The 13th APEC-Khon Kaen International Symposium:
Bridge 12 years of APEC Lesson Study to InMside High Quality Curriculum for Digital Economies
8-11 September 2018

STATISTICAL THINKING

AS A PART OF HUMAN'S CULTURE

IVAN VYSOTSKIY, MCCME, MOSCOW

CONTENT

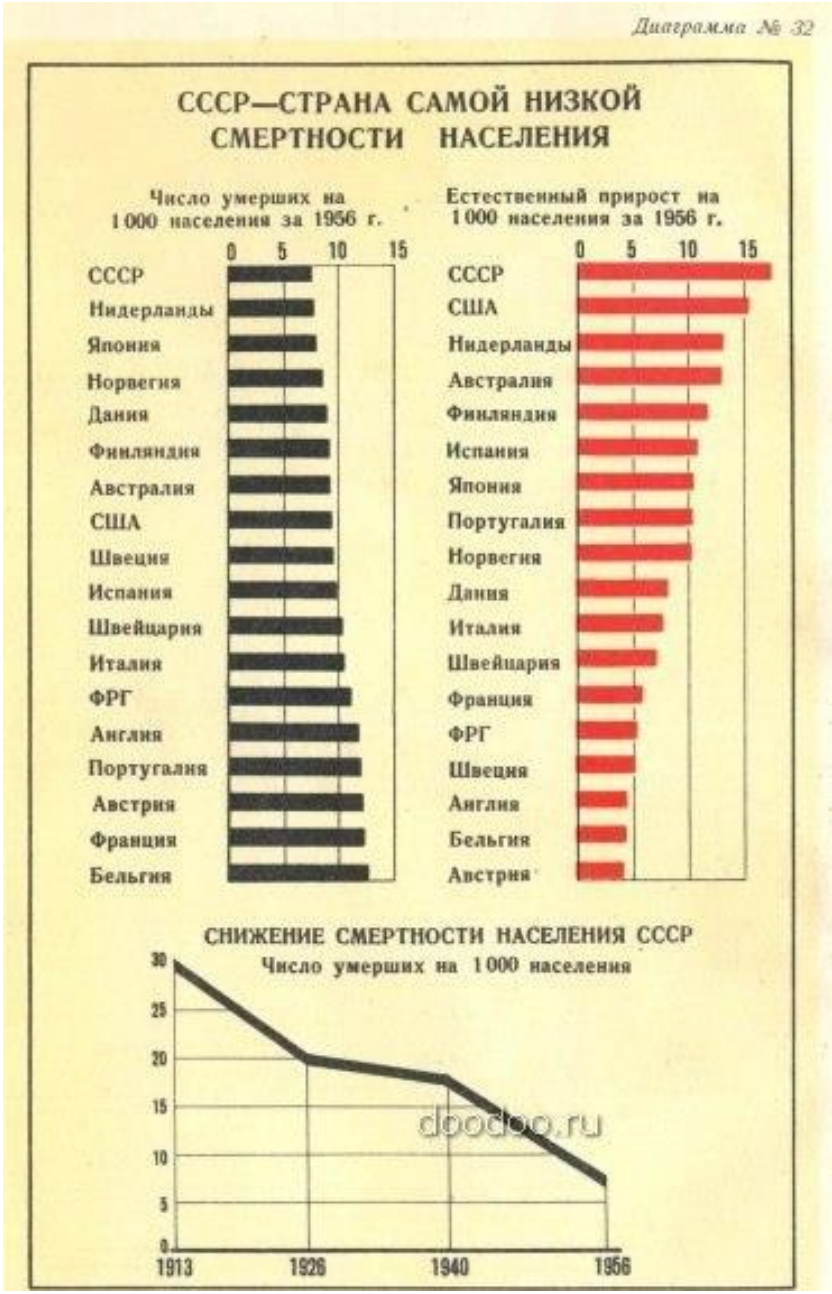
- STATISTICS AS A SCIENCE / INFO / CULTURAL PHENOMENA
- STATISTICAL THINKING
- STATISTICS (& PT) IN SCHOOL

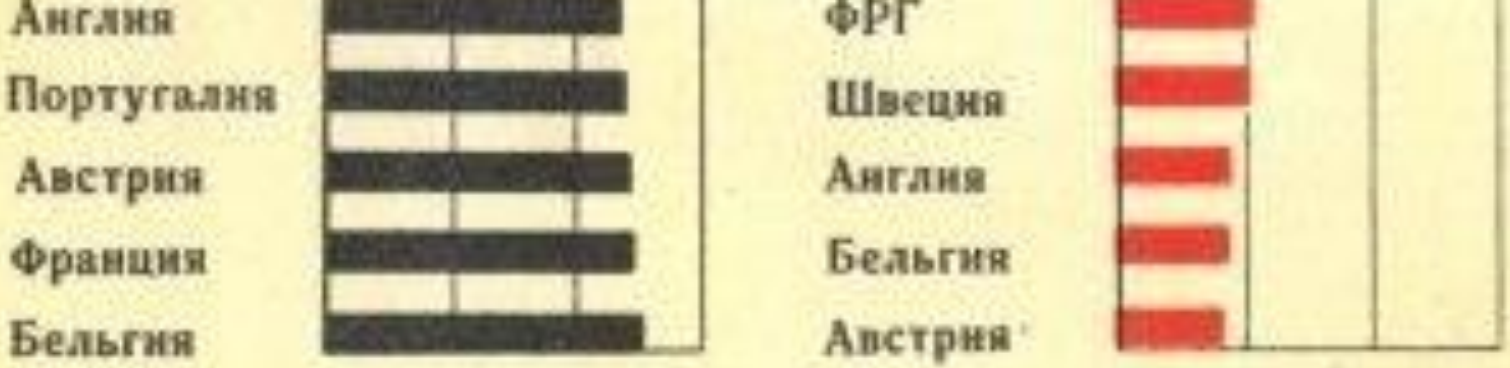
THERE ARE THREE

kinds of lies:

lies,
damned lies,
and statistics

Sir Charles Wentworth Dilke





СНИЖЕНИЕ СМЕРТНОСТИ НАСЕЛЕНИЯ СССР

Число умерших на 1000 населения

STATISTICS – FROM "STATE"

One of the main tasks of statistics is providing information for managing whatever you're going to manage

Statistics – way to make conclusions and decisions

STATISTICS GIVES VITAL INFO

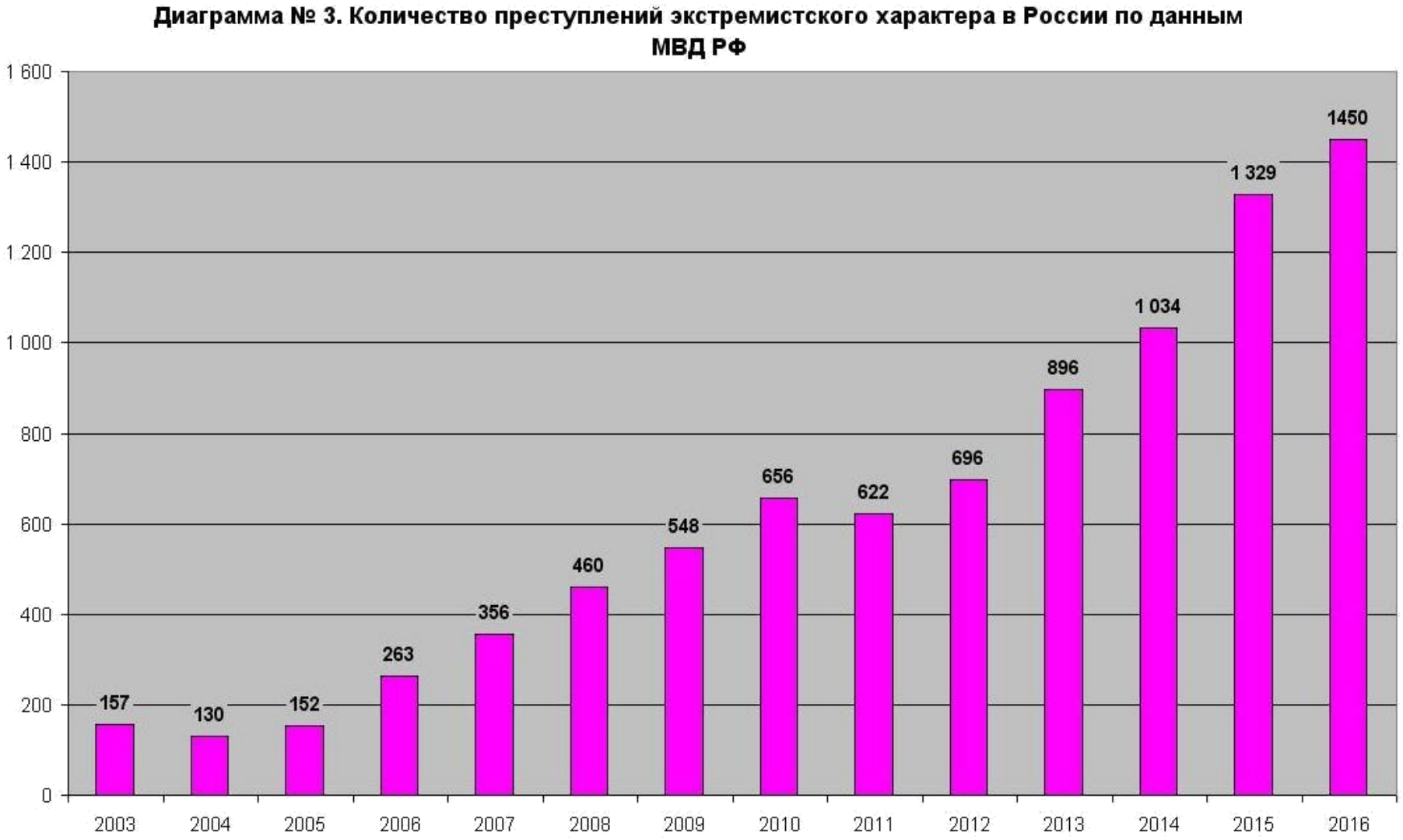
WHEN DRIVING, YOU NEED INFO FROM THE DASHBOARD ABOUT SPEED, LOCATION AND STATE OF YOUR CAR AND ITS SYSTEMS

WHEN MANAGING YOUR BUSINESS, YOU NEED INFO ABOUT STATE OF YOUR COMPANY AND ITS PROGRESS

NEGLECTING OR MISUSING STATISTICS MAY LEAD TO COLLAPS

- Eternal desire to wishful thinking (Russian default 1998)
- ‘Apparent’ but false correlations (red vine and health index)
- Survivor’s mistakes
- Mechanical using methods and formulas without understanding their applicability (W-for-H index)
- Intentional (evil- or well-minded) manipulations

MANIPULATING



COMMON MISTAKES

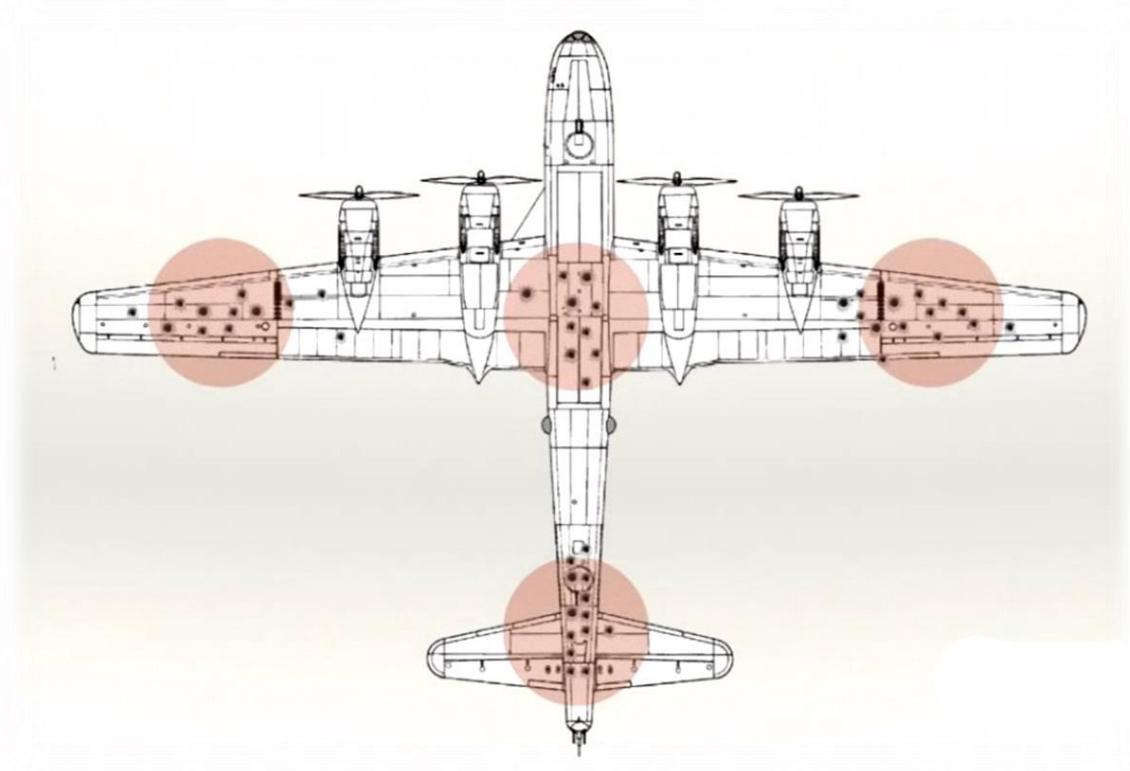
Well known researches on benefits and harm of red wine

1. In Denmark since 1973 the health index goes up as after joining the country EEC, consumption of red wine had significantly increased.

2. A research group from Med.School at University of Johns Hopkins had scrutinized useful properties of antioxidant resveratrol which red wine contains. 800 thousand took part in the research. Results were negative – no connection was found between health level and index of resveratrol in blood.

SURVIVOR'S MISTAKE

During the WWII, the command of the American and British air forces instructed Abraham Wald to find out what parts of the fuselage of the aircraft need to protect additional armor. Wald studied the planes returning from sorties, noting the place of holes concentrating.



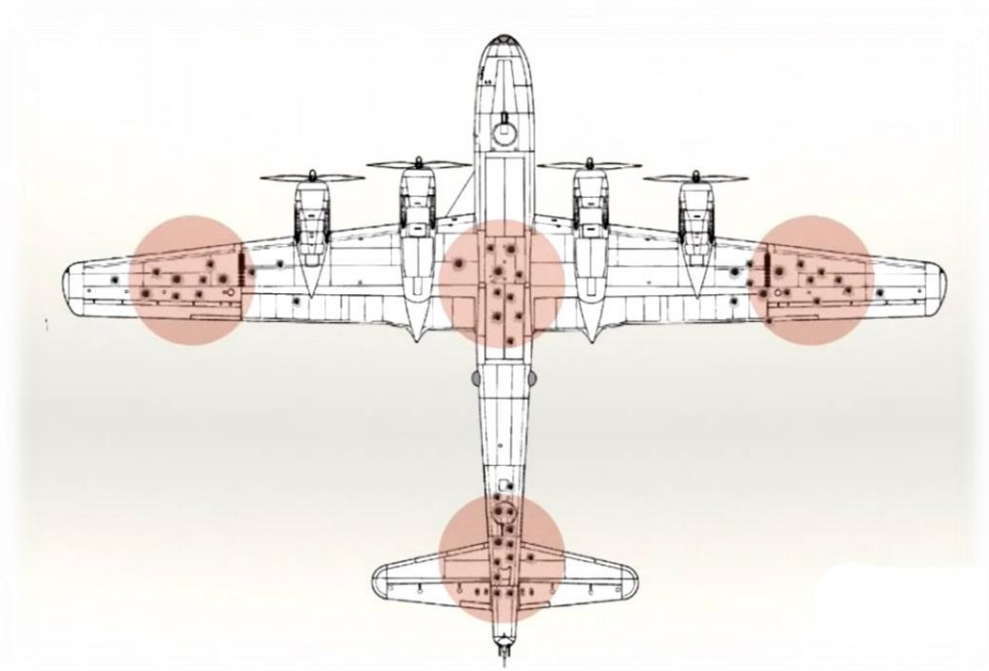
SURVIVOR'S MISTAKE

What do you think Abraham Wald suggested?

SURVIVOR'S MISTAKE

Wald's recommendation was to put the additional protection on those areas, where the number of holes was minimal.

The recommendation was based on the conclusion that it is necessary to protect against those hits that Wald did not see-the planes that received them, did not return



EXAMPLE OF MISUNDERSTANDING

Table 4.1: The results of multiple regression analysis on the atmospheric aerosols concentration data (DustTrak) and the data of reflectance from atmospheric aerosols (Logitech C922 webcam) for best algorithm determination.

Algorithm	R^2	RMS ($\mu g/m^3$)
$P = a_0 + a_1 R_R + a_2 R_R^2$	0.6528	13
$P = a_0 + a_1 R_G + a_2 R_G^2$	0.6823	11
$P = a_0 + a_1 R_B + a_2 R_B^2$	0.6727	12
$P = a_0 + a_1 \ln R_R + a_2 (\ln R_R)^2$	0.4568	18
$P = a_0 + a_1 \ln R_G + a_2 (\ln R_G)^2$	0.4659	17
$P = a_0 + a_1 \ln R_B + a_2 (\ln R_B)^2$	0.4897	17
$P = a_0 + a_1 (R_R/R_G) + a_2 (R_R/R_G)^2$	0.2999	25
$P = a_0 + a_1 (R_R/R_B) + a_2 (R_R/R_B)^2$	0.3345	24
$P = a_0 + a_1 (R_G/R_B) + a_2 (R_G/R_B)^2$	0.3173	25

EXAMPLE OF MISUNDERSTANDING

$P = a_0 + a_1(R_R + R_G)/R_B + a_2[(R_R + R_G)/R_B]^2$	0.3943	20
$P = a_0 + a_1(R_R + R_B)/R_G + a_2[(R_R + R_B)/R_G]^2$	0.4153	18
$P = a_0 + a_1(R_G + R_B)/R_R + a_2[(R_G + R_B)/R_R]^2$	0.3896	21
$P = a_0 + a_1(R_R - R_G)/R_B + a_2[(R_R - R_G)/R_B]^2$	0.4226	19
$P = a_0 + a_1(R_R - R_B)/R_G + a_2[(R_R - R_B)/R_G]^2$	0.4328	19
$P = a_0 + a_1(R_G - R_B)/R_R + a_2[(R_G - R_B)/R_R]^2$	0.4542	18
$P = a_0 + a_1(R_R - R_G) + a_2(R_R - R_G)^2$	0.6453	13
$P = a_0 + a_1(R_R - R_B) + a_2(R_R - R_B)^2$	0.6756	11
$P = a_0 + a_1(R_G - R_B) + a_2(R_G - R_B)^2$	0.6652	12
$P = a_0 + a_1R_R + a_2R_G$	0.7358	9
$P = a_0 + a_1R_R + a_2 R_B$	0.7225	10
$P = a_0 + a_1R_G + a_2 R_B$	0.7334	9
* $P = a_0 + a_1R_R + a_2R_G + a_3R_B$	**0.8493**	**5**

*The best algorithm identified is $P = a_o + a_1R_R + a_2R_G + a_3R_B$ with $R^2 = 0.8493$ and RMS = 5.

STATISTICS IN SCHOOL

- Statistics and probabilities are inalienable subjects like drawing and geometry
- Statistics is a primary of those two
- Since a certain age (when students are ready for concept of changebility)

CURRICULUM

MATHEMATICAL VERTICAL

Statistics. Grade 7. (1 — 2 h per week. 34 — 68 in total)
(based on FES and tentative curriculum on math)

Topics	Ac. hours *	
	Basic (1–1,5 h/week)	Deepened (1,5–2 h/week)
1. Stat data. Data representation and search in tables. Practical calculation in tables	2–3	3
2. Practical work ** № 1 «Tables»	1–2	2
3. Graphical representations: pie- and toll charts. Examples (demography, family's budget distribution, nutritious elements distribution etc.)	2–3	3
4. *Scattering charts for paired observations****. *Examples (height and weight, temperature and blood pressure)*	0–1	1–3
5. Practical work № 2 «Charts»	1	2

CURRICULUM

7. Diagnostic work «Data representation»	1	1
8. The descriptive statistics. Number arrays, mean. Examples (yearly income, mean height, mean temperature, frequency of an event as a mean)	3–4	4
9. *The geometric mean. Examples (deposit or credit present, mean GDP yearly growth)*	0–1	1–2
10. *The harmonic mean. Examples. (velocity, labor capacity)*	0–1	1–2
11. The median. Median's stability. *A median representative of a value.* Examples (populations, rivers' lenght)	2	2–3
12. Max and min. The span. Examples (sports records, spring floods levels, minimal salary level)	2	2–3
13. Practical work № 3 «Means»	1–2	2
14. *Data scattering. Deviation from arithmetic mean. Variance (dispersion) of a number array. Examples (school marks, height).*	0–2	2–4

CURRICULUM

15. *Properties of the arithmetic mean and variance. Solving problems*	0–2	2–4
16. Practical work № 4 «Scattering measures»	0–2	2
17. Diagnostic work «Descriptive statistics»	1	1
18. Everyday examples of random changeability. Human's height, precision (weighting, measuring height, size…)	2	2–4
19. Practical work № 5 «Random changeability. Measurements»	1	2
20. Repeating and generalization of the topics learnt	4	4–6
21. Control work on the chapter "Statistics"	1	1

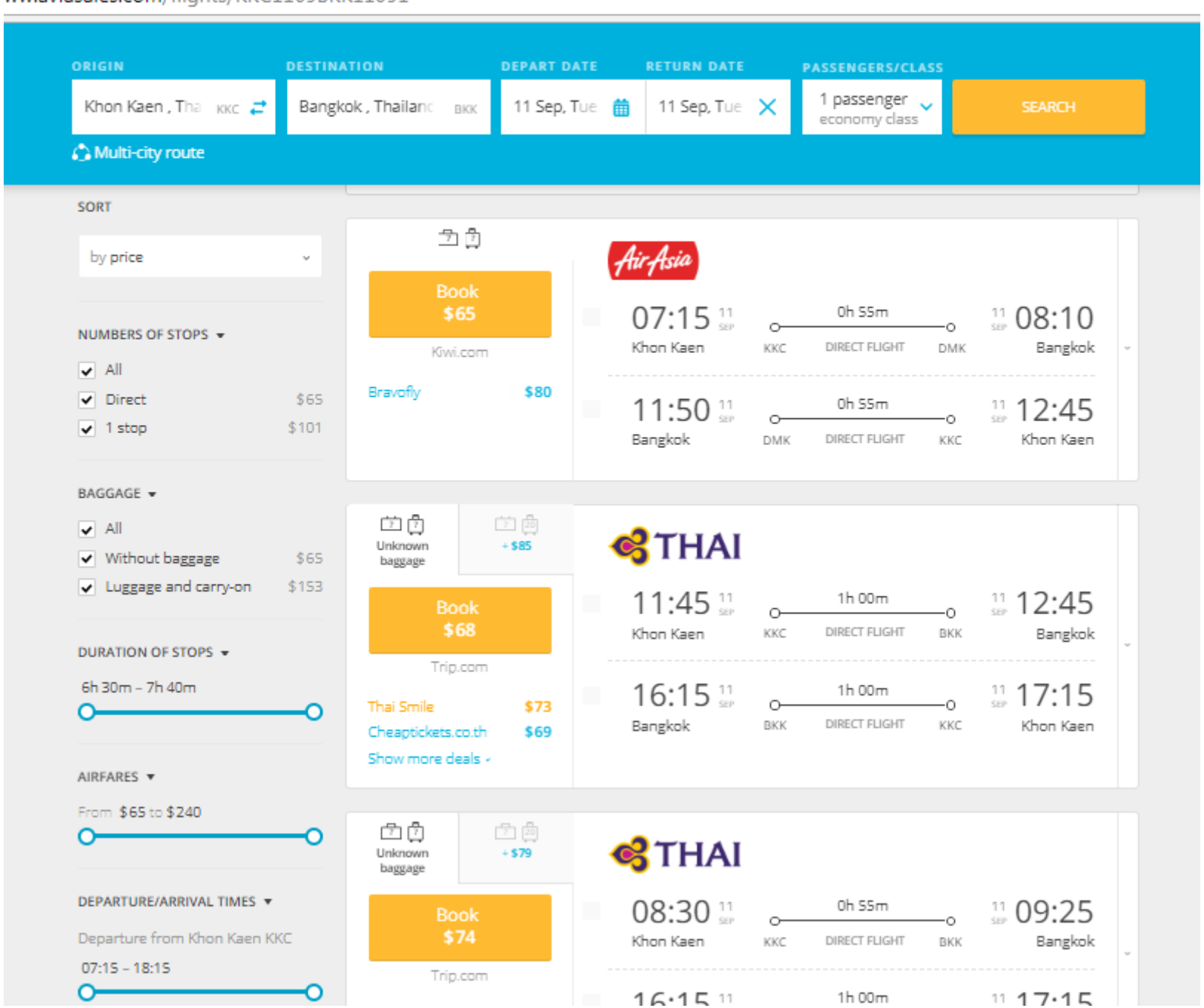
* Number of hours could differ from recommended due to time reserve
** Practical works designed for applying computers
*** *Italic* marks topics which aren't compulsory for the basic course

MAKING THE CONTENT MEANS MAKING PROBLEMS

Suppose, tomorrow you have a very important meeting in Bangkok. It is appointed in Radisson Hotel at 11 am on dot.

It will last one hour after what you have to get back to the afternoon APEC session in KKU

Problem 1. Meeting in Bangkok



Problem 1. Meeting in Bangkok

SELECTED PROBLEMS

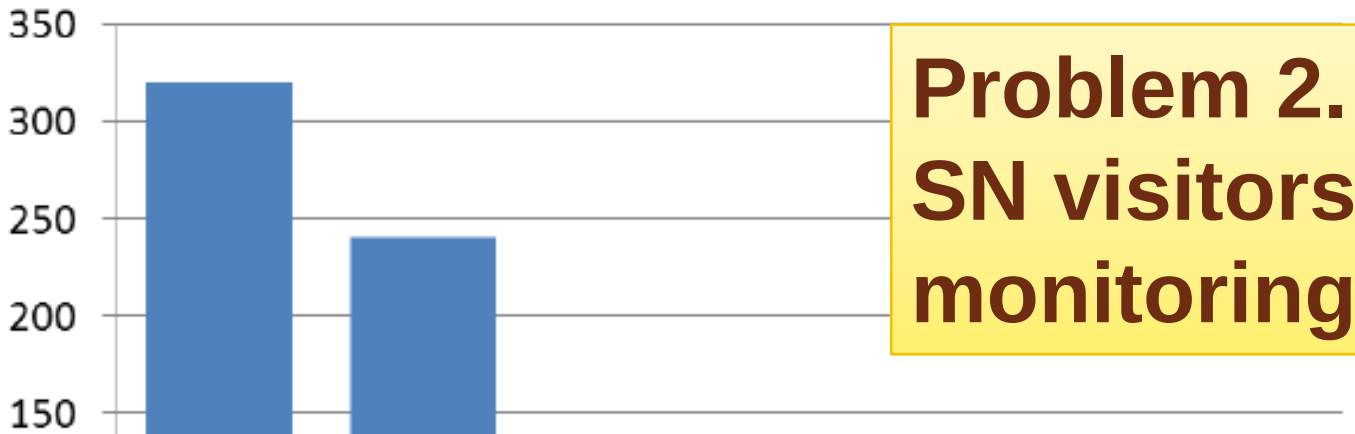
The table shows the number of visitors to one of the groups of the social network "VKontakte" from different cities on April 7.

A diagram (below) was plotted using the data from the table.

a) Determine what percent of visitors is from Omsk.

b) Determine to which city the second column of the diagram corresponds.

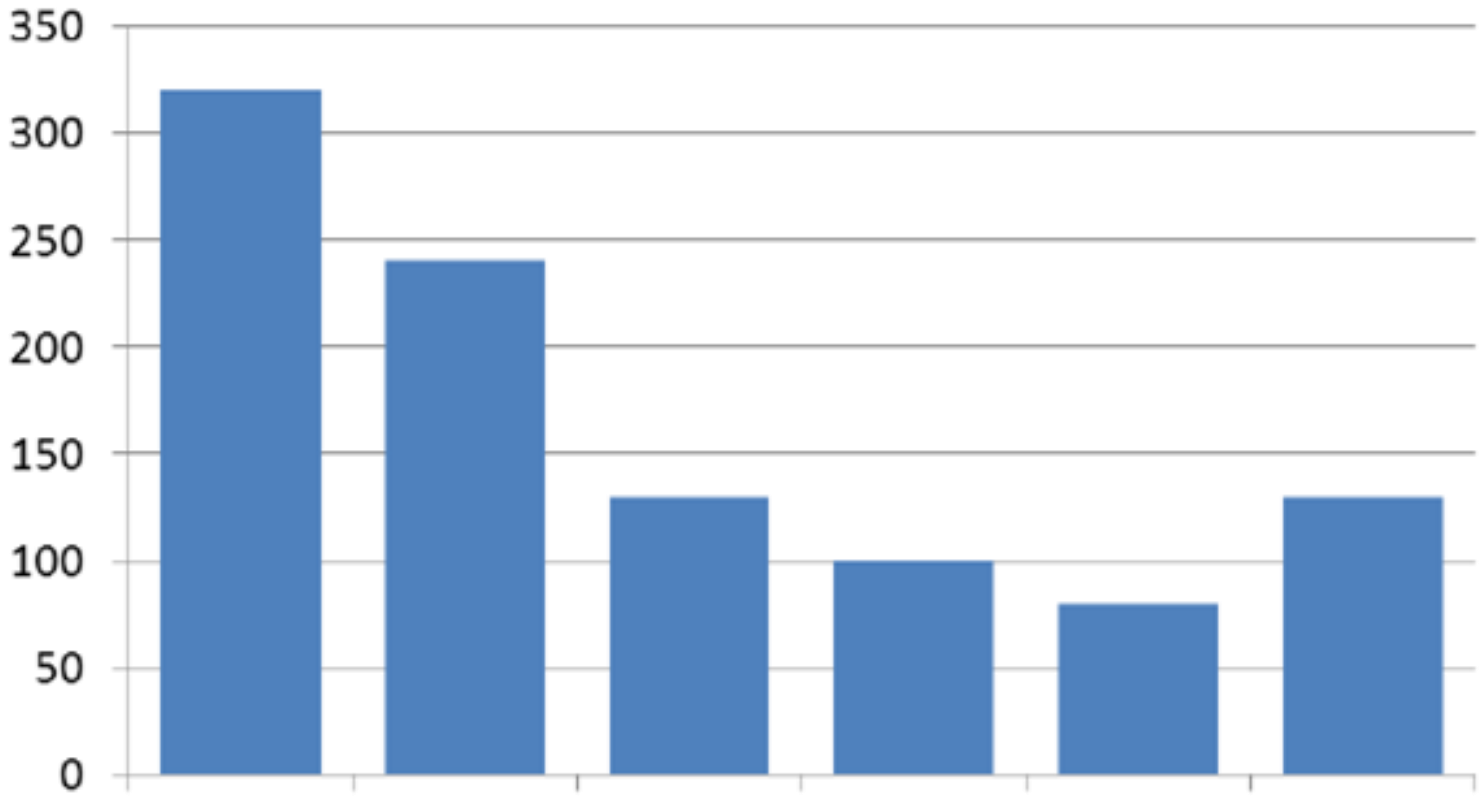
City	Number of visitors
Moscow	320
Omsk	130
Nizhny Novgorod	80
Tyumen	100
Saint Petersburg	240
Other cities	130



Problem 2. SN visitors monitoring

SELECTED PROBLEMS

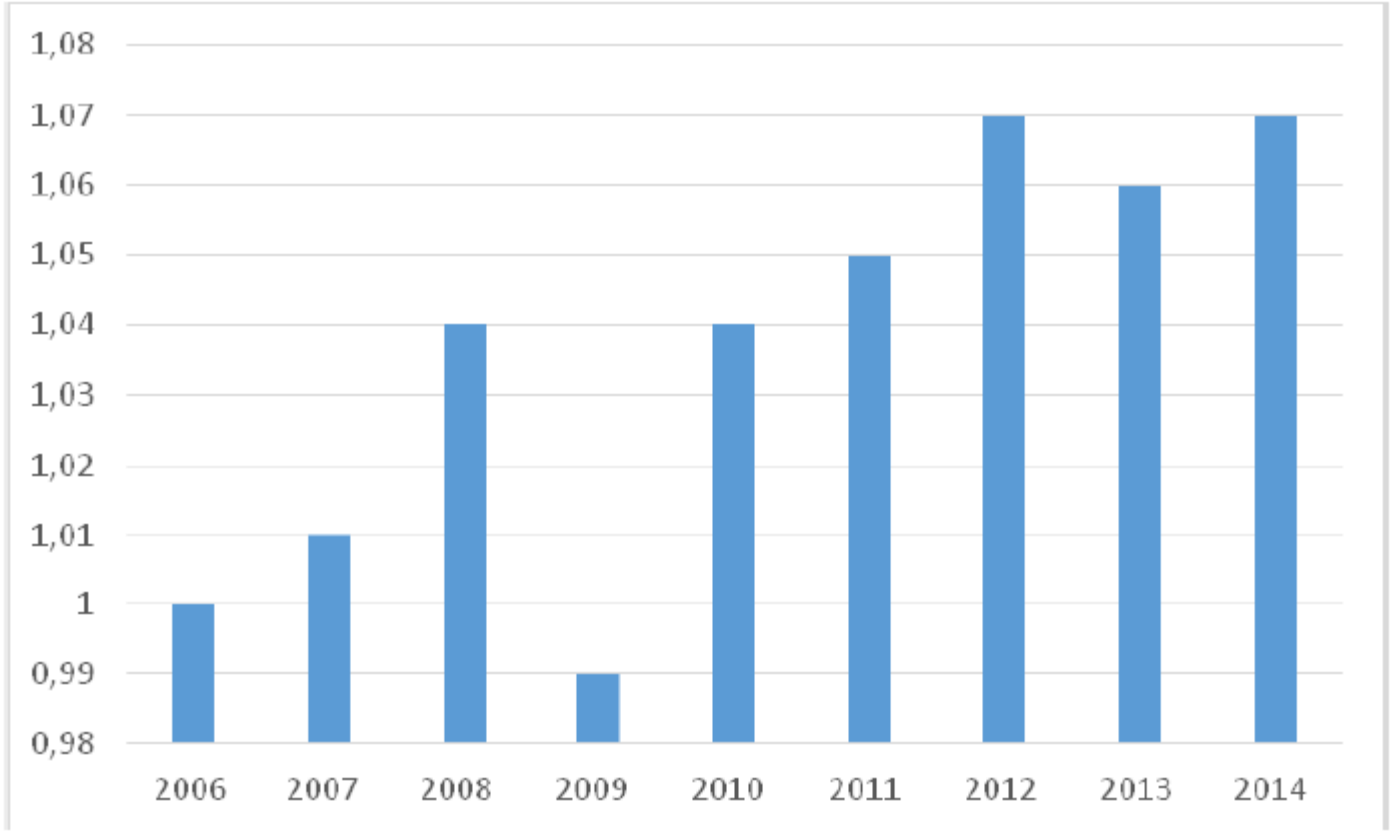
Problem 2. SN visitors monitoring



SELECTED PROBLEMS

The annual production of electricity is the total amount of electricity in Gigawatt-Hours (GW·h), produced at all power plants of a country for one year. The diagram shows the production of electricity in one large technologically advanced country for several years. The vertical axis indicates production in millions of gigawatt-hours. Data is rounded to the nearest hundredth.

Problem 3. Production of electricity



SELECTED PROBLEMS

a) By how many percent did the electricity production grow in 2012 comparing to 2009? Round your answer to the tenth of a percent.
b) If the production of electricity in some year grows more than 3% comparing to previous year, we say that this year shows ***a sharp increase*** in the electricity production. In which years did the electricity production in this country show a sharp increase?
в) Give a probable reason for a sharp decrease of the electricity production in the country in 2009.

Problem 3. Electricity production

SELECTED PROBLEMS

Problem 4. Spring floods on Volga

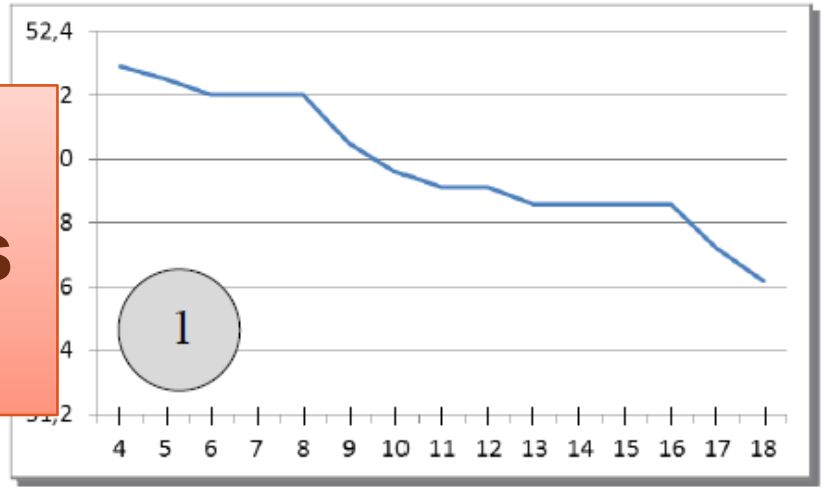
3. ***Volga*** is the main river of the European part of Russia. It starts in Tver Oblast and, ng a wide arc, flows to the Caspian ny large industrial centers are located a, as the river not only provides water necessary for all of us, but also works pping way. The scheme shows some ties located on Volga shores.

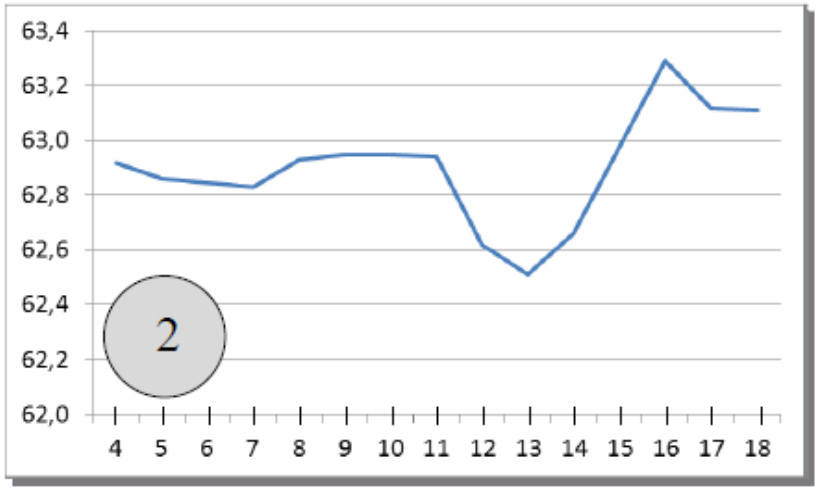
It is especially important to monitor a water level in spring. A water level in Russian rivers is measured in ***meters of Baltic system*** (mBS). The level of Baltic sea in Kronstadt is taken as the zero level.

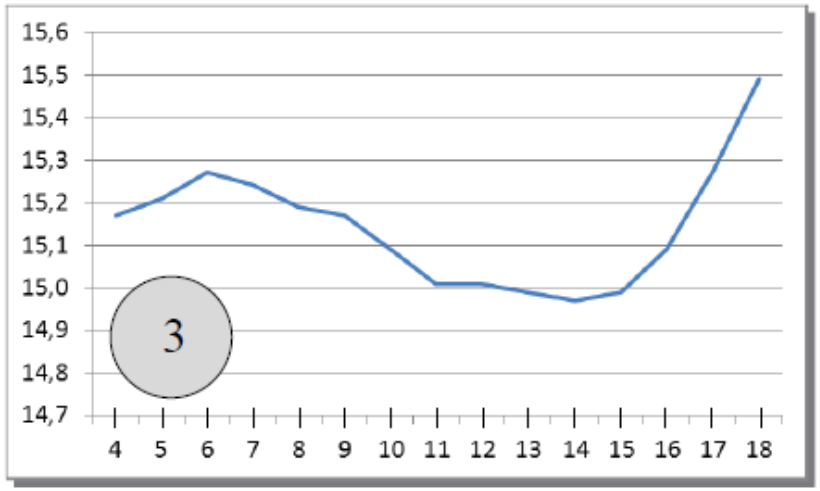
Four graphs below show levels of Volga water for four cities: **Nizhny Novgorod, Saratov, Cheboksary** and **Samara** during the period 4-18 April 2018. The horizontal axis shows days of April, while the vertical axis shows the water levels in mBS.

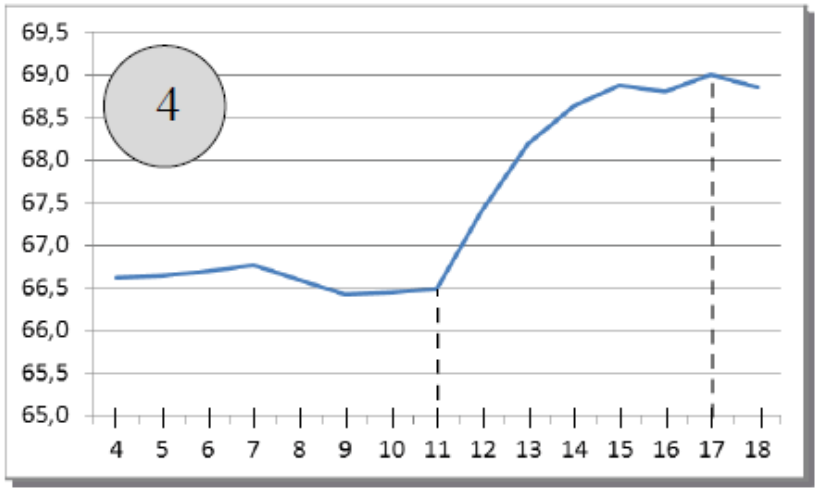
SELECTED PROBLEMS

Problem 4. Spring floods on Volga









SELECTED PROBLEMS

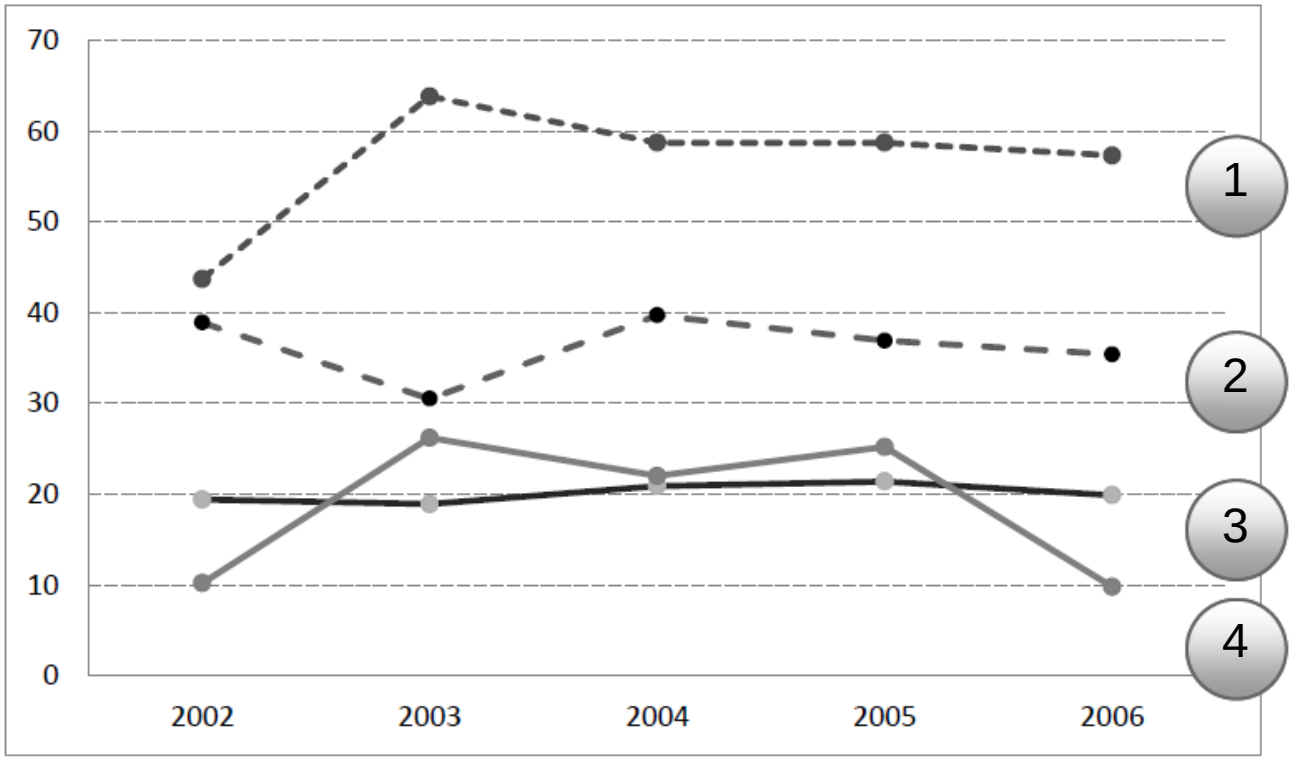
a) To which of the cities does the graph 3 correspond? Explain how you determined it.
b) The graph 4 shows a sharp increase of the water level starting from April 11. By how many meters per day in average did Volga water level near this city rise for 6 days from 11 to 17 April (round your result to the nearest hundredth)?
c) In your opinion, how the significant fluctuations of the Volga water level in spring can be explained? Provide two possible reasons.

Problem 4. Spring floods on Volga

SELECTED PROBLEMS

4. ***An annual production of wheat*** is the total mass of all varieties of wheat grown in a country during one year. It is usually measured in million tons. ***A yield of wheat*** (in centners per hectare) is the ratio of the mass of wheat in centners to the total area of cultivated areas in hectares. The diagram below shows the production of wheat in million tons in four countries: France, Turkey, the United States and Australia for five years, starting 2002. Review the diagram and read the accompanying article.

Problem 5. Annual production of wheat



SELECTED PROBLEMS

In 2002, in the USA, especially in the southern states, the heat and severe drought lasted whole summer, forest fires raged. Such weather conditions adversely affected the harvest of cereals, and in particular, of wheat. In the same year, incredibly dry weather happened in Australia; this was due to the return of the climatic phenomenon of El Niño. But the worst drought in Australia happened four years later, in 2006. This year, the production of agricultural cultures in this country decreased by 20%, and the production of wheat suffered more than others.

Since May 2003, the whole southwestern Europe has been experiencing abnormally hot weather; by mid-summer the temperature reached 40 degrees. In some c and July a drop of rain did not fall. The country suffered from severe fire of the fields. It was the hottest summer in France for the last 100 years. 2003 led to a drop in yield and a low harvest of cereals in France co subsequent years.

Problem 5. Annual production of wheat

SELECTED PROBLEMS

a) Based on the article, determine which country corresponds to each of the four lines in the diagram.
b) Name one or two other factors besides weather conditions, which can affect the production of wheat in a particular country.
c) On line 1, in 2003, there is a sharp increase in wheat production. A similar situation is on line 4 and on line 2 in 2004. How can you explain such production peaks after unsuccessful years?
d) We call a decrease of production ***significant***, if it makes more than 10% from the level of previous year. In what country in 2004 the decrease in wheat production was significant?

Problem 5. Annual production of wheat

CHANGEABILITY AND REGULARITY

- Since very beginning we are constantly noticing some regularities
- What are those regularities?
- Can we understand some of them?
- Can we estimate mean values having got only small amount of observation? May be not for sure but with an acceptable mistake?
- If we saw a regularity of a kind, could we say that some data observed conform to that regularity (or reject such possibility)?

- Statistical thinking and reasoning lead us to a new thought: we need a science to search and research regularities in random changeability and uncertainty of the world we live in. So, we need

PROBABILITIES THEORY

We need the probabilities theory as a mathematical theory modelling randomness and changeability using few mathematically defined (undefined) objects:
Events and probabilistic measure
Random values and their quantitative characteristics.

We goes to PT in grade 8th (m.b., very end of 7th) with intention to get back to statistics with newly developed probabilistic instruments and views.

LARGE NUMBERS LAW

The top of the whole stochastic part of school math is LNL.

Even if we don’t know its mathematical reasons we have to understand the main principles:

1. LNL allows us to measure unknown probabilities using frequencies.
2. Real values and means that we observe in experiments could differ. The more observations, the less the difference could be almost for sure
3. LNL gives us stability in life

CONCLUSION

The statistical thinking – is a culture (subculture) allowing successful planning and managing our life, business, even states and peoples. The statistical thinking is tightly connected to critical thinking.

It would be mistake to think that statistics operate big data. Professor Yuriy Tyurin said that if you have a big data you don’t need any science. Statistics is an art to derive reasonable conclusion on small samples.

CONCLUSION

The statistical thinking now couldn’t be torn from probabilistic thinking. Statistics without PT sounds like draftsmanship without geometry.

School statistics ought not to be mathematical statistics. What is more important is the concept of random changeability and readiness of students to accept it. Next point is that in the changeability some regularities are covered (or have covered), being described with some objective laws. That leads us to PT. Main law is the LNL.

CONCLUSION

In our present understanding, learning LNL and its operating way on simple instances, learning about its general nature and role in medicine, technique, social phenomena is the main objective of statistical and probabilistic course in school mathematics.

CONCLUSION

In our present understanding, learning LNL and its operating way on simple instances, learning about its general nature and role in medicine, technique, social phenomena is the main objective of statistical and probabilistic course in school mathematics.

THANK YOU