritisk in podobno, podajanje porazdelitve verjetnosti v tabelah ni praktično (teoretično niti ni možno, ampak o tem tu ne bomo).

Variabilnost v teh primerih podajamo s **(kumulativno) porazdelitveno funkcijo**. Za slučajno spremenljivko X je njena porazdelitvena funkcija $F(x)$ definirana takole

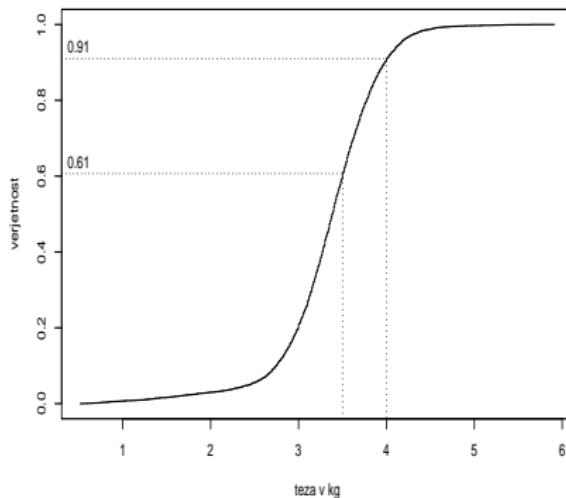
$$F(x) = P(X \leq x),$$

Pri dani vrednosti x je $F(x)$ torej verjetnost, da so vrednosti slučajne spremenljivke X manjše od x . Z drugimi besedami, $F(x)$ je delež vrednosti X , ki so manjše od x .

Vrnimo se k primeru s porodno težo. Spodnja slika kaže (empirično) porazdelitveno funkcijo porodne teže. Iz nje razberemo, da je 61% novorojenčkov lažjih od 3,5 kg in kar 91% lažjih od štirih kilogramov. Med 3,5 in 4 kg pa je

$$\begin{aligned} P(\text{teža} \leq 4\text{kg}) - P(\text{teža} \leq 3,5\text{kg}) &= F(4) - F(3,5) \\ &= 0,91 - 0,61 = 0,3, \end{aligned}$$

torej 30% novorojenčkov.



Verjetnostna porazdelitev in gostota verjetnosti

Diskretne porazdelitve

Kot smo videli na primerih, za diskrette spremenljivke ponavadi lahko navedemo verjetnosti pojava posameznih vrednosti, torej

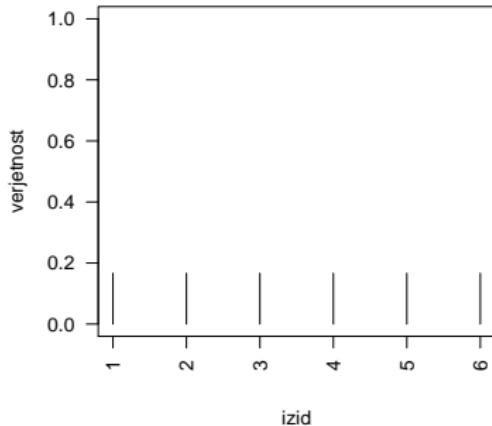
$$p(x) = P(X = x),$$

čemur rečemo **verjetnostna porazdelitev**. Porazdelitvena funkcija pri x je potem definirana kot vsota verjetnosti vseh izidov, manjših ali enakih x

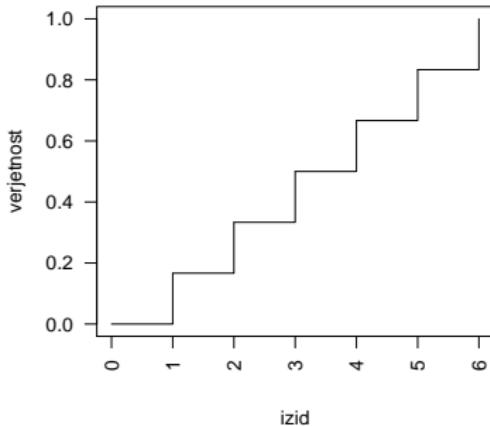
$$F(x) = \sum_{y \leq x} p(y).$$

Primer: Met kocke

porazdelitev



porazdelitvena funkcija



Za verjetnostno porazdelitev velja:

1. $\sum_{\text{vsi } x} p(x) = 1$.
2. $p(x) \geq 0$.
3. $P(a < X \leq b) = \sum_{a < x \leq b} p(x)$.

Zvezne porazdelitve

Pri zveznih porazdelitvah ne moremo govoriti o verjetnostih posameznih vrednostih, pač pa lahko govorimo o **gostoti verjetnosti**. To je funkcija, ponavadi jo označujemo z $f(x)$, ki pove, kako goste so vrednosti okrog danega x . Verjetnost, da je vrednost spremenljivke v intervalu $(x, x + \Delta x)$ je približno $f(x)\Delta x$, natančna definicija pa pravi

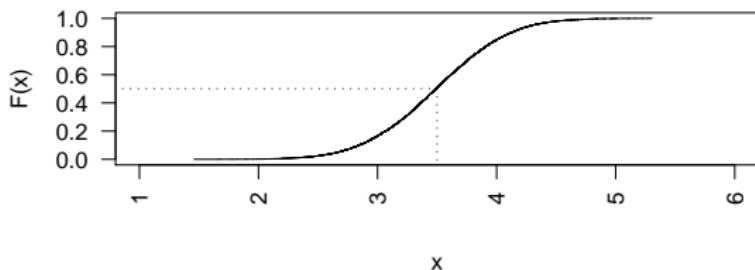
$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Za gostoto velja:

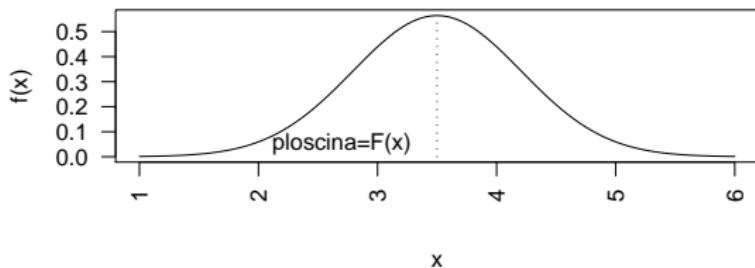
1. $\int_{-\infty}^{\infty} f(x)dx = 1.$
2. $f(x) \geq 0.$
3. $P(a < X \leq b) = \int_a^b f(x)dx.$

Povezava med porazdelitveno funkcijo in gostoto je ilustrirana na spodnji sliki.

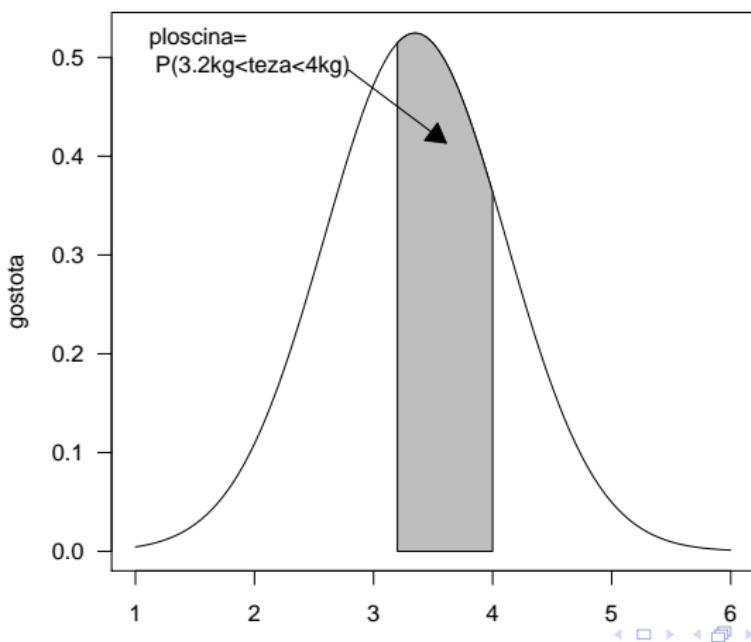
porazdelitvena funkcija



gostota porazdelitve



Če je porazdelitev zvezna, je verjetnost, da je vrednost spremenljivke v nekem intervalu (a,b) enaka ustreznri ploščini pod krivuljo, ki predstavlja gostoto porazdelitve. Na spodnji sliki je ilustrirana verjetnost, da je porodna teža otroka med 3,2 in 4 kilogrami.



OPISOVANJE VARIABILNOSTI NA VZORCU

Porazdelitvena funkcija, verjetnostna funkcija in gostota so **teoretične funkcije**, ki opisujejo variabilnost v populaciji.

Za opisovanje variabilnosti na vzorcih uporabljamo analogno definirane **empirične funkcije**.

Diskretne porazdelitve

Histogram: Graf frekvenc, ali relativnih frekvenc za vsako vrednost spremenljivke.

Empirična porazdelitvena funkcija: Graf kumulativne relativne frekvence.

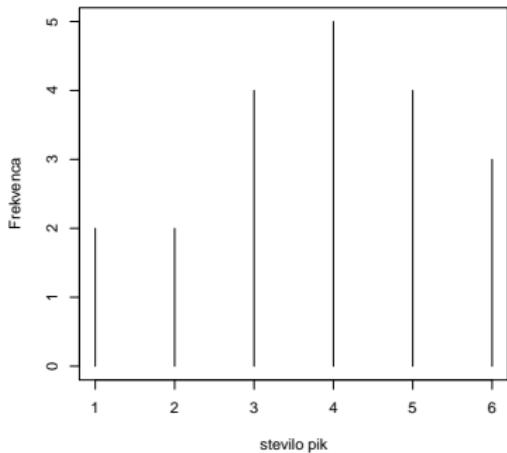
Primer: Metanje kocke. Kocko vržemo 20-krat.

Met	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Izid	1	3	3	6	5	5	4	1	2	3	4	5	4	3	4	6	6	2	4	5

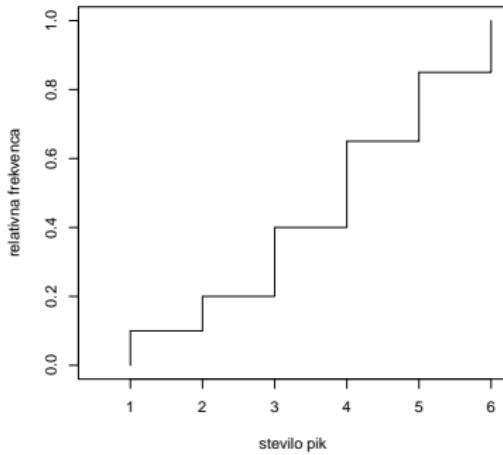
Izide lahko povzamemo v tabeli

Izid	1	2	3	4	5	6
Relativna frekvenca	10%	10%	20%	25%	20%	15%

ali narišemo histogram



Empirična porazdelitvena funkcija pa je videti takole

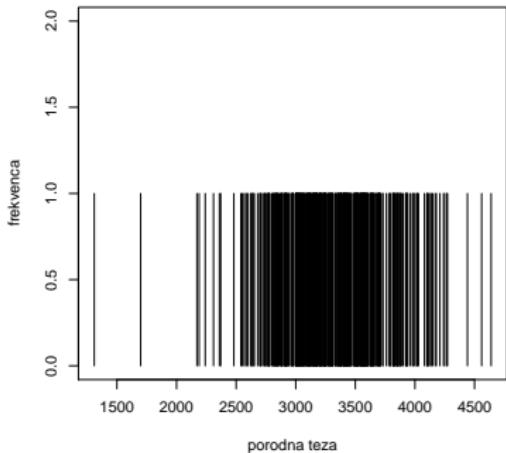


Opomba: Pri majhnem številu vrednosti (kot v našem primeru) je tabela ponavadi najprimernejša.

Zvezne porazdelitve

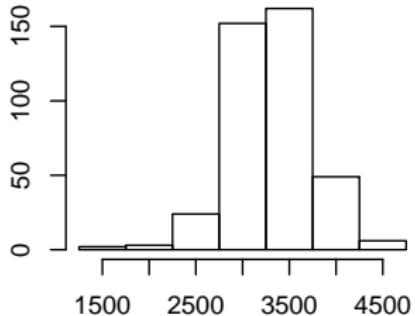
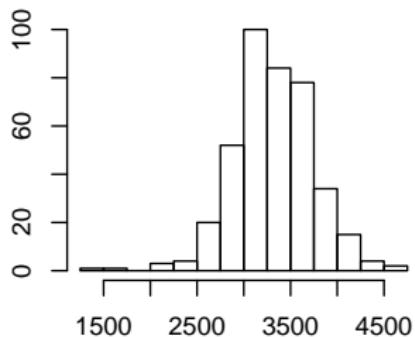
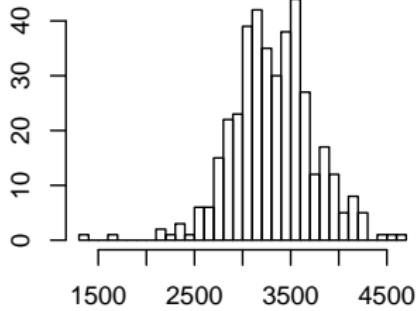
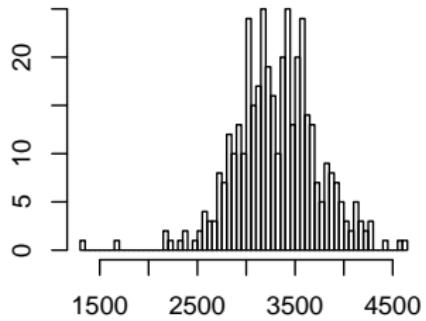
Primer: *Porodna teža dečkov, rojenih v 38. tednu nosečnosti*

Spodnji graf (histogram) prikazuje frekvence porodne teže 398 dečkov. Teža je merjena na gram natančno.



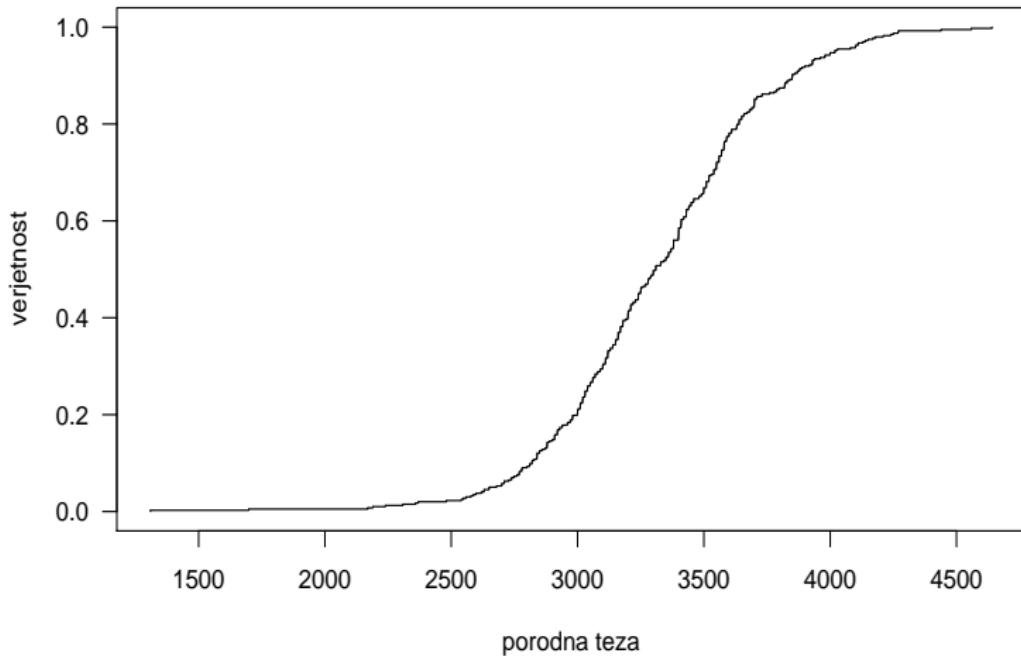
Ker so vse teže različne (frekvence so 1!), graf ne pove dosti. Ločimo le bolj goste predele od manj gostih.

Smiselno je torej grupiranje podatkov. Vrednosti spremenljivke zato razdelimo v intervale in nad vsakim intervalom narišemo pravokotnik, čigar ploščina je sorazmerna (relativni) frekvenci vrednosti v tistem intervalu. Tudi takšnemu grafu rečemo **histogram**. Oblika histograma je odvisna od tega, koliko intervalov (razredov) smo izbrali. Na sliki so histogrami za 4 različna grupiranja (50g, 100g, 250g in 500g).



Na spodnji sliki pa je **empirična porazdelitvena funkcija**.

Čeprav je v smislu vsebine, ki jo prikazuje, enakovredna histogramu, pa informacijo s takšnega grafa težje razberemo.



Seveda histogram in empirična porazdelitvena funkcija nista edina možna načina za opisovanje variabilnosti na vzorcu. Vsaj še en graf je treba omeniti, a preden to storimo, moramo spoznati kvantile.

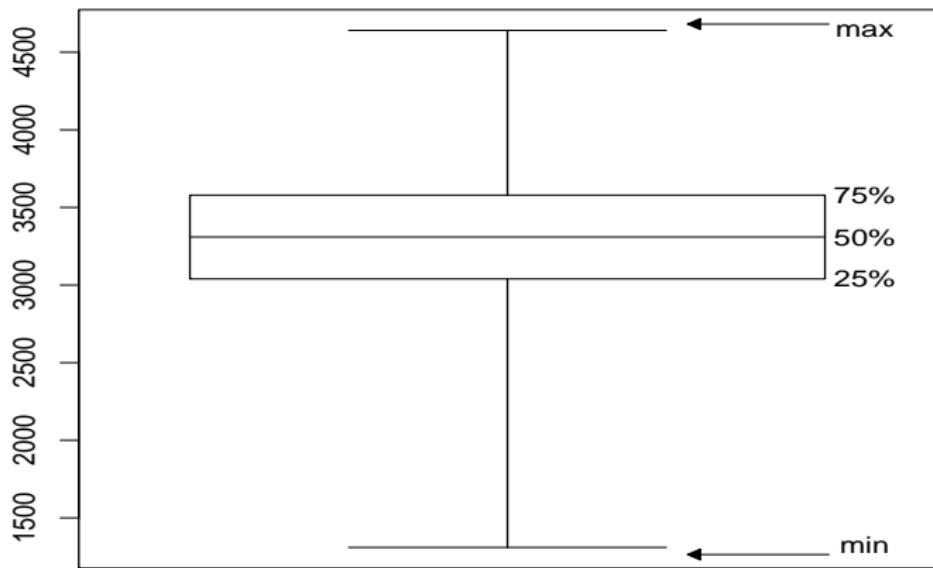
Kvantili

Vrednost, pod katero leži določen delež podatkov se imenuje kvantil. Ponavadi govorimo kar o **percentilih** (ali centilih). Teh je 99 in razdelijo vse po velikosti urejene podatke na 100 delov, v vsakem delu torej leži 1 odstotek vseh vrednosti. Na primer, pod 10. percentilom leži 10% vrednosti, nad njim pa 90% vrednosti. Navajanje percentilov pogosto ni praktično, največkrat nam za opis variabilnosti zadoščajo **kvartili**. To so vrednosti, ki po velikosti urejene podatke razdelijo na 4 kose. Kvartili so torej trije, prvi, drugi in tretji, drugi ima še posebno ime - **mediana**. Pod mediano torej leži polovica vseh vrednosti, nad njo pa tudi polovica.

Graf kvantilov

Graf kvantilov v (angleščini *box-and-whiskers plot*) v svoji osnovni obliki prikazuje 5 števil: minimum, tri kvartile in maksimum. Razpon od prvega do tretjega kvartila je prikazan s pravokotnikom (box), do minimuma in maksimuma pa od pravokotnika segata daljici (whiskers). Znotraj pravokotnika je s črto označena mediana. Nekatere verzije grafa drugače definirajo pomen daljic, zato je pametno vedno preveriti, kaj nam program nariše. Razdaljo od prvega do tretjega kvartila imenujemo **interkvartilni razmik**, ki tako po definiciji vedno zajema srednjih 50% podatkov.

Spodnja slika prikazuje enake podatke kot gornji histogrami, torej teže dečkov, rojenih v 38. tednu nosečnosti.



POVZEMANJE GLAVNIH ZNAČILNOSTI PORAZDELITVE

Porazdelitvena funkcija in gostota dajeta sicer popolno informacijo o variabilnosti v podatkih, a pogosto želimo informacijo strniti z nekaj značilnostmi. Za ta namen uporabljamo predvsem **mere središčnosti** in **mere razpršenosti**.

Od mer središčnosti si bomo ogledali samo dve, ki ju največkrat uporabljamo. Kar tu pa omenimo še tretjo, imenovano **modus**, ki je na vzorcu najpogostejša vrednost, v populaciji pa najverjetnejša vrednost.

Mere središčnosti

Na vzorcu velikosti n (empirična vrednost)	V populaciji (teoretična vrednost)
<p>Aritmetična sredina (ali vzorčno povprečje)</p> $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n [x_i \cdot \frac{1}{n}]$ <p>V kakšnem smislu je \bar{x} sredina?</p> <p>Ker je $\sum_{i=1}^n (x_i - \bar{x}) = 0$, sta vsoti levih in desnih odmikov enaki!</p> <p>Mediana = srednja vrednost glede na range (pri sodem n aritm. sredina srednjega para).</p>	<p>Pričakovana vrednost (populacijsko povprečje)</p> <p>Diskretna porazdelitev</p> $E(X) = \mu = \sum_x [x \cdot p(x)]$ <p>Zvezna porazdelitev</p> $E(X) = \mu = \int xf(x)dx$ <p>Mediana = tista vrednost, za katero velja $F(x) = 0,5$. Podobno definiramo druge kvantile. Za diskretne porazdelitve je stvar nekoliko nerodna.</p>

Mere razpršenosti

Na vzorcu velikosti n (empirična vrednost)	V populaciji (teoretična vrednost)
Vzorčna varianca = povprečen kvadrirani odmik od aritmetične sredine $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ (Zakaj ne delimo z n bo jasno kasneje!)	Varianca = pričakovana vrednost kvadririranega odmika od aritmetične sredine $Var(X) = \sigma^2 = \sum_x [x - E(X)]^2 \cdot p(x)$ oziora $Var(X) = \sigma^2 = \int [x - E(X)]^2 f(x) dx$
Vzorčni standardni odklon (deviacija) = kvadratni koren variance $s = \sqrt{s^2}$	Standardni odklon = kvadratni koren variance $\sigma = \sqrt{\sigma^2}$

Mere razpršenosti - nadaljevanje

Na vzorcu velikosti n (empirična vrednost)	V populaciji (teoretična vrednost)
Interkvartilni razmik = razlika med 3. in 1. kvartilom empirične porazdelitve	Interkvartilni razmik = razlika med 3. in 1. kvartilom teoretične porazdelitve

Primer: Met kocke (nadaljevanje)
najprej izračunajmo empirične vrednosti.

$$\bar{x} = \frac{1}{20} \cdot (1 + 3 + 3 + 6 + 5 + \dots + 6 + 6 + 2 + 4 + 5) = \frac{1}{20} \cdot 76 = 3,8.$$

$$\begin{aligned}s^2 &= \frac{1}{20 - 1} \cdot [(1 - 3,8)^2 + (3 - 3,8)^2 + \dots + (5 - 3,8)^2] \\&= \frac{1}{19} \cdot 45,20 = 2,3789.\end{aligned}$$

Če opazovanja uredimo po velikosti, ugotovimo, da sta srednji vrednosti dve štirici (imamo sodo število opazovanj!) in je torej

$$\text{mediana} = \frac{4 + 4}{2} = 4.$$

Primer: Met kocke (nadaljevanje)

Če bi bila kocka popolnoma simetrična, bi bile verjetnosti vseh metov enake, torej

$$p(1) = p(2) = \dots = p(6) = \frac{1}{6}$$

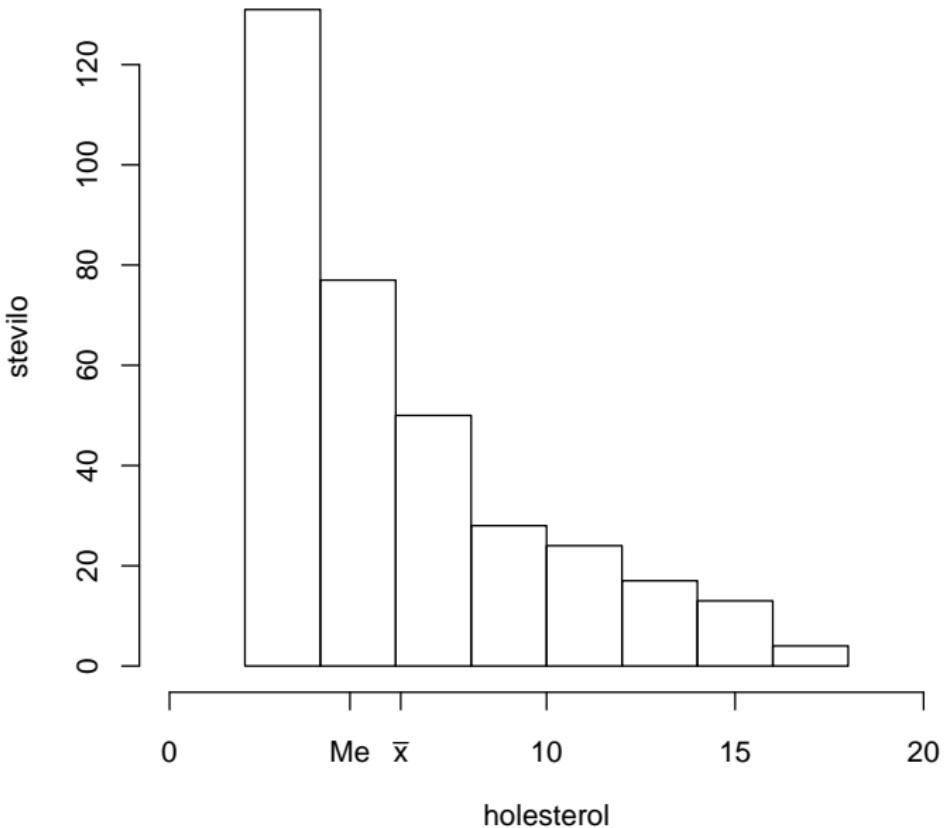
pričakovana vrednost in varianca pa

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5$$

$$\begin{aligned}Var(X) &= (1 - 3,5)^2 \cdot \frac{1}{6} + (2 - 3,5)^2 \cdot \frac{1}{6} + \dots + (6 - 3,5)^2 \cdot \frac{1}{6} \\&= 2,9167.\end{aligned}$$

Primerjava aritmetične sredine in mediane

Mediana razdeli vse podatke na dva dela, v vsakem je 50% podatkov. Aritmetična sredina je teoretično enaka mediani, kadar je porazdelitev simetrična. Za naš primer dečkov, rojenih v 38. tednu nosečnosti, sta tako aritmetična sredina 3314 gramov in mediana 3310 gramov. Na sliki na naslednji strani pa vidimo, da je holesterol zelo nesimetrično porazdeljen, kar ima za posledico precejšnjo razliko med aritmetično sredino in mediano (6,13 proti 4,79).



Primerjava aritmetične sredine in mediane - nadaljevanje

Razlike lahko povzročijo tudi tujki, to so nenavadne vrednosti daleč od 'sredine'. Te vplivajo na aritmetično sredino, ne pa na mediano.

Mediano bomo torej rajši uporabljali pri nesimetričnih porazdelitvah, če ne bomo posebej želeli, da mera središčnosti upošteva dejanske vrednosti spremenljivke. Pa še eno prednost ima mediana. Da bi jo določili moramo poznati le 50% vrednosti, za ostale je dovolj, če vemo, da so večje od teh. To nam pride posebej prav v analizi preživetja.

Povsem idealna seveda tudi mediana ni. Kot že rečeno, ne upošteva dejanskih vrednosti spremenljivke (razen pri rangiranju), zelo nerodna pa je za računanje. Tako se recimo mediana dveh združenih vzorcev ne da izraziti z medianama posameznih vzorcev.

LASTNOSTI PRIČAKOVANE VREDNOSTI IN VARIANCE

1. Za vse slučajne spremenljivke velja

$$E(a_0 + a_1 X_1 + \cdots + a_n X_n) = a_0 + a_1 E(X_1) + \cdots + a_n E(X_n)$$

2. Za **neodvisne** slučajne spremenljivke velja

$$\begin{aligned}Var(a_0 + a_1 X_1 + \cdots + a_n X_n) \\= (a_1)^2 Var(X_1) + \cdots + (a_n)^2 Var(X_n)\end{aligned}$$

Slučajni vzorec

Če so slučajne spremenljivke X_1, X_2, \dots, X_n paroma neodvisne in identično porazdeljene (iid), govorimo o **slučajnem vzorcu**.

Vzorčno povprečje $\bar{X} = (\sum X_i)/n$ je seveda spet slučajna spremenljivka.

Primer: Pričakovana vrednost in varianca vzorčnega povprečja

Slučajni vzorec

Če so slučajne spremenljivke X_1, X_2, \dots, X_n paroma neodvisne in identično porazdeljene (iid), govorimo o **slučajnem vzorcu**.

Vzorčno povprečje $\bar{X} = (\sum X_i)/n$ je seveda spet slučajna spremenljivka.

Primer: Pričakovana vrednost in varianca vzorčnega povprečja

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \cdot [X_1 + X_2 + \dots + X_n]\right) \\ &= \frac{1}{n} E(X_1 + X_2 + \dots + X_n) \\ &= \frac{1}{n} \cdot [E(X_1) + E(X_2) + \dots + E(X_n)] \\ &= \frac{1}{n} \cdot [\mu + \mu + \dots + \mu] = \mu \end{aligned}$$

$$\begin{aligned}Var(\bar{X}) &= Var\left(\frac{1}{n} \cdot [X_1 + X_2 + \dots + X_n]\right) \\&= \frac{1}{n^2} Var(X_1 + X_2 + \dots + X_n) \\&= \frac{1}{n^2} \cdot [Var(X_1) + Var(X_2) + \dots + Var(X_n)] \\&= \frac{1}{n^2} \cdot [\sigma^2 + \sigma^2 + \dots + \sigma^2] = \frac{\sigma^2}{n}\end{aligned}$$

$$sd(\bar{X}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Standardnemu odklonu vzorčnega povprečja rečemo
standardna napaka.

Risanje grafov

“... drawing graphs, like motor-car driving and love-making, is one of those activities which almost every researcher thinks he or she can do well without instruction.”

Wainer & Thissen, 1991 Annual Review of Psychology

Nekateri principi konstrukcije grafov

Nekateri principi konstrukcije grafov

- 1.** Izločite nepotrebne dimenzije

Nekateri principi konstrukcije grafov

1. Izločite nepotrebne dimenzije
2. Oznake meritve lestvice (“tick marks”) naj kažejo navzven

Nekateri principi konstrukcije grafov

1. Izločite nepotrebne dimenzije
2. Oznake meritve lestvice (“tick marks”) naj kažejo navzven
3. Razmislite o vključevanju ničle v graf (včasih dobro, včasih ne)

Nekateri principi konstrukcije grafov

1. Izločite nepotrebne dimenzije
2. Oznake meritve lestvice ("tick marks") naj kažejo navzven
3. Razmislite o vključevanju ničle v graf (včasih dobro, včasih ne)
4. Zaznavanje relativnih razdalj je najbolj natančno - ploščine so težje

Nekateri principi konstrukcije grafov

1. Izločite nepotrebne dimenzije
2. Oznake meritve lestvice ("tick marks") naj kažejo navzven
3. Razmislite o vključevanju ničle v graf (včasih dobro, včasih ne)
4. Zaznavanje relativnih razdalij je najbolj natančno - ploščine so težje
5. Izogibajte se odvečnosti - mislite na razmerje črnilo : informacija

Nekateri principi konstrukcije grafov

1. Izločite nepotrebne dimenzije
2. Oznake meritve lestvice ("tick marks") naj kažejo navzven
3. Razmislite o vključevanju ničle v graf (včasih dobro, včasih ne)
4. Zaznavanje relativnih razdalij je najbolj natančno - ploščine so težje
5. Izogibajte se odvečnosti - mislite na razmerje črnilo : informacija
6. Koristni grafi so lahko zahtevni - ne nujno preprosti in takoj dojemljivi

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Décrété par M. Minard, Inspecteur Général des Ponts et Chaussées en date du 6 juillet 1886.

Une autre fois juge à l'œil la diminution de l'âme, j'y suppose que les corps de Louis XIV et du Maréchal D'Artagnan qui avaient été dérobés sur l'île de Wight, ont rejoint les Cieux au-delà de l'au-delà, unis éternellement avec l'âme.

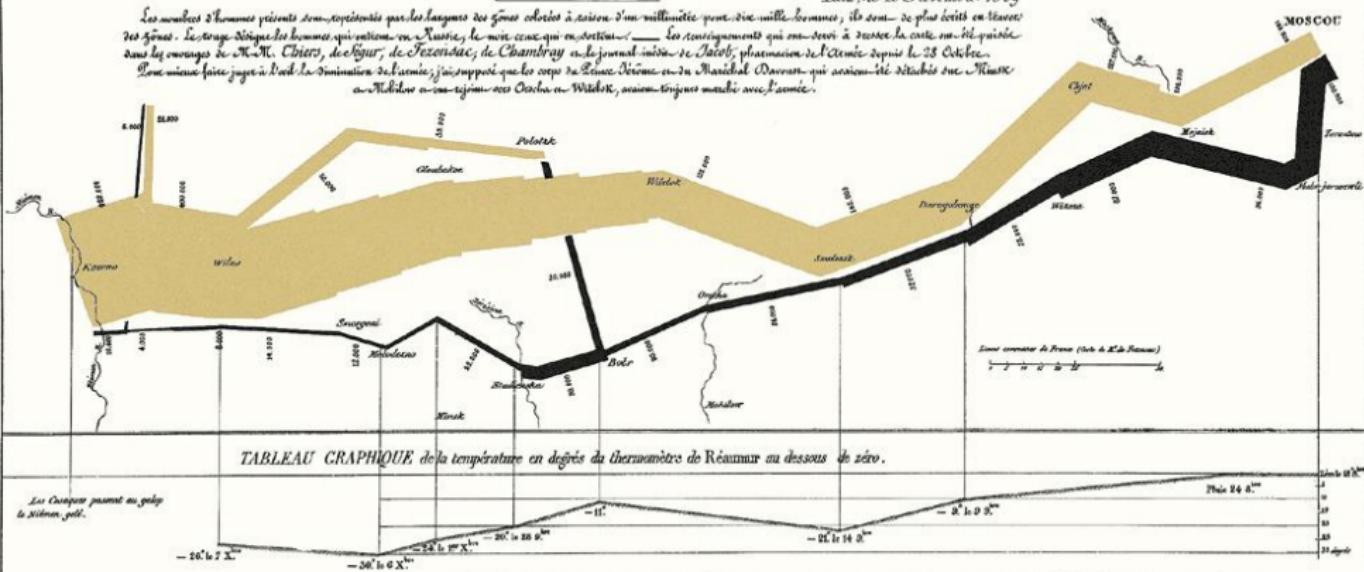


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

*Les Européens paient au gris
le Néerlandais, gold.*

Strukturni krog (torta) in stolpčni diagram

An Electronic Companion to Biostatistics TM

Diet Motivators

Goals Americans say would motivate them to diet if they were overweight:

A pie chart titled "Diet Motivators" showing the percentage distribution of reasons Americans would diet if overweight. The categories and percentages are: To be healthier (67%), To look better (21%), For better love life (6%), To get better job (3%), and Don't know (3%).

Motivator	Percentage
To be healthier	67%
To look better	21%
For better love life	6%
To get better job	3%
Don't know	3%

A bar chart titled "Diet Motivators" showing the percentage distribution of reasons Americans would diet if overweight. The categories and percentages are: To be healthier (67%), To look better (21%), For better love life (6%), To get better job (3%), and Don't know (3%).

Motivator	Percentage
To be healthier	67%
To look better	21%
For better love life	6%
To get better job	3%
Don't know	3%

Plots for **qualitative (categorical)** data must show two things: outcomes (categories), and how often each occurs (counts or percentages).

Bar charts show categories as bars, and counts or percentages as heights.

A bar chart whose bars are ordered from tallest (most frequent) to shortest is called a **Pareto chart**.

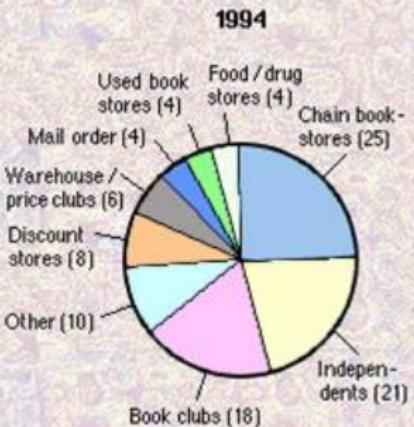
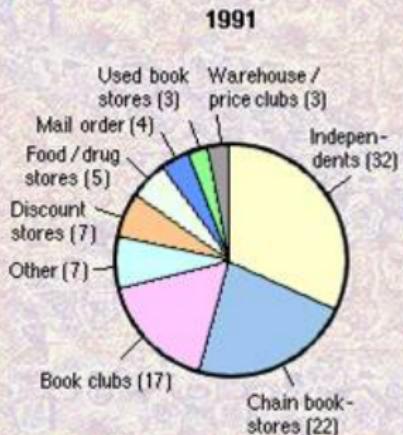
Pie charts show categories as wedges of a pie, percentages as areas (or angles).

Pie Charts

Bar Charts

Graphical Description Plots for Qualitative Data Topics Testing Options

Where People Buy Books About Diets



Pie charts are popular, but they are overused.

They show percentages using areas, which are proportional to angles. Because angles are hard to compare by eye, they don't convey numbers very well.

The example at left, which was published in a newspaper, is typical of poorly done pie charts. You get your information from the text, by reading the numbers. The picture doesn't help much.

Computer programs that produce pie charts automatically order the categories by size. Unfortunately, that changes the placement of categories, making the two pie charts harder to compare.

- Pie Charts
- Bar Charts

Graphical Description

Plots for Qualitative Data

Topics

Testing

Options

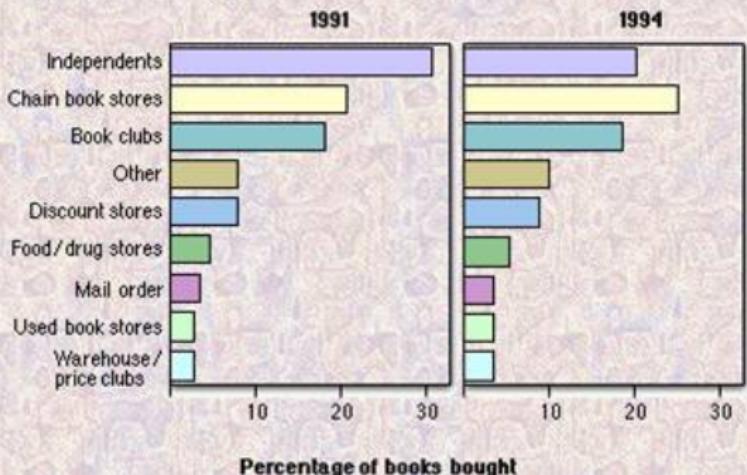
Where People Buy Books About Diets

Bar charts are usually more effective than pie charts.

They show counts or percentages by length, which our eyes find easier to compare than angles.

We can produce two bar charts. Each is easy to grasp, unlike the pie charts for the same data, but comparing the two is difficult.

One chart with both sets of data in it, such as the last chart here, can make the comparison easier.



- Poor presentation
- Better presentation

Pie Charts

Bar Charts

Graphical Description

Plots for Qualitative Data

Topics

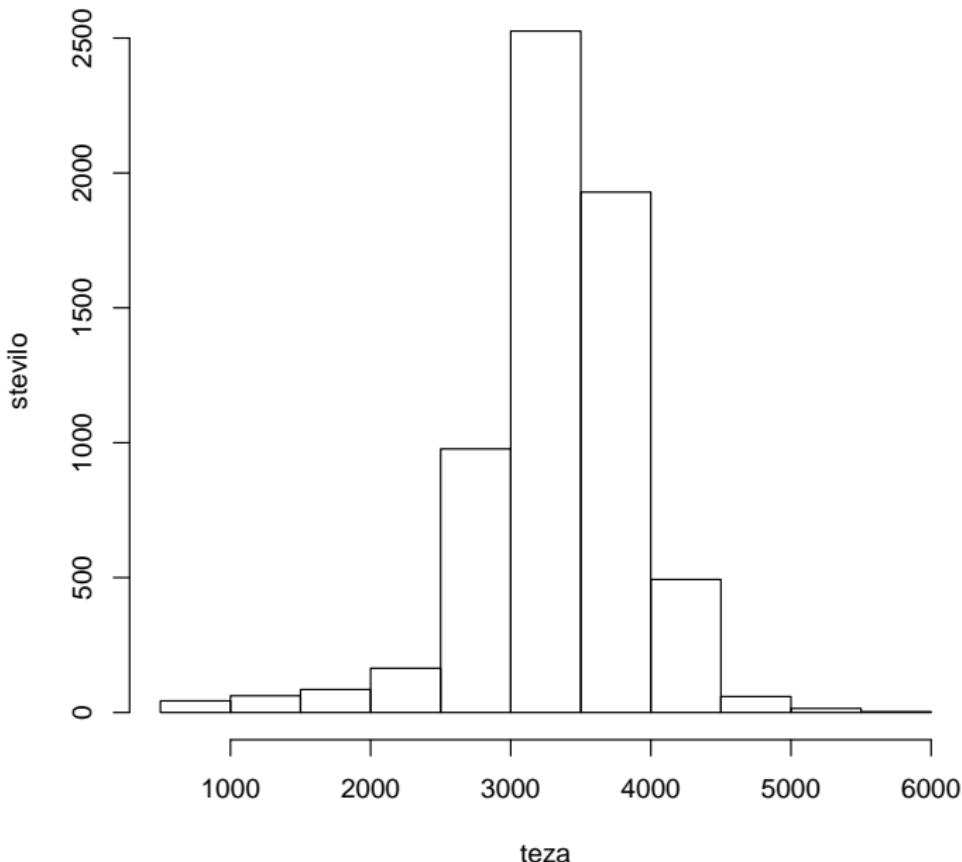
Testing

Options

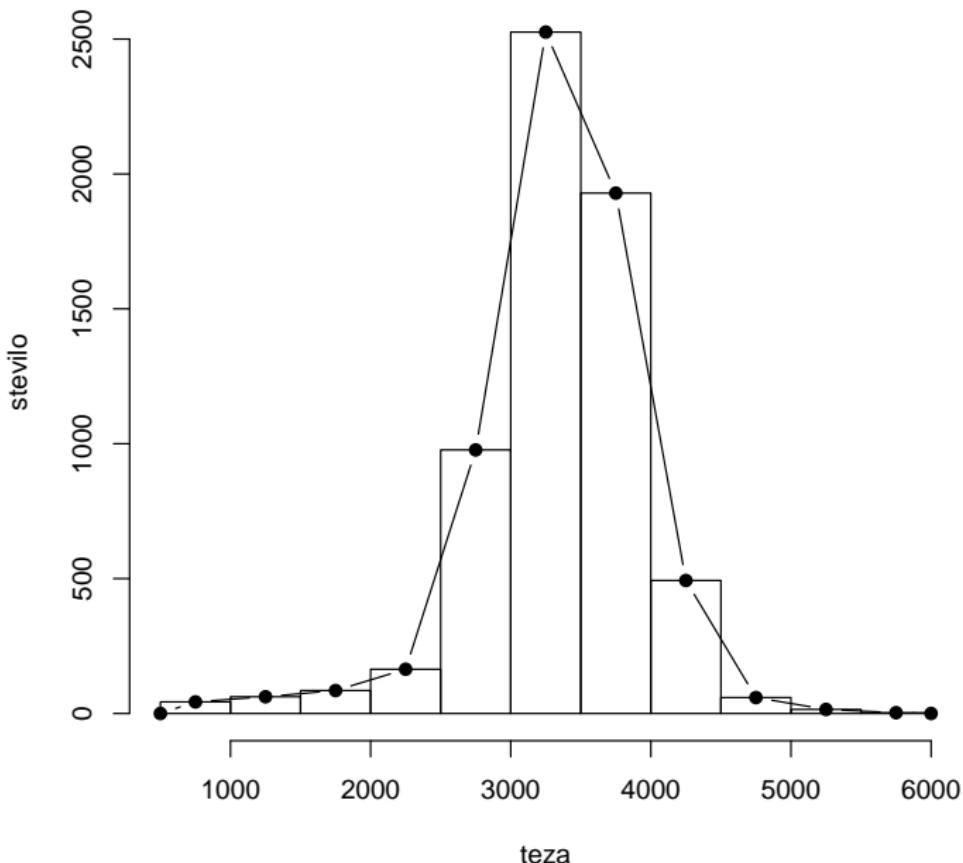
Točkovni diagram



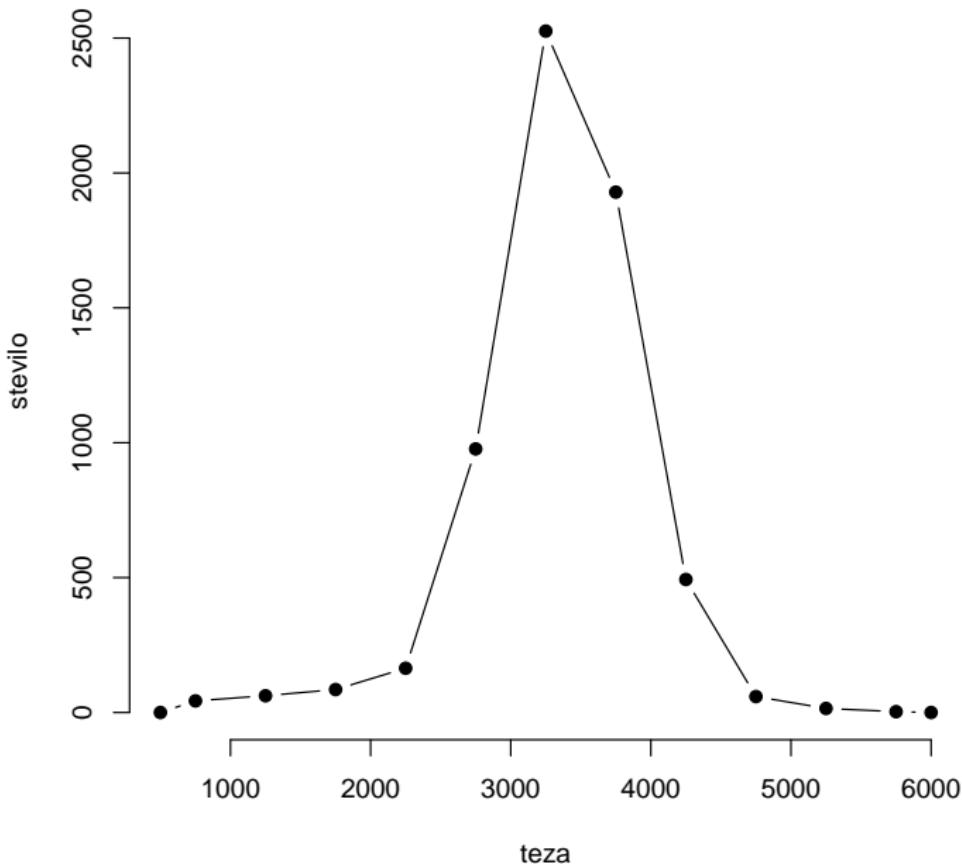
Histogram in frekvenčni poligon



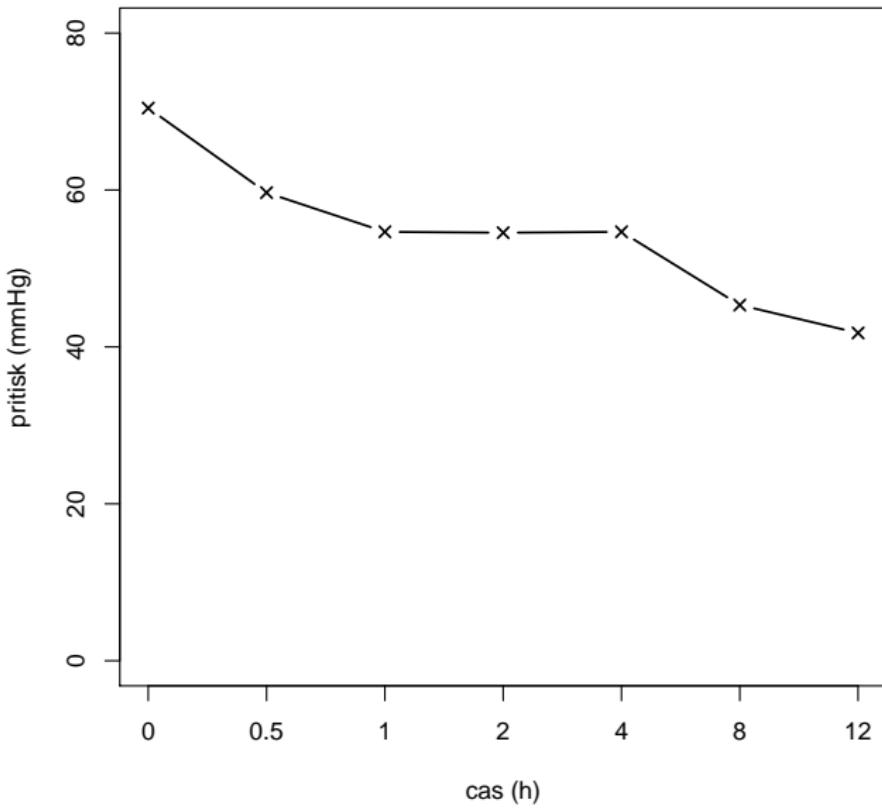
Histogram in frekvenčni poligon



Histogram in frekvenčni poligon



Linijski diagram



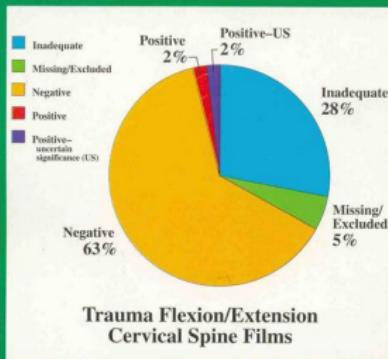


LIPPINCOTT
WILLIAMS & WILKINS

Volume 52 • Number 1 • January 2002

Full-Text Online from 1993
www.jtrauma.com

The Journal of
TRAUMA®
Injury, Infection, and Critical Care



FORT LEE NJ 07024-2546
CENTRAL MEDICINSKA KNUJZ JNR
000101288689 Th 12/10/00 0
CENTRAL MEDICAL LIBRARY
C/O AMERICAN SCIENTIFIC PUBLISHERS
5075 MAIN ST
FORT LEE NJ 07024-2546

CENTRALNA MEDICINSKA KNUJZ JNR

D26/I-IV
J Trauma
617



998 52 1

American Association for the Surgery of Trauma
Eastern Association for the Surgery of Trauma
Trauma Association of Canada/L'Association
Canadienne de Traumatologie
Western Trauma Association

www.jtrauma.com





The Journal of
TRAUMA®
Injury, Infection, and Critical Care

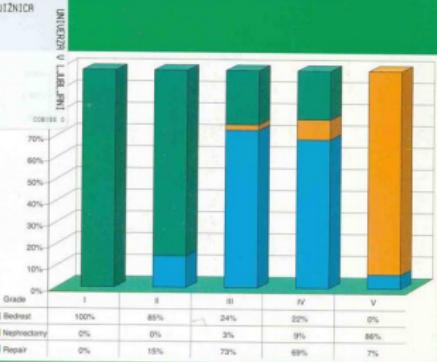
CENTRALNA MEDICINSKA KNIJIZNICA

D26/I-IV
J Trauma
617



998 50, 2

AUTO *****
0001012889890 TH 12/01/00 01 00 0002
CENTRAL MED LIB SLOMENJH
AMERICAN SCIENTIFIC PBL INC
507C MAIN ST
FORT LEE NJ 07024 2540



American Association for the Surgery of Trauma

Eastern Association for the Surgery of Trauma

Trauma Association of Canada/L'Association

Canadienne de Traumatologie

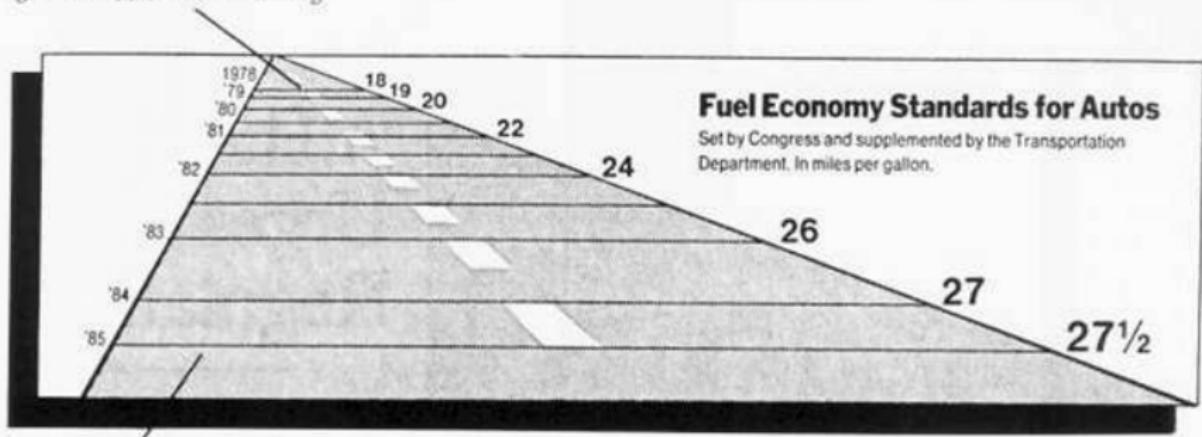
Western Trauma Association



www.jtrauma.com

En lažniv graf

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

New York Times, August 9, 1978, p. D-2.

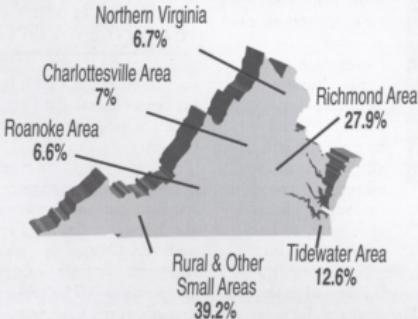
Who are Virginia's state employees?

Virginia's employees oversee the health and well-being of the Commonwealth's residents, its infrastructure, and its cultural, historical and natural resources.

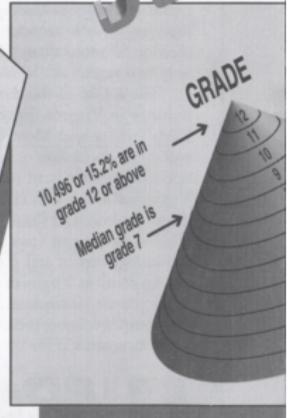
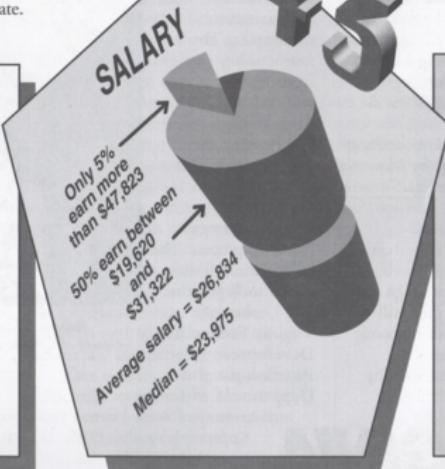
The statistics represented here provide an overview of the diversity of the state's work force by highlighting several areas of employee information, including location, salary, race, sex, occupational group, grade, age, and length of service with the state.

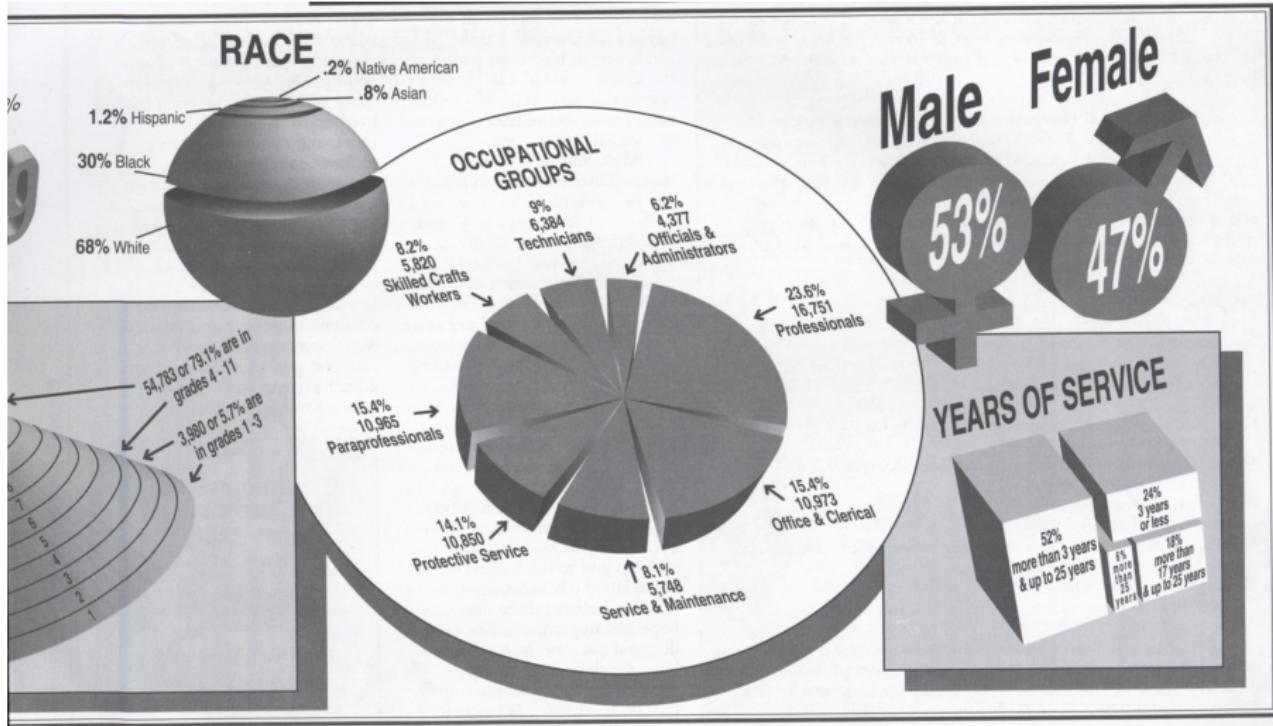
These figures represent non-faculty, classified state employees only.

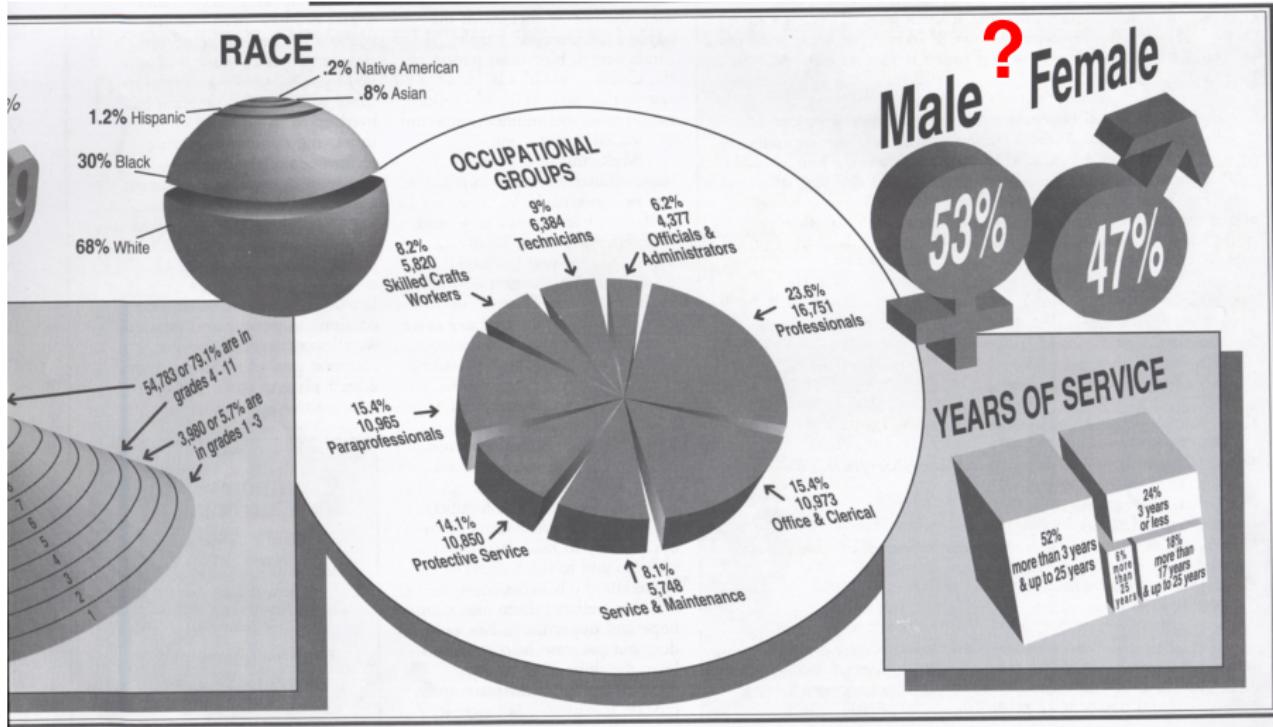
LOCATION



ONLY 7% ARE
27
AND UNDER
MEDIAN AGE IS
43
THE AGE OF 50
IS BETWEEN
35-44





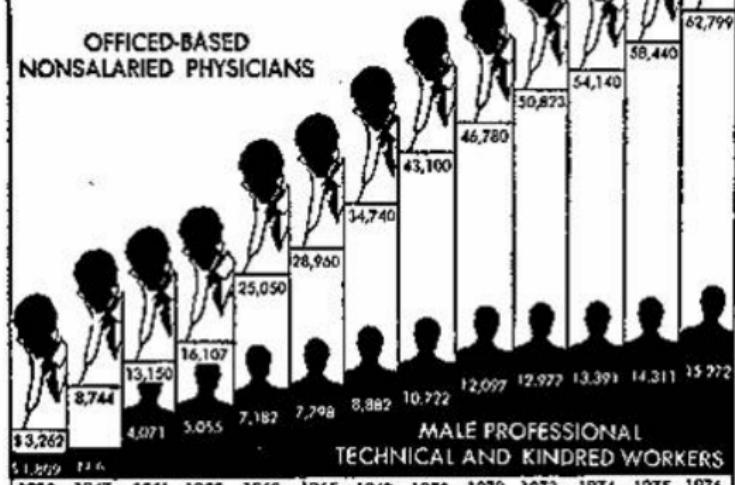


Incomes of Doctors Vs. Other Professionals

(MEDIAN NET INCOMES)

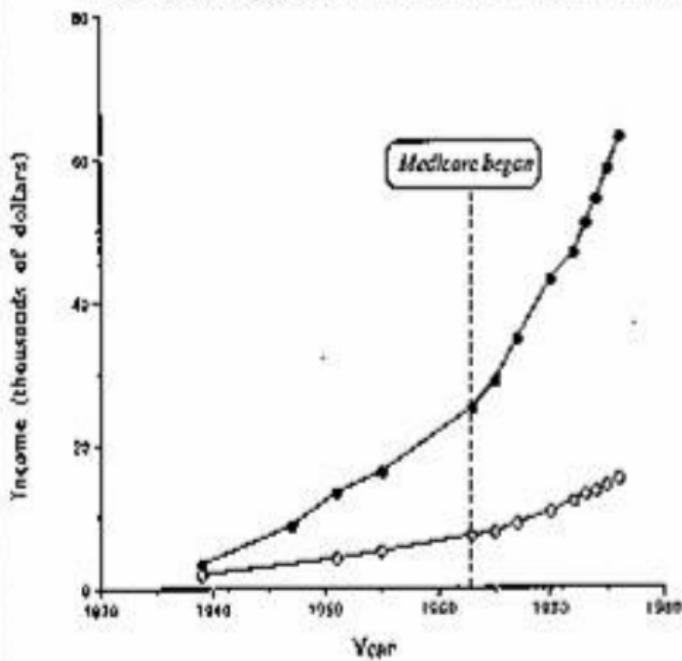
SOURCE: Council on Wage and Price Stability

OFFICED-BASED NONSALARIED PHYSICIANS



MALE PROFESSIONAL TECHNICAL AND KINDRED WORKERS

Physicians' income has grown exponentially since 1939
Whereas other professionals' income has gone up linearly



NORMALNA PORAZDELITEV

Normalna, ali Gaussova, porazdelitev je najpomembnejša teoretična porazdelitev. Njena gostota je podana s funkcijo

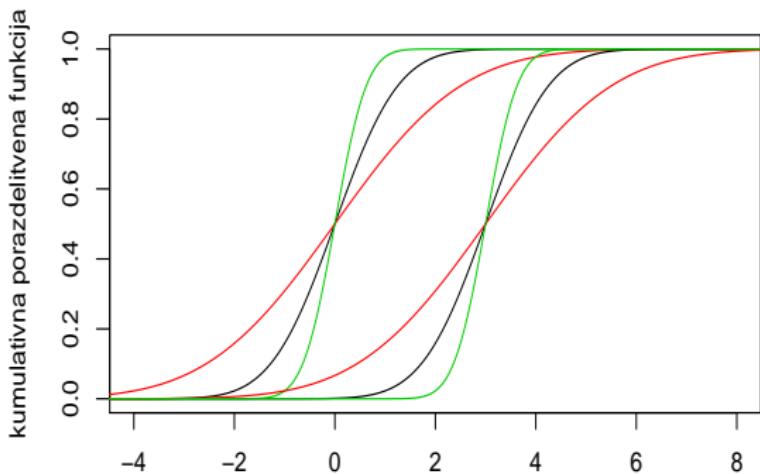
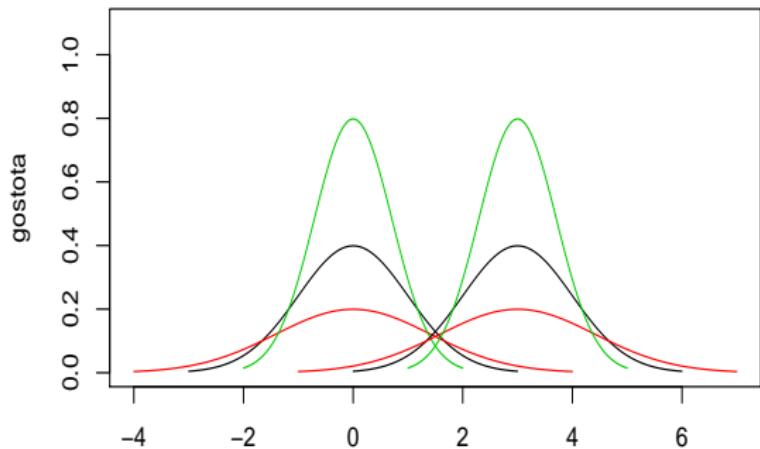
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Funkcija je torej popolnoma določena z dvema parametroma, **povprečjem (pričakovano vrednostjo) μ** in **varianco σ^2** .

Dejstvo, da je slučajna spremenljivka X normalno porazdeljena s povprečjem μ in varianco σ^2 , zapišemo kot $X \sim \mathcal{N}(\mu, \sigma^2)$.

Funkcija je **simetrična** okrog μ , torej je povprečje enako mediani. Če spremojamo μ , se graf gostote premika po x osi, če pa spremojamo σ^2 , se spreminja oblika zvona na sredini.

Naslednja slika ilustrira ti dve preprosti dejstvi. Dodani so tudi grafi porazdelitvene funkcije. Na levih delih grafov so predstavljene funkcije s povprečjem 0 in različnimi variancami, na desnih je povprečje 3.



Če je X normalno porazdeljena z $E(X) = \mu$ in $Var(X) = \sigma^2$, je

$$Y = a + b \cdot X$$

tudi normalno porazdeljena z

$$E(Y) = a + b\mu \quad \text{in} \quad Var(Y) = b^2\sigma^2.$$

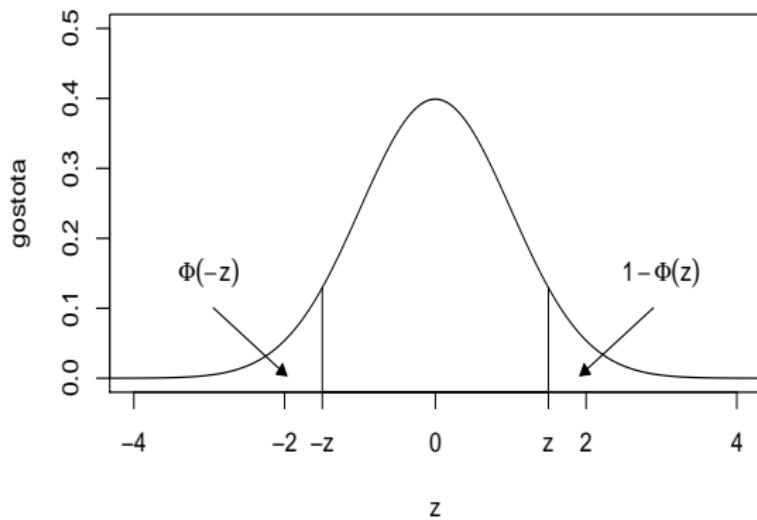
Standardizirana spremenljivka

$$Z = \frac{X - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma} \cdot X$$

je potem normalno porazdeljena z

$$E(Z) = 0 \quad \text{in} \quad Var(Z) = 1.$$

Gostoto standardizirane normalne porazdelitve ponavadi označujemo s φ , njeno porazdelitveno funkcijo pa s Φ .



Standardizirana normalna porazdelitev je tabelirana v statističnih tabelah, tipično za pozitivne z med 0 in 3 in z podanim na dve decimalki natančno. Tabele se razlikujejo po tem, kateri del ploščine, ki pripada z , tabelirajo: nekatere ploščino levo od z , druge desno od z , spet tretje ploščino v repih porazdelitve skupaj (torej desno od z in levo od $-z$) in tako naprej. Upoštevaje dejstvo, da je φ simetrična okrog 0 in da je celotna ploščina pod njo enaka 1, lahko vsako tabelo sprememimo v vsako drugo in so si torej enakovredne. In zato nam tudi ni treba tabelirati vrednosti za negativne z . Naslednja prosojnica prikazuje eno takšnih tabel.

Površina pod standardizirano normalno krivuljo



Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09		0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359	0	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753	1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141	1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517	1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879	1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224	2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549	2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852	2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133	2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389	2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621	2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830	2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015	2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177	2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319	2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441	3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4990	0,4990	0,4990

Poglejmo si **primer**:

Naj bo X normalno porazdeljena z $\mu = 10$ in $\sigma = 2$, torej $X \sim \mathcal{N}(10, 2^2)$. **Vprašanje 1:** Koliko je $P(7,9 < X \leq 11)$?

Za izračun uporabimo dejstvo, da je $Z = \frac{X - \mu}{\sigma}$ standardizirano normalno porazdeljena!

$$\begin{aligned} P(7,9 < X \leq 11) &= P\left(\frac{7,9 - 10}{2} < \frac{X - 10}{2} \leq \frac{11 - 10}{2}\right) \\ &= P(-1,05 < Z \leq 0,5) \\ &= P(Z \leq 0,5) - P(Z \leq -1,05) \\ &= P(Z \leq 0,5) - [1 - P(Z \leq 1,05)] \\ &= 0,6915 - [1 - 0,8531] = 0,5446. \end{aligned}$$

Iz naše tabele smo za ustrezne verjetnosti sicer dobili 0,1915 in 0,3531, ki pa smo jima prišteli 0,5.

Vprašanje 2: Koliko je tretji kvartil za X ?

Določiti moramo torej vrednost a , za katero velja

$$P(X \leq a) = 0,75.$$

Zopet uporabimo dejstvo, da je $Z = (X - 10)/2$ standardizirana normalna spremenljivka. Potem je gornja zahteva isto kot

$$P\left(Z \leq \frac{a - 10}{2}\right) = 0,75.$$

Iz tabele vidimo, da je približno 25% vrednosti med 0 in 0,67, kar pomeni, da jih je levo od 0,67 približno 75%. Ustrezen z je torej 0,67, se pravi, da iz $(a - 10)/2 \approx 0,67$ sledi, da je

$$a = 10 + 2 \cdot 0,67 = 11,34.$$

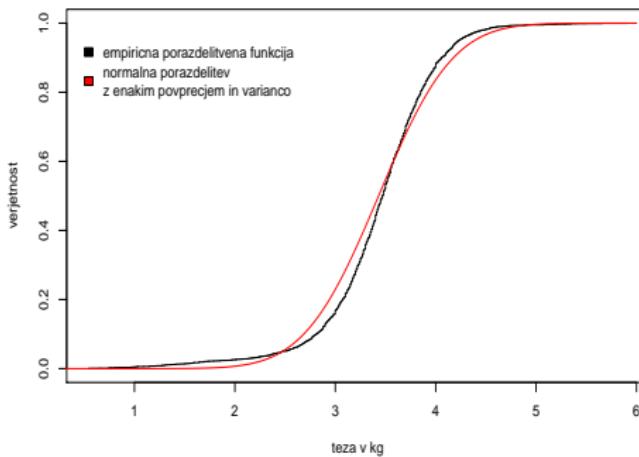
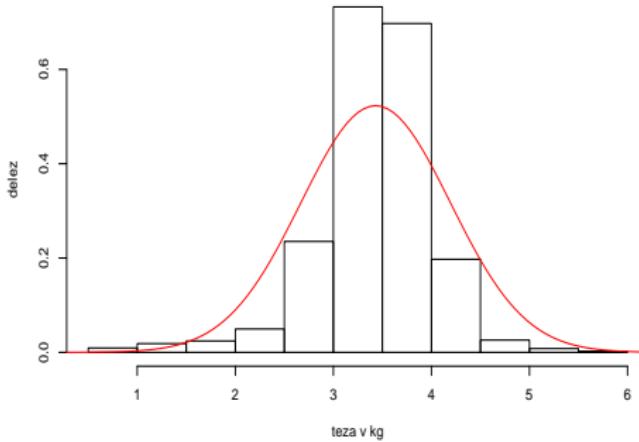
Kot bomo videli kasneje, nas v statistiki posebej zanimajo nekatere vrednosti z . Najpogosteje se srečamo z vrednostjo, pri kateri v repih porazdelitve ostane še 5% vseh vrednosti. To se zgodi pri $z = \mathbf{1,96}$, kar je številka, ki si jo velja zapomniti. V spodnji tabeli je poleg te še nekaj zanimivih vrednosti.

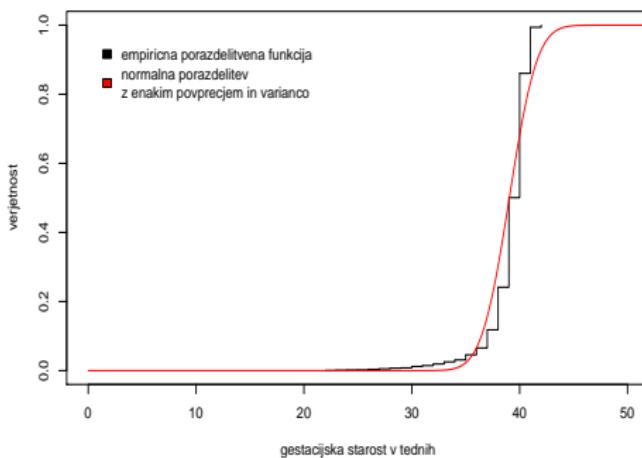
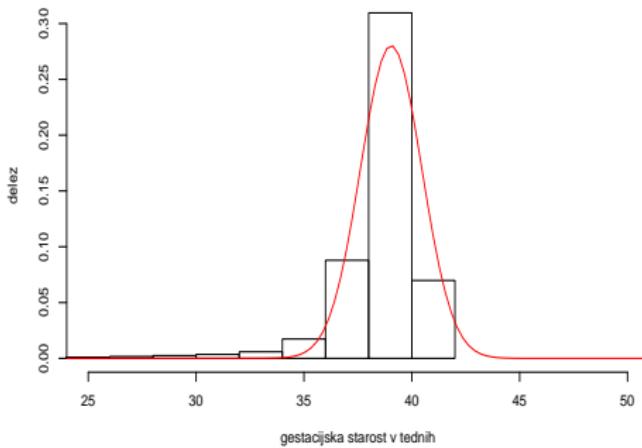
x	$P(X > \mu + x) + P(X < \mu - x)$
0	1
σ	0,3174
$1,96 \cdot \sigma$	0,05
$2 \cdot \sigma$	0,0455
$3 \cdot \sigma$	0,0027
$4 \cdot \sigma$	0,00006334

Q-Q grafi

Q-Q graf je eden od grafičnih načinov preverjenja predpostavke, da je neka spremenljivka normalno porazdeljena. Načeloma naj bi bilo to razvidno tudi iz grafa kumulativne porazdelitvene funkcije ali iz histograma, vendar oba načina nista preveč občutljiva na odstope od normalne porazdelitve.

Naslednja dva primera to ilustrirata.

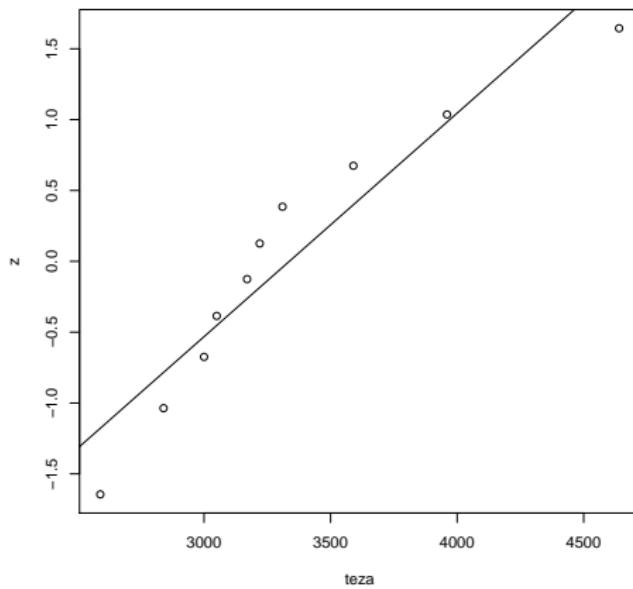




Princip Q-Q grafov je na kratko tale: na abscisno os nanesemo izmerjene vrednosti, na ordinatno pa ustreerne z vrednosti, ki bi pripadale x -om glede na njihov rang ob predpostavki, da je X normalno porazdeljena spremenljivka. Ustrezni graf mora biti približno linearen, ker je teoretično $z = (x - \mu)/\sigma$.

Primer: Tule je 10 naključno izbranih vrednosti porodnih tež dečkov, rojenih v 38. tednu nosečnosti: 2590, 2840, 3000, 3050, 3170, 3220, 3310, 3590, 3960, 4640. Če naj bi prihajale te vrednosti iz normalne porazdelitve, potem prvi vrednosti pripada tisti z , pod katerim leži $1/10$ vseh vrednosti, drugi tisti, pod katerim je $2/10$ vseh vrednosti in tako naprej. No, ker bi za zadnjo vrednost dobili $1 (= 10/10)$, stvar nekoliko popravimo in namesto $i/10$ uporabljamo $(i - 0,5)/10$ (obstajajo tudi bolj komplikirani pristopi).

Q-Q graf



Histogram teze

