



Financováno
Evropskou unií
NextGenerationEU



VETUNI pro 21. století: Rozvoj VETUNI v oblasti digitalizace činností, profesionálního vzdělávání a flexibilních forem vzdělávání

Projekt NPO registrační číslo NPO_VETUNI_MSMT-16594/2022

Specifický cíl A3: Tvorba nových profesně zaměřených studijních programů

Studijní program:

Veterinární asistence

Návaznost na výstup:

☐ č. 1 ☐ č. 2 ☐ č. 3 ☒ č. 5 ☐ č. 6 ☐ č. 7 ☐ č. 8 ☐ č. 10

Předmět: (zkratka a název dle studijního plánu)

I. Bakalářská práce - Metodologie a biostatistika V2BP1

Typ výstupu: (konkretizovat, např. přednáška, návod na praktické cvičení aj.)

Studijní opory pro teoretickou výuku

Autor výstupu: Mgr. Linhart Petr, Ph.D.



Financováno
Evropskou unií
NextGenerationEU



Bakalářská práce I – Metodologie a biostatistika

Biostatistika



Biostatistika

- Popisuje biologické vlastnosti živých organismů na základě vybraných statistických znaků (stres-hladina kortizolu)
- Znaky zpravidla nabývají číselných hodnot
- Z velké části kvantifikují sledované vlastnosti

Statistické znaky

Kvalitativní (nominální):

- Jen dva stavy ANO NE
- Přítomnost nebo nepřítomnost znaku

Kvantitativní:

- Ordinální – vzestupné/sestupné, je možno je uspořádat (subjektivní)
- Kardinální – přesná číselná hodnota (objektivní – měřené přístrojem)
- Z formálního hlediska: Diskrétní x spojité

Statistický soubor

Základní soubor (ZS, populace):

- Neznámá velikost, $N=\infty$
- Všichni jedinci, u kterých se může sledovaný znak vyskytovat

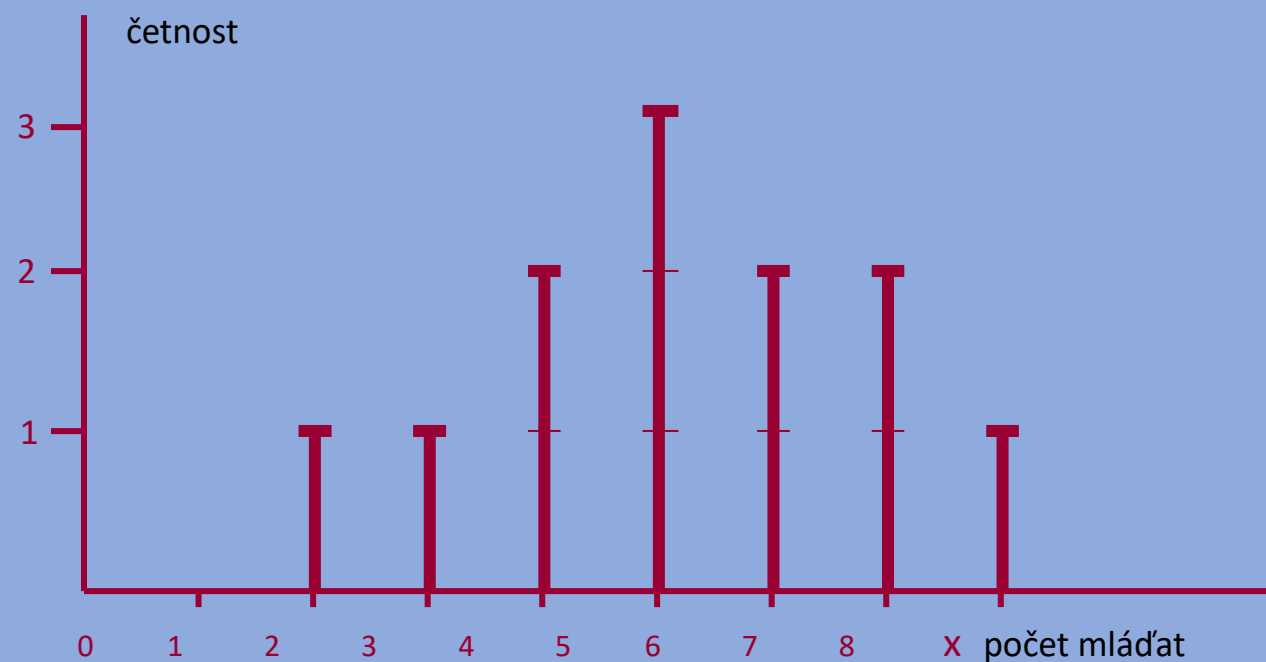
Výběrový soubor (VS, vzorek):

- Omezený a známý počet jedinců – $n=$
- Nepřesnost výpočtů, pomocí statistické analýzy se snažíme minimalizovat chyby

Náhodná veličina

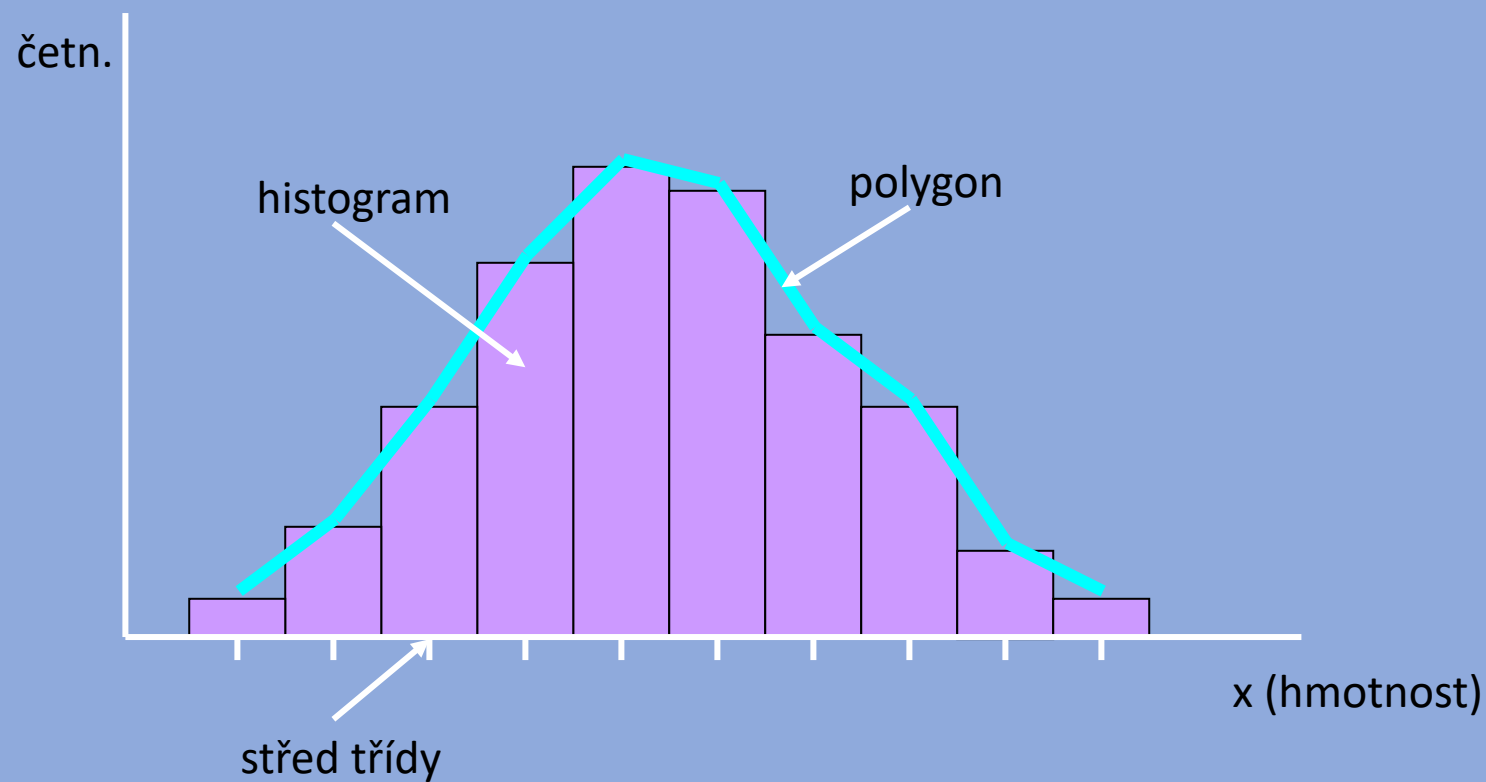
- Diskrétní, spojitá
- Variační řada – vzestupně nebo sestupně uspořádané hodnoty datového souboru např.: 2,3,4,4,5,5,5,6,6,7,7,8 (diskrétní veličina)
- Četnost varianty - počet opakování stejné hodnoty ve variační řadě
- Rozdělení četností náhodné veličiny - grafické vyjádření rozložení hodnot v datovém souboru

Rozdělení četností – diskrétní veličina:

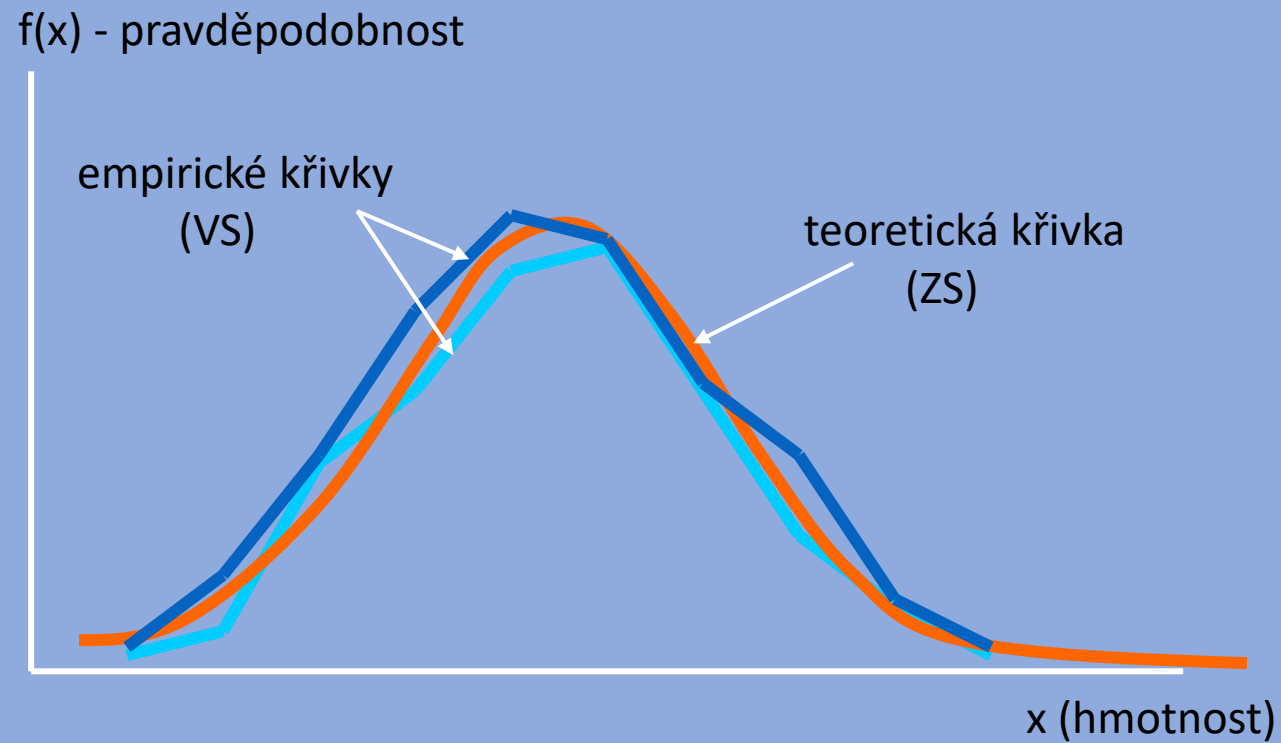


Diskrétní veličina – počet mláďat ve vrhu: 2,3,4,4,5,5,5,6,6,7,7,8

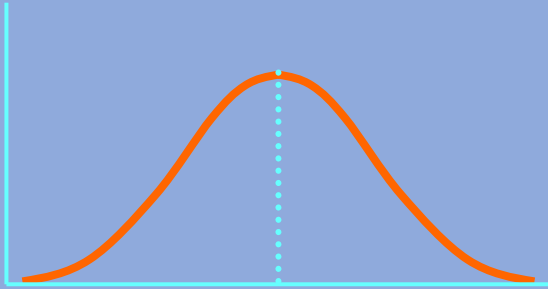
Rozdělení četností – spojitá veličina:



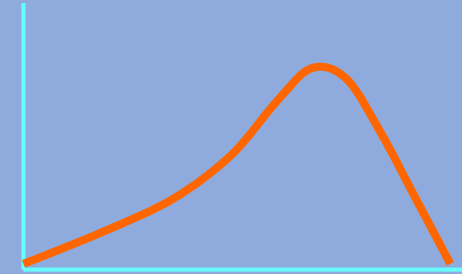
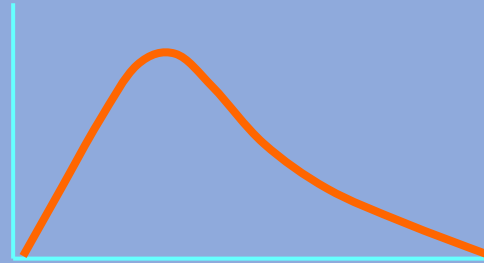
Rozdělení četností (pravděpodobností)



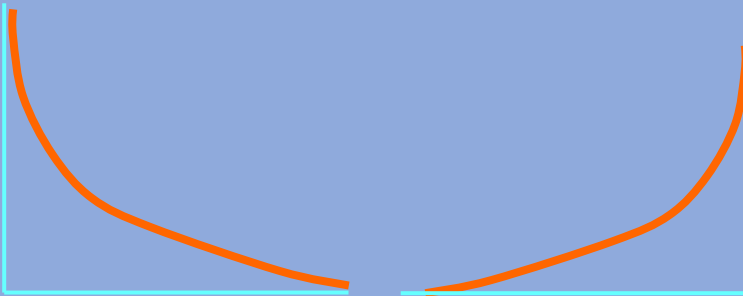
Typy teoretických křivek rozdělení



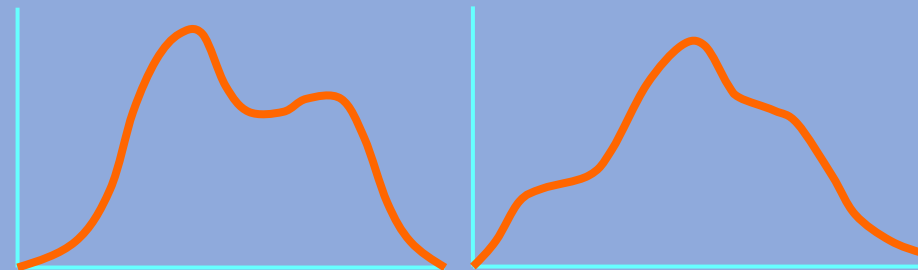
a) normální (Gaussovo)



b) asymetrické



c) extrémní

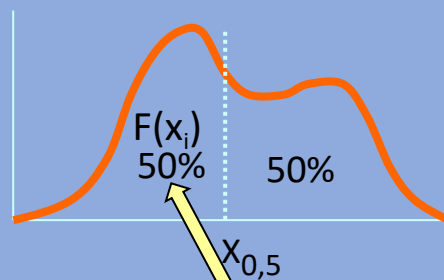
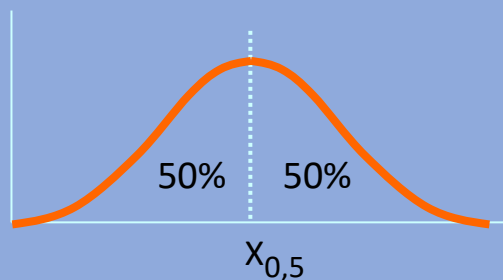


d) neznámé (nepravidelné)

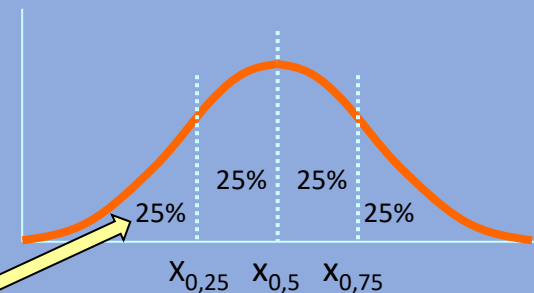
Kvantil

- hodnota kvantitativního znaku, která rozdělí soubor na 2 části:
 - jedinci s hodnotami nižšími než kvantil
 - jedinci s hodnotami vyššími než kvantil

50% kvantil – $x_{0,5}$ (medián) dělí soubor na poloviny:



Kvartily (1/4) :



Distribuční funkce – $F(x_i)$

-pravděpodobnost, že X nabude hodnot menších než je x_i

Využití : - kritické hodnoty při testování hypotéz
- koeficienty při výpočtech

Tabulka III

Distribuční funkce normovaného normálního rozdělení

u	$F(u)$	u	$F(u)$	u	$F(u)$	u	$F(u)$
1,40	0,919 24	1,85	0,967 84	2,30	0,989 28	3,00	0,998 65
1,41	0,920 73	1,86	0,968 56	2,31	0,989 56	3,02	0,998 74
1,42	0,922 20	1,87	0,969 26	2,32	0,989 83	3,04	0,998 82
1,43	0,923 64	1,88	0,969 95	2,33	0,990 10	3,06	0,998 89
1,44	0,925 07	1,89	0,970 62	2,34	0,990 36	3,08	0,998 97
1,45	0,926 47	1,90	0,971 28	2,35	0,990 61	3,10	0,999 03
1,46	0,927 86	1,91	0,971 93	2,36	0,990 86	3,12	0,999 16
1,47	0,929 22	1,92	0,972 57	2,37	0,991 11	3,14	0,999 16
1,48	0,930 56	1,93	0,973 20	2,38	0,991 34	3,16	0,999 21
1,49	0,931 89	1,94	0,973 81	2,39	0,991 58	3,18	0,999 26
1,50	0,933 19	1,95	0,974 41	2,40	0,991 80	3,20	0,999 31
1,51	0,934 48	1,96	0,975 00	2,41	0,992 02	3,22	0,999 36
1,52	0,935 74	1,97	0,975 58	2,42	0,992 24	3,24	0,999 40
1,53	0,936 99	1,98	0,976 15	2,43	0,992 45	3,26	0,999 44
1,54	0,938 22	1,99	0,976 70	2,44	0,992 66	3,28	0,999 48
1,55	0,939 43	2,00	0,977 25	2,45	0,992 86	3,30	0,999 52
1,56	0,940 62	2,01	0,977 78	2,46	0,993 05	3,32	0,999 55
1,57	0,941 79	2,02	0,978 31	2,47	0,993 05	3,34	0,999 58
1,58	0,942 95	2,03	0,978 82	2,48	0,993 43	3,36	0,999 61
1,59	0,944 08	2,04	0,979 32	2,49	0,993 48	3,38	0,999 64
1,60	0,945 20	2,05	0,979 82	2,50	0,993 79	3,40	0,999 66
1,61	0,946 30	2,06	0,980 30	2,52	0,994 13	3,42	0,999 69
1,62	0,947 38	2,07	0,980 77	2,54	0,994 46	3,44	0,999 71
1,63	0,948 45	2,08	0,981 24	2,56	0,994 77	3,46	0,999 73
1,64	0,949 50	2,09	0,981 69	2,58	0,995 06	3,48	0,999 75
1,65	0,950 53	2,10	0,982 14	2,60	0,995 34	3,50	0,999 77
1,66	0,951 54	2,11	0,982 57	2,62	0,995 60	3,55	0,999 81
1,67	0,952 54	2,12	0,983 00	2,64	0,995 85	3,60	0,999 84
1,68	0,953 52	2,13	0,983 41	2,66	0,996 09	3,65	0,999 87
1,69	0,954 49	2,14	0,983 82	2,68	0,996 32	3,70	0,999 89

Popisné charakteristiky statistických souborů

- **Střední hodnoty**
- **Míra variability**

Cíle statistiky

- Na základě dat vzorku získat informace o vlastnostech celé populace
- **ZS** – přesné parametry a informace (nelze je zjistit pomocí měření – nekonečná velikost populace)
- **VS** – výběrové charakteristiky (slouží jako odhad skutečných parametrů celé populace, není tak přesný)

Střední hodnoty

Aritmetický průměr

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

Vlastnosti:

- je ovlivněn extrémními hodnotami !
- lze ho použít jen u souborů s normálním rozdělením

Střední hodnoty

Medián

$$Poř.č._{\tilde{x}} = \frac{n+1}{2}$$

Vlastnosti:

- není ovlivněn extrémními hodnotami

Střední hodnoty

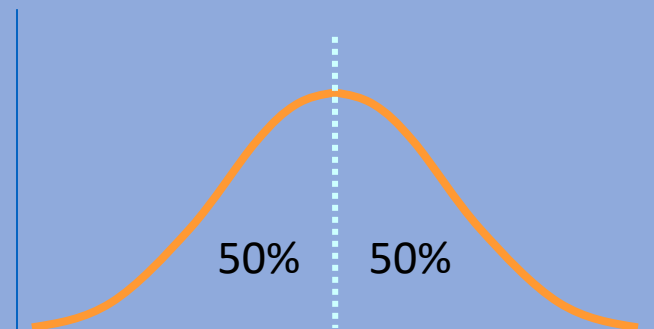
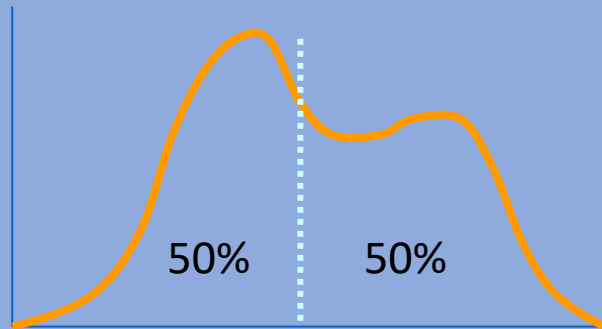
Medián

$$Poř.č._{\tilde{x}} = \frac{n+1}{2}$$

prostřední hodnota variační řady (průměr 2 prostředních)

Vlastnosti:

- není ovlivněn extrémními hodnotami



Příklad: Tělesná hmotnost u 2 variet laboratorních myší:

Varieta A
 x_i (g)

34

36

37

39

40

41

42

43

79

 $n = 9$

medián: $x_5 = 40 \text{ g}$

$\bar{x} = 43.4 \text{ g}$

Varieta B

x_i (g)

34

36

37

39

40

41

42

43

44

45

 $n = 10$

medián: $x_{5.5} = \frac{40 + 41}{2} = 40.5 \text{ g}$

$\bar{x} = 40.1 \text{ g}$

Střední hodnoty

Modus

Je nejčastější hodnota variační řady

Vlastnosti:

- není ovlivněn extrémními hodnotami

Míry variability

Rozptyl

Průměr čtverců odchylek jednotlivých hodnot x_i od aritmetického průměru souboru.

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

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

Míry variability

Směrodatná odchylka (SD-Standard Deviation)

Vypočítáme ji jako odmocninu z rozptylu

Má stejný rozměr jako měřená veličina

Míry variability

Variační koeficient

$$V = \frac{s}{\bar{x}} \cdot 100 \quad [\%]$$

Relativní směrodatná odchylka

Není závislý na absolutních hodnotách znaku

Míry variability

Střední chyby průměru SEM (Standard Error of Mean)

$$SEM = \frac{s}{\sqrt{n}}$$

Vyjadřuje přesnost výběrového průměru

Kolísání výběrových průměrů kolem skutečné střední hodnoty populace



Financováno
Evropskou unií
NextGenerationEU

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Národní
plán
obnovy

Bakalářská práce I – Metodologie a biostatistika

Typy rozdělení náhodné veličiny



Rozdělení pro ZS (populace):

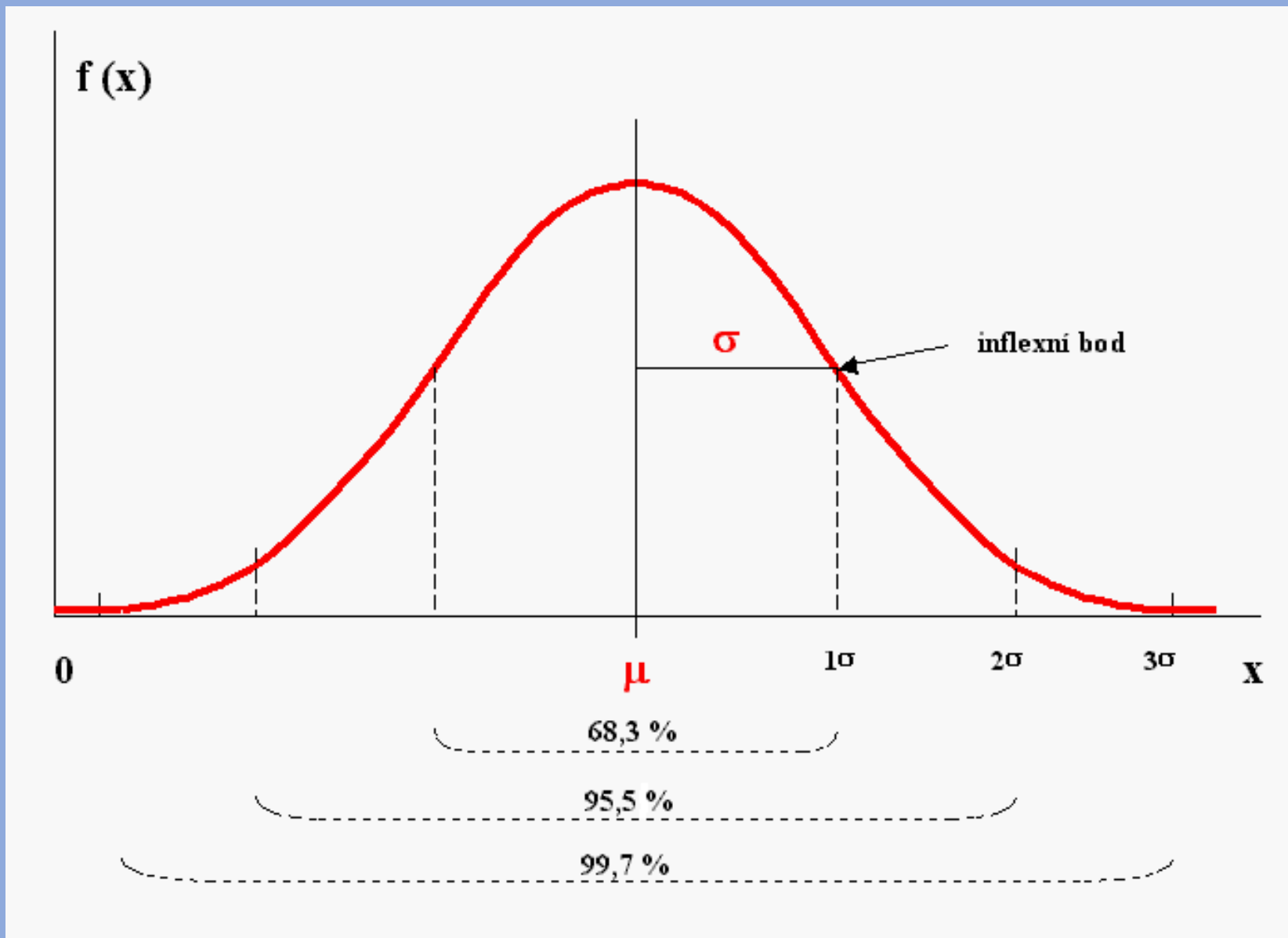
- Gaussovo normální rozdělení
- Normované normální rozdělení
- Neznáme rozdělení

Výběrová rozdělení (VS, vzorek):

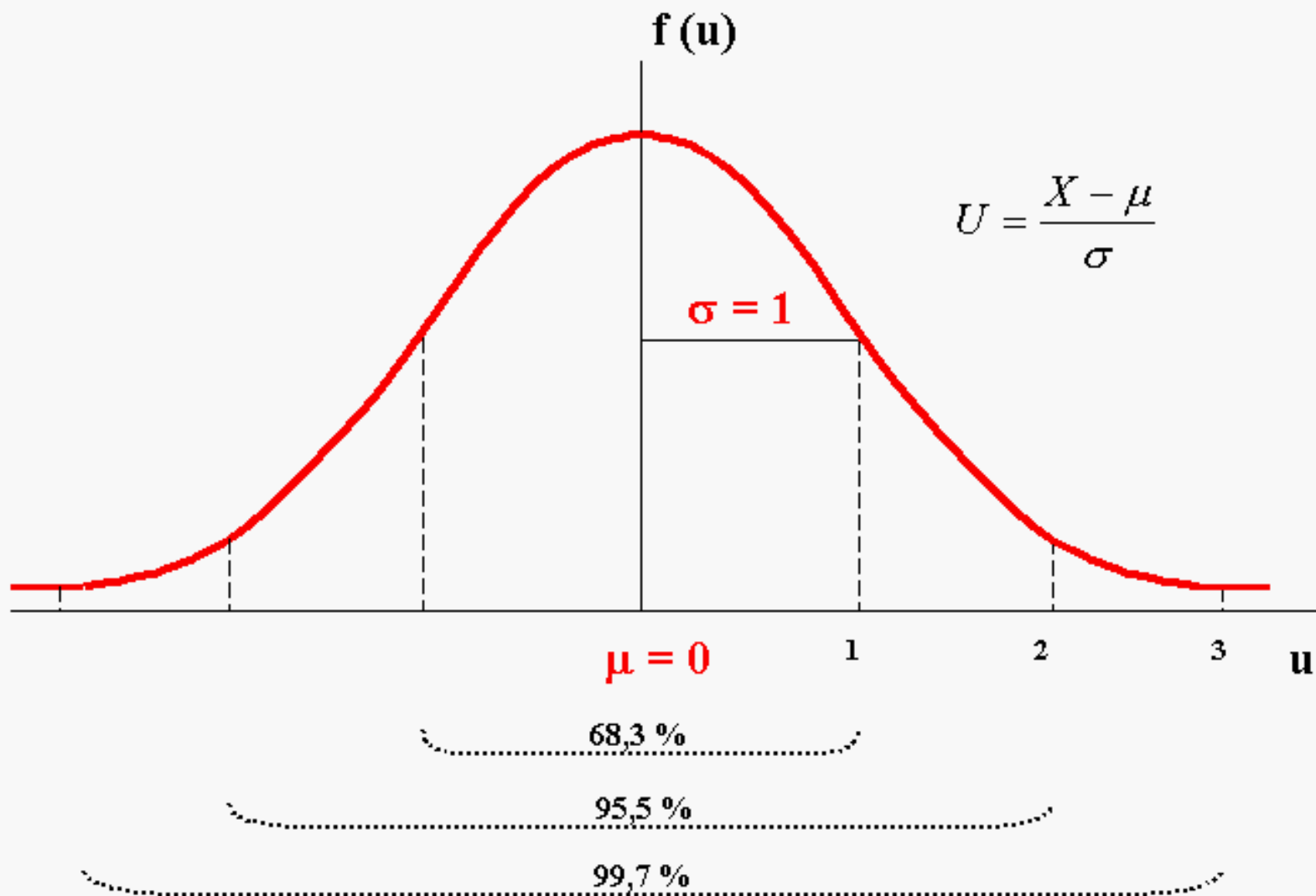
- t - rozdělení
- χ^2 - rozdělení
- F - rozdělení

Rozdělení pro ZS

Gaussovo normální rozdělení



Normované normální rozdělení

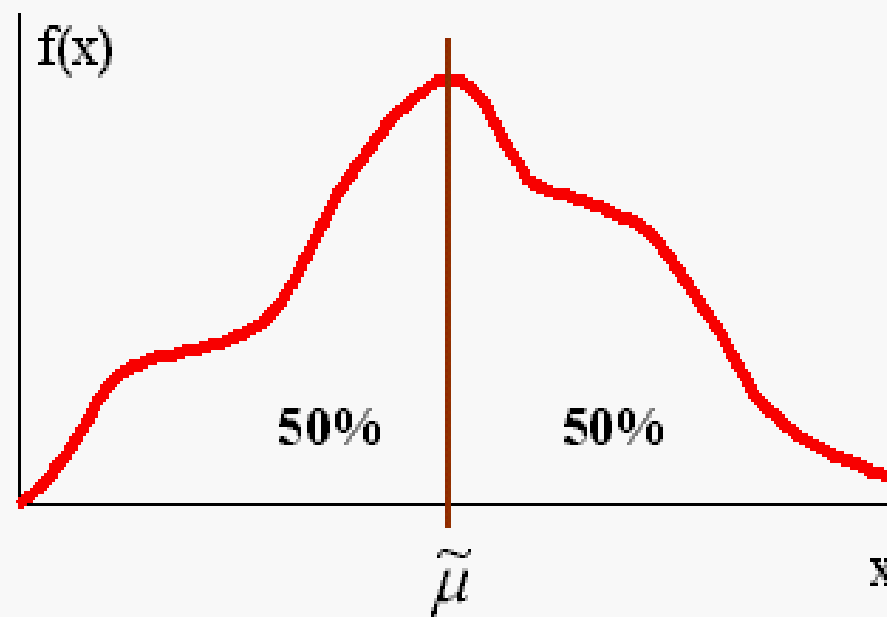
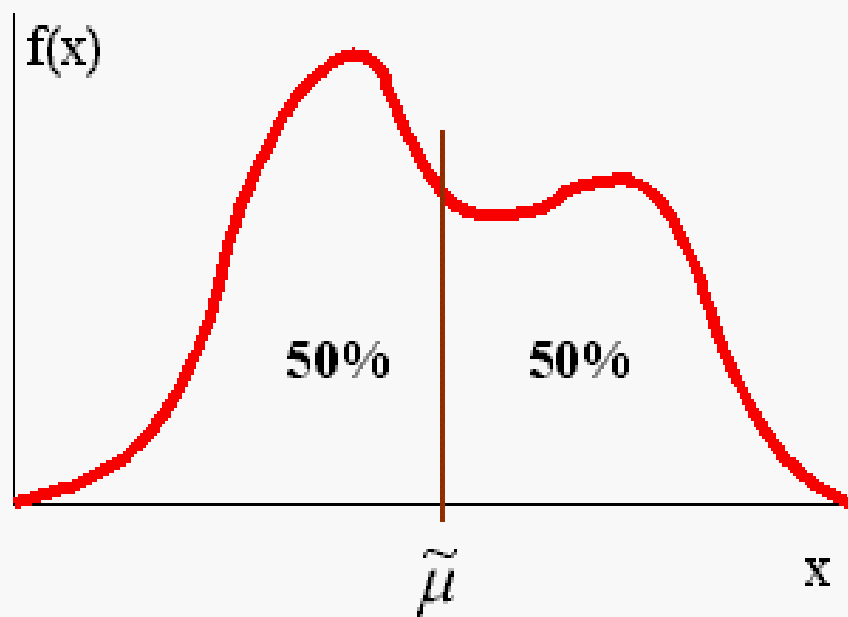


Tabulka III

Distribuční funkce normovaného normálního rozdělení

u	$F(u)$	u	$F(u)$	u	$F(u)$	u	$F(u)$
1,40	0,919 24	1,85	0,967 84	2,30	0,989 28	3,00	0,998 65
1,41	0,920 73	1,86	0,968 56	2,31	0,989 56	3,02	0,998 74
1,42	0,922 20	1,87	0,969 26	2,32	0,989 83	3,04	0,998 82
1,43	0,923 64	1,88	0,969 95	2,33	0,990 10	3,06	0,998 89
1,44	0,925 07	1,89	0,970 62	2,34	0,990 36	3,08	0,998 97
1,45	0,926 47	1,90	0,971 28	2,35	0,990 61	3,10	0,999 03
1,46	0,927 86	1,91	0,971 93	2,36	0,990 86	3,12	0,999 16
1,47	0,929 22	1,92	0,972 57	2,37	0,991 11	3,14	0,999 16
1,48	0,930 56	1,93	0,973 20	2,38	0,991 34	3,16	0,999 21
1,49	0,931 89	1,94	0,973 81	2,39	0,991 58	3,18	0,999 26
1,50	0,933 19	1,95	0,974 41	2,40	0,991 80	3,20	0,999 31
1,51	0,934 48	1,96	0,975 00	2,41	0,992 02	3,22	0,999 36
1,52	0,935 74	1,97	0,975 58	2,42	0,992 24	3,24	0,999 40
1,53	0,936 99	1,98	0,976 15	2,43	0,992 45	3,26	0,999 44
1,54	0,938 22	1,99	0,976 70	2,44	0,992 66	3,28	0,999 48
1,55	0,939 43	2,00	0,977 25	2,45	0,992 86	3,30	0,999 52
1,56	0,940 62	2,01	0,977 78	2,46	0,993 05	3,32	0,999 55
1,57	0,941 79	2,02	0,978 31	2,47	0,993 05	3,34	0,999 58
1,58	0,942 95	2,03	0,978 82	2,48	0,993 43	3,36	0,999 61
1,59	0,944 08	2,04	0,979 32	2,49	0,993 48	3,38	0,999 64
1,60	0,945 20	2,05	0,979 82	2,50	0,993 79	3,40	0,999 66
1,61	0,946 30	2,06	0,980 30	2,52	0,994 13	3,42	0,999 69
1,62	0,947 38	2,07	0,980 77	2,54	0,994 46	3,44	0,999 71
1,63	0,948 45	2,08	0,981 24	2,56	0,994 77	3,46	0,999 73
1,64	0,949 50	2,09	0,981 69	2,58	0,995 06	3,48	0,999 75
1,65	0,950 53	2,10	0,982 14	2,60	0,995 34	3,50	0,999 77
1,66	0,951 54	2,11	0,982 57	2,62	0,995 60	3,55	0,999 81
1,67	0,952 54	2,12	0,983 00	2,64	0,995 85	3,60	0,999 84
1,68	0,953 52	2,13	0,983 41	2,66	0,996 09	3,65	0,999 87
1,69	0,954 49	2,14	0,983 82	2,68	0,996 32	3,70	0,999 89
1,70	0,955 43	2,15	0,984 22	2,70	0,996 53	3,75	0,999 91
1,71	0,956 37	2,16	0,984 61	2,72	0,996 74	3,80	0,999 93
1,72	0,957 28	2,17	0,985 00	2,74	0,996 93	3,85	0,999 94
1,73	0,958 18	2,18	0,985 37	2,76	0,997 11	3,90	0,999 95
1,74	0,959 07	2,19	0,985 74	2,78	0,997 28	3,95	0,999 96
1,75	0,959 94	2,20	0,986 10	2,80	0,997 44	4,00	0,999 97
1,76	0,960 80	2,21	0,986 45	2,82	0,997 60	4,05	0,999 97
1,77	0,961 64	2,22	0,986 79	2,84	0,997 74	4,10	0,999 98
1,78	0,962 46	2,23	0,987 13	2,86	0,997 88	4,15	0,999 98
1,79	0,963 27	2,24	0,987 45	2,88	0,998 01	4,20	0,999 99
1,80	0,964 07	2,25	0,987 78	2,90	0,998 13	4,25	0,999 99
1,81	0,964 85	2,26	0,988 09	2,92	0,998 25	4,30	0,999 99
1,82	0,965 62	2,27	0,988 40	2,94	0,998 36	4,35	0,999 99
1,83	0,966 38	2,28	0,988 70	2,96	0,998 46	4,40	0,999 99
1,84	0,967 12	2,29	0,988 99	2,98	0,998 56	4,45	1,000 00

Neznámé rozdělení



Výběrová rozdělení

t- rozdělení (Studentovo)

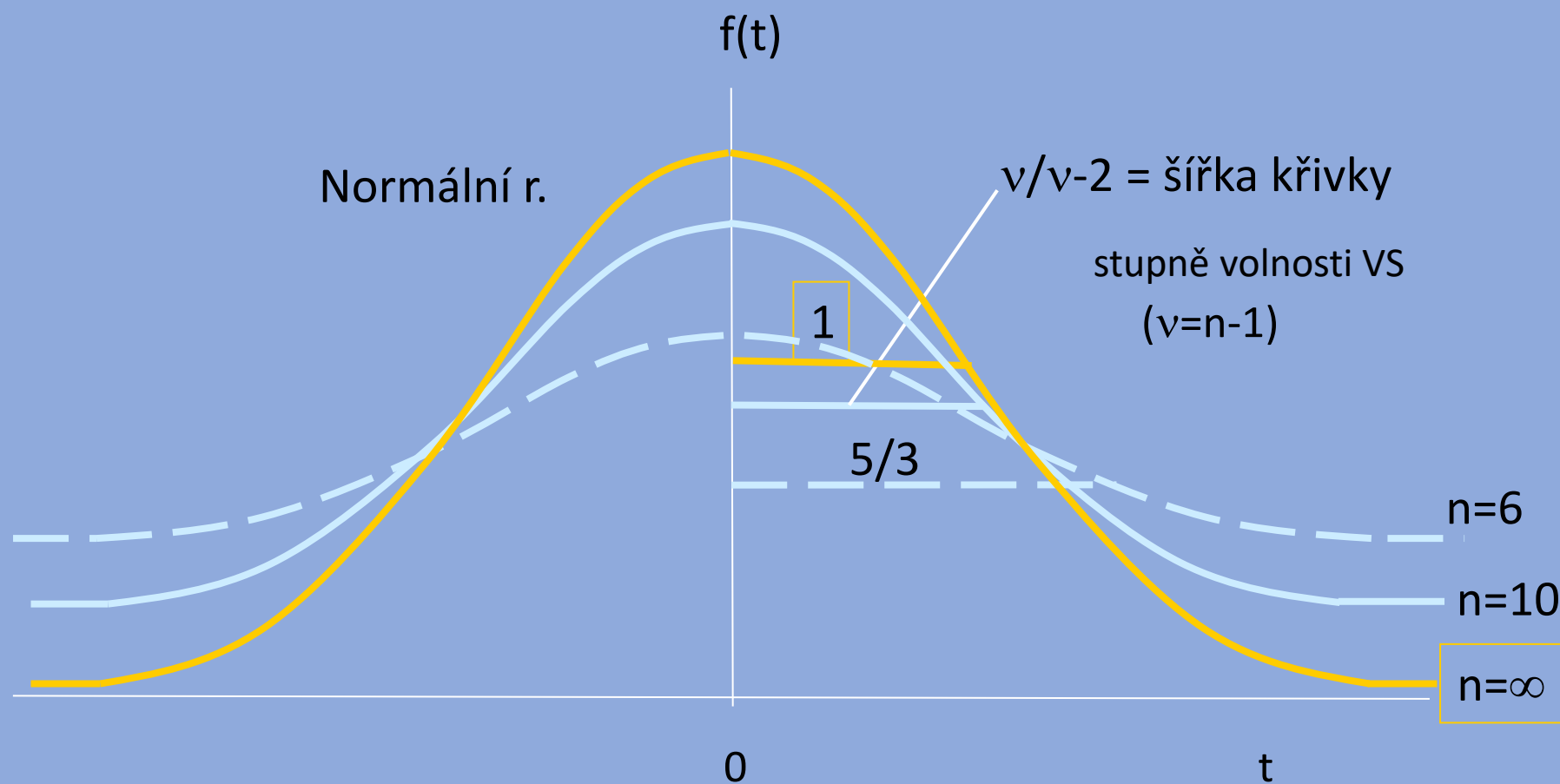


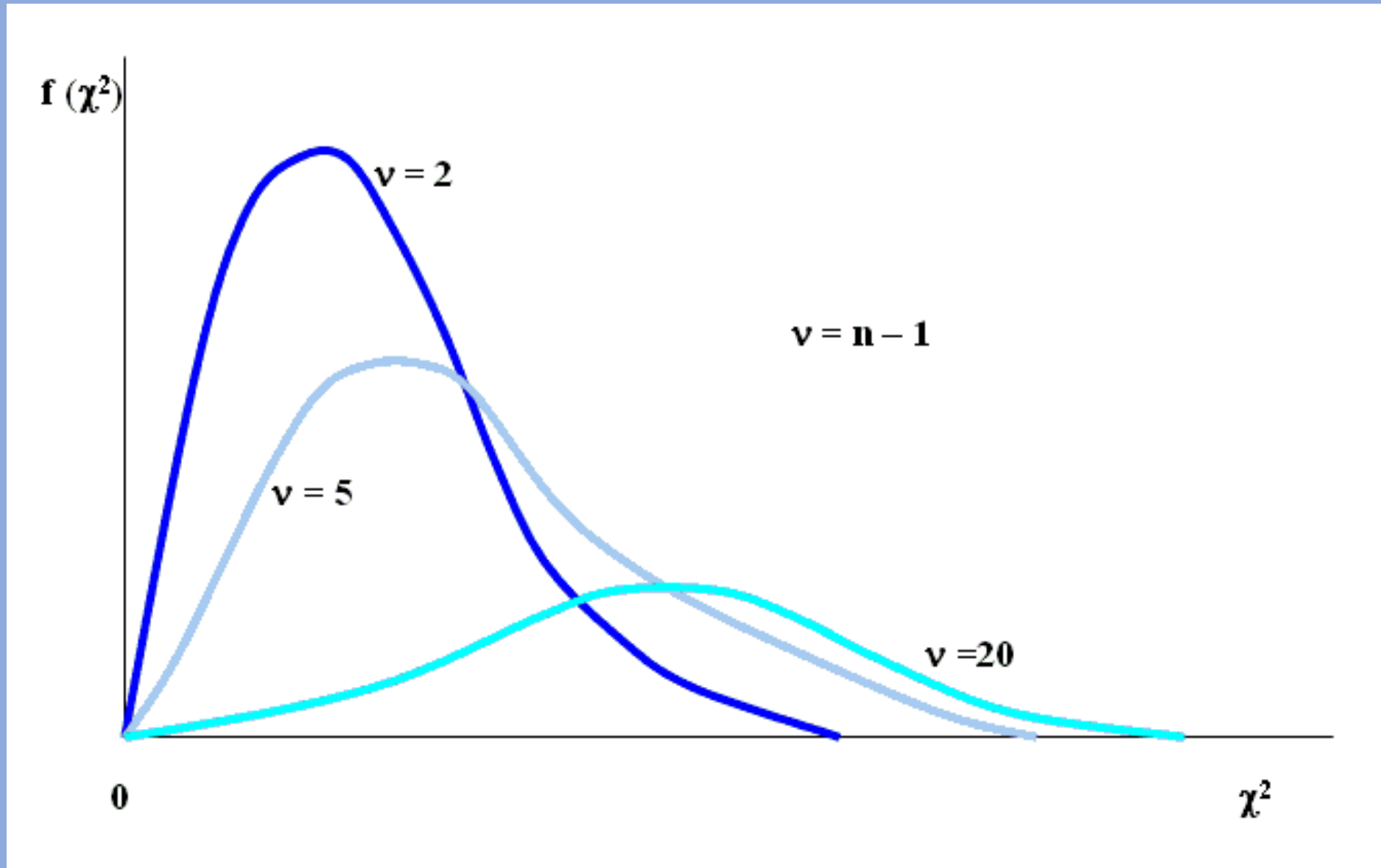
TABLE B.3 Critical Values of the *t* Distribution

α	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
ν									
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.693	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140

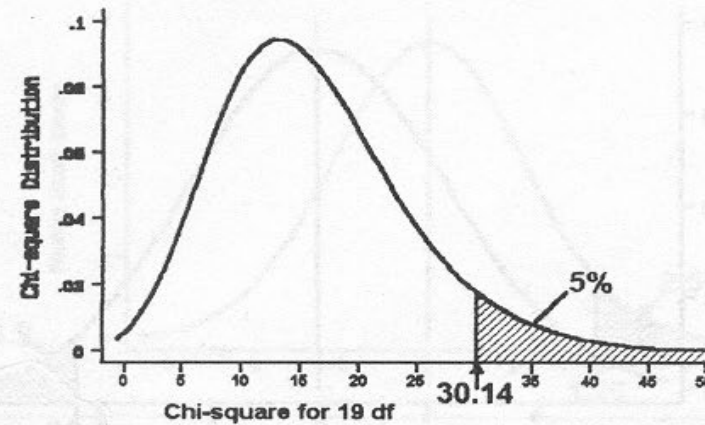
Využití:

- kritické hodnoty pro testování rozdílu mezi 2 průměry (*t*-test)
- koeficienty při výpočtu intervalů spolehlivosti průměru

χ^2 -rozdělení (Pearsonovo)



Chi-Square Distribution, Right Tail^a



α (area in right tail)	.10	.05	.025	.01	.005	.001	.0005
$1-\alpha$ (except right tail)	.90	.95	.975	.99	.995	.999	.9995
df= 1	2.71	3.84	5.02	6.63	7.88	10.81	12.13
2	4.61	5.99	7.38	9.21	10.60	13.80	15.21
3	6.25	7.81	9.35	11.34	12.84	16.26	17.75
4	7.78	9.49	11.14	13.28	14.86	18.46	20.04
5	9.24	11.07	12.83	15.08	16.75	20.52	22.15
6	10.64	12.59	14.45	16.81	18.54	22.46	24.08
7	12.02	14.07	16.01	18.47	20.28	24.35	26.02
8	13.36	15.51	17.53	20.09	21.95	26.10	27.86
9	14.68	16.92	19.02	21.67	23.59	27.88	29.71

Využití:

- kritické hodnoty pro testování rozdílu mezi četnostmi

17	24.77	27.59	30.19	33.41	35.72	40.78	42.89
18	25.99	28.87	31.53	34.80	37.16	42.32	44.47

F-rozdělení (Fischer-Snedecorovo)

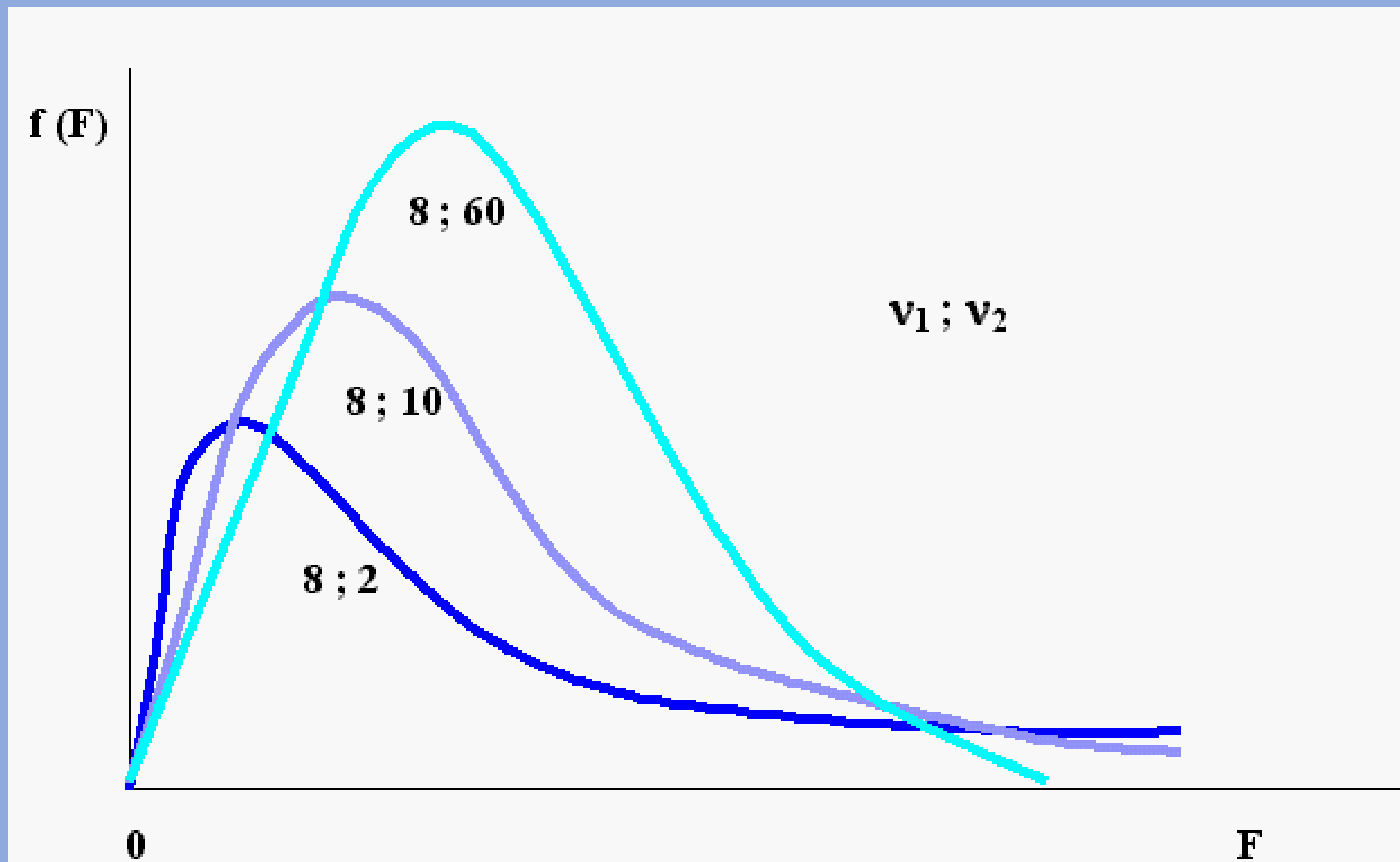


TABLE B.4 (cont.) Critical Values of the F Distribution

Numerator DF = 9

Denom. DF	$\alpha(2): 0.50$ $\alpha(1): 0.25$	0.20 0.10	0.10 0.05	0.05 0.025	0.02 0.01	0.01 0.005	0.005 0.0025	0.002 0.001	0.001 0.0005
1	9.26	59.9	241.	963.	6020.	24100.	96400.	602000.	2410000.
2	3.37	9.38	19.4	39.4	99.4	199.	399.	999.	2000.
3	2.44	5.24	8.81	14.5	27.3	43.9	70.1	130.	207.
4	2.08	3.94	6.00	8.90	14.7	21.1	30.3	48.5	69.0
5	1.89	3.32	4.77	6.68	10.2	13.8	18.5	27.2	36.3
6	1.77	2.96	4.10	5.52	7.98	10.4	13.4	18.7	23.9
7	1.69	2.72	3.68	4.82	6.72	8.51	10.7	14.3	17.8
8	1.63	2.56	3.39	4.36	5.91	7.34	9.03	11.8	14.3
9	1.59	2.44	3.18	4.03	5.35	6.54	7.92	10.1	12.1
10	1.56	2.35	3.02	3.78	4.94	5.97	7.14	8.96	10.6
11	1.53	2.27	2.90	3.59	4.63	5.54	6.56	8.12	9.48
12	1.51	2.21	2.80	3.44	4.39	5.20	6.11	7.48	8.66
13	1.49	2.16	2.71	3.31	4.19	4.94	5.76	6.98	8.03
14	1.47	2.12	2.65	3.21	4.03	4.72	5.47	6.58	7.52
15	1.46	2.09	2.59	3.12	3.89	4.54	5.23	6.26	7.11
16	1.44	2.06	2.54	3.05	3.78	4.38	5.04	5.98	6.77

Využití:

- kritické hodnoty pro testování rozdílu mezi 2 rozptyly (F -test)

27	1.37	1.87	2.25	2.63	3.15	3.56	3.98	4.57	5.04
28	1.37	1.87	2.24	2.61	3.12	3.52	3.93	4.50	4.96
29	1.36	1.86	2.22	2.59	3.09	3.48	3.89	4.45	4.89
30	1.36	1.85	2.21	2.57	3.07	3.45	3.85	4.39	4.82

Testování hypotéz

- Využití: vyhodnocování experimentů
- VS – platnost hypotézy o celé populaci (ZS)

Statistická hypotéza

- Tvrzení o sledovaném znaku

Nulová - H_0 ($\mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$...)

Alternativní - H_1 (popírá platnost H_0)

- Vyhodnocení experimentů: pokus x kontrola
- Při výpočtu testu si zvolíme standardní **chybu $\alpha = 0,05$ (0,01)**

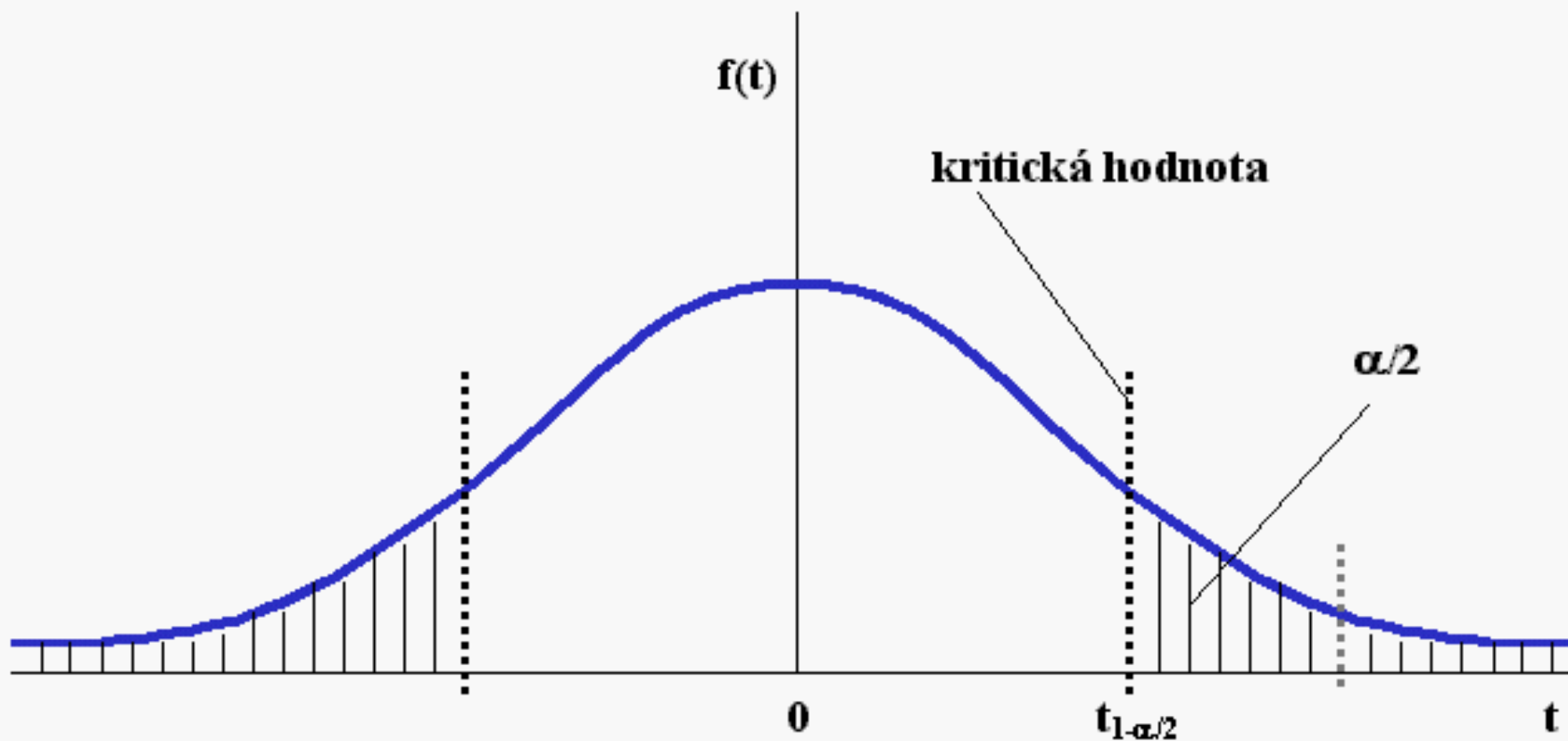
Testovací kritérium

- Rozhodnutí o platnosti H_0 provádíme na základě výpočtu testovacího kritéria

např.

- t – testování rozdílu 2 průměrů (t-test)
- F – testování rozdílu 2 rozptylů (F-test)
- χ^2 – testování rozdílu četností (χ^2 - test)

Testovací kritérium - t



obor přijetí H_0 **kritický obor**
(zamítáme H_0)

Kritická hodnota = kvantil $t_{1-\alpha/2}$ Studentova t -rozdělení (tabulky)

TABLE B.3 Critical Values of the *t* Distribution

α	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
ν									
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.702	1.382	1.823	2.262	2.821	3.250	3.690	4.297	4.781

Tab. č. 3 Kvantily $t_{1-\alpha/2}(\nu)$ Studentova *t* rozdělení

St. volnosti ν	0,80	0,90	0,95	0,975	0,9875	0,995
1	1,376	3,078	6,314	12,706	25,452	63,657
2	1,061	1,886	2,920	4,303	6,205	9,925
3	0,978	1,638	2,353	3,182	4,176	5,841

$\alpha = 0,05$ - hladina významnosti testu
(0,01)

9	,883	1,383	1,833	2,262	2,685	3,250
10	,879	1,372	1,812	2,228	2,634	3,169
11	,876	1,363	1,796	2,201	2,593	3,106
12	,873	1,356	1,782	2,179	2,560	3,055
13	,870	1,350	1,771	2,160	2,533	3,012

Parametrické testy

(použití u souborů s normálním rozdělením)

Hypotéza se týká průměru a SD

F - test

- testování rozdílu 2 rozptylů – $H_0: \sigma_1^2 = \sigma_2^2$

Testovací kritérium

$$F = \frac{\textit{větší}(s_1^2, s_2^2)}{\textit{menší}(s_1^2, s_2^2)}$$

- Je-li vypočítané $F > F_{\text{krit.}} \Rightarrow \sigma_1^2 \neq \sigma_2^2$ (statisticky významný rozdíl mezi rozptyly)
- Je-li vypočítané $F \leq F_{\text{krit.}} \Rightarrow \sigma_1^2 = \sigma_2^2$ (statisticky nevýznamný rozdíl mezi rozptyly)

Příklad: Zjistěte, zda se liší variabilita počtu vajec ve snášce domácích a divokých kachen.

Velikost snášky (počet vajec):

Domácí Divoká

10	9
11	8
12	11
11	12
10	10
11	13
11	11
	10
	12

$v = 6$

$n_1=7$

$n_2=9$

$v = 8$

$v = n-1$

$s_1^2= 0.48$

$s_2^2= 2.50$

$$F = \frac{2.50}{0.48} = 5.21$$

Tab. č. 7 Kvantily $F_{0,975}(v_V, v_M)$ Fisher-Snedecorova rozdělení

($\alpha = 0,05$)

vV \ vM	Počet stupňů volnosti čitatele								
	1	2	3	4	5	6	7	8	9
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28
2	38,506	39,000	39,165	39,248	39,298	39,331	39,355	39,373	39,387
3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,473
4	12,218	10,649	9,979	9,605	9,365	9,197	9,074	8,980	8,905
5	10,007	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681
6	8,813	7,260	6,599	6,227	5,988	5,820	5,696	5,600	5,523
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823
8	7,571	6,060	5,416	5,053	4,817	4,652	4,529	4,433	4,357
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049
17	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880
20	5,872	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837
21	5,827	4,420	3,819	3,475	3,250	3,090	2,969	2,874	2,798
22	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763
23	5,750	4,349	3,751	3,408	3,184	3,023	2,902	2,808	2,731
24	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703
25	5,686	4,291	3,694	3,353	3,129	2,969	2,848	2,753	2,677

Příklad: Zjistěte, zda se liší variabilita počtu vajec ve snášce domácích a divokých kachen.

Velikost snášky (počet vajec):

Domácí Divoká

10	9
11	8
12	11
11	12
10	10
11	13
11	11
	10
	12

$$F = \frac{2.50}{0.48} = 5.21$$

$v = 6$

$n_1=7$

$n_2=9$

$v = 8$

$v = n-1$

$$s_1^2 = 0.48$$

$$s_2^2 = 2.50$$

$$F_{\text{krit.}(8,6)} = 5.6$$

Závěr:

$F < F_{\text{krit.}} \Rightarrow H_0: \sigma_1^2 = \sigma_2^2 \Rightarrow$ Variabilita počtu vajec ve snášce domácích a divokých kachen se statisticky významně neliší.



Financováno
Evropskou unií
NextGenerationEU

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Národní
plán
obnovy

Bakalářská práce I – Metodologie a biostatistika

Statistická analýza biologických dat -souhrn statistických metod



Biostatistika

- souhrn vědeckých metod sběru, třídění, zpracování a analyzování dat včetně vyvozování závěrů (získání informace o celé populaci-základní soubor)
- aplikace statistických metod pro řešení biologické a klinické problematiky
- použití statistických metod pro získávání informací je nezbytné a není možné se bez nich obejít



Získání dat (pozorování, experiment...)



Data, zápis, kontrola



Statistické vyhodnocení



Vyhodnocení výsledků, vytvoření závěrů

Metodika sběru a analýzy dat

- opomíjená součást v rámci klinických i experimentálních studií – dobře naplánovaná metodika získání dat a jejich následné analýzy
- je třeba vědět, jaká data je možné v rámci studie získat a jak budou následně statisticky vyhodnocena ještě před zahájením studie
- data sbíráme za účelem ověření hypotézy – náhodný sběr náhodných čísel je noční můrou pro ty, kdo data následně zpracovávají

Příprava dat, MS Excel

- Správný a přehledný zápis dat je základem jejich úspěšné pozdější analýzy
- Nejvhodnějším způsobem je zápis dat ve formě tabulky, kde ve sloupcích jsou proměnné (např. věk, hmotnost, výška, tepová frekvence...) a v řádcích pozorování (např. jednotlivá zvířata nebo jednotlivá vyšetření)

Příprava dat, MS Excel

- Každý sloupec obsahuje pouze jeden typ dat, popsáný hlavičkou sloupce
- Neměly by se kombinovat v jednom sloupci číselné a textové hodnoty
- Popisy jsou uloženy v samostatných sloupcích
- U textových dat je nezbytné kontrolovat překlepy v názvech kategorií
- Specifickým typem dat je datum, kde je nutné kontrolovat, zda jsou uložena ve správném formátu.

Příprava dat, MS Excel

[illegible]

Statistická analýza

- **Popisná** – výpočet základních statistických parametrů souboru (průměr, medián, min., max., SD, variační koeficient, SEM), četnosti výskytu, % apod., grafická prezentace
- **Indukční** – testování hypotéz, vyvozování závěrů včetně zhodnocení jejich spolehlivosti
 - a) Srovnávání skupin – porovnávání souborů (párový, nepárový pokus, jednovýběrový, vícevýběrový)
 - b) Korelace – hodnocení a kvantifikace vztahu mezi statistickými znaky

Typ dat $\Rightarrow$ volba testu

1) Kvantitativní data

- **Spojitá:** nabývají jakýchkoliv hodnot v určitém rozmezí – reálná čísla (hmotnost, výška, biochemické a hematologické parametry...)
- **Diskrétní:** hodnoty jsou celá čísla – typicky počty (počet králíků v králíkárně, počet snesených vajec...)

Před vlastním porovnáním výběrových souborů je nutné ověření normality dat (Unistat): normální rozdělení x nenormální rozdělení

Normální



Parametrické testy

(H_0 : shoda μ, σ^2)

Kontrola normality rozdělení dat
Shapiro-Wilks

Nenormální



Neparametrické testy

(H_0 : shoda rozdělení)

Parametrické testy – pro normálně rozdělená data

- 2 výběrové soubory

- a) F-test - testování rozdílu 2 rozptylů ($H_0: \sigma_1^2 = \sigma_2^2$), MS Excel, Unistat
- b) t-test - testování rozdílu 2 průměrů ($H_0: \mu_1 = \mu_2$), MS Excel, Unistat

- 3 a více výběrových souborů

- a) ANOVA - testování rozdílů více průměrů ($H_0: \mu_1 = \mu_2 = \mu_3 \dots = \mu_k$), Unistat

F-test

- vliv pokusného zásahu na rozptýlení hodnot sledované veličiny (součást metody ANOVA)
- porovnání přesnosti 2 metod měření
- před nepárovým t-testem (nepárový t-test pro shodné/různé rozptyly)

F-test

MS Excel

Vložit funkci

Vyhledat funkci:

Zadejte stručný popis požadované činnosti a poté klikněte na tlačítko Přejít.

Vybrat kategorii: Statistické

Vybrat funkci:

- F.DIST.RT
- F.INV
- F.INV.RT
- F.TEST**
- FISHER
- FISHERINV
- FORECAST.ETS

F.TEST(matice1;matice2)

Vrátí výsledek F-testu, tj. pravděpodobnosti (dva chvosty), že se rozptýly v argumentech matice1 a matice2 výrazně neliší.

[Nápověda k této funkci](#)

Argumenty funkce

F.TEST

Matice1 F1:F10 = {3.24;3.07;2.98;2.94;3.52;2.69;3.16;3....}

Matice2 G1:G10 = {3.51;3.69;2.99;4.01;3.22;4.5;4.62;3.9...}

= 0.046204394

Vrátí výsledek F-testu, tj. pravděpodobnosti (dva chvosty), že se rozptýly v argumentech matice1 a matice2 výrazně neliší.

Matice2 je druhá matice nebo oblast dat. Hodnoty argumentu mohou být čísla, matice nebo odkazy obsahující čísla. (Prázdné buňky jsou ignorovány.).

Výsledek = 0.046204394

[Nápověda k této funkci](#)

OK Storno

$p > 0,05$ nevýznamný rozdíl mezi rozptyly

$p < 0,05$ významný rozdíl mezi rozptyly

t-test

- **Jednovýběrový t-test** - porovnání základního a výběrového souboru, známe střední hodnotu (např. fyziologická hodnota)
- **Dvouvýběrový t-test** – porovnání dvou výběrových souborů
 - a) párový – závislé výběry (1 skupina zvířat, před a po pokusném zásahu)
 - b) nepárový – nezávislé výběry (2 skupiny pokus x kontrola), v závislosti na výsledku F-testu nepárový t-test pro shodné nebo různé rozptyly

t-test

MS Excel

B15  =

	A	B	C
1		Kortizol (ng/ml)	
2	Č. zv.	Před nakládkou	Po nakládce
3	1	86.9	98.3
4	2	92.6	102.7
5	3	99.6	124.8
6	4	85.3	144.2
7	5	100.3	125.6
8	6	75.9	115.3
9	7	86.3	111.0
10	8	101.6	114.3
11	9	100.2	129.1
12	10	86.6	117.3
13	Průměr	91.53	118.26
14	SD	8.666	13.357
15	t-test	=	

Vložit funkci

Vyhledat funkci:

Zadejte stručný popis požadované funkce, klikněte na tlačítko Přejít.

Vybrat kategorii: Statistické

Vybrat funkci:

T.INV.2T
T.TEST
TRIMMEAN

Argumenty funkce

T.TEST

Matice1 B3:B12 = {86.9;92.6;99.6;85.3;100.3;75.9;86.3;101.6;100.2;86.6}

Matice2 C3:C12 = {98.3;102.7;124.8;144.2;125.6;115.3;111.0;114.3;129.1;117.3}

Chvosty 2 = 2

Typ 1 = 1

= 0.000270112

Vrátí pravděpodobnost odpovídající Studentovu t-testu.

Typ je typ t-testu: spárované výběry = 1, dva výběry se shodným rozptylem = 2, dva výběry s různým rozptylem = 3.

Výsledek = 0.000270112

[Nápověda k této funkci](#)

OK Storno

$p > 0,05$ nevýznamný rozdíl mezi rozptyly
 $p < 0,05$ významný rozdíl mezi rozptyly
 $p < 0,01$ vysoce významný rozdíl mezi rozptyly

ANOVA (Analýza rozptylu)

- zjišťuje rozdíl mezi průměry tří a více skupin
- skupiny jsou na sobě navzájem nezávislé
- je splněna podmínka normality (Shapiro-Wilk)
- je splněna podmínka homogeneity rozptylů (shodný rozptyl ve všech skupinách), Levenův test

ANOVA testy mnohonásobného porovnávání

- pokud jsme zjistili pomocí analýzy rozptylu statistický rozdíl mezi průměry testovaných skupin, musíme pomocí post-hoc testu zjistit statistické rozdíly mezi jednotlivými skupinami
- test pro mnohonásobné porovnávání – Tukey HSD, Dunett...
- výsledkem jsou hodnoty pravděpodobnosti pro jednotlivé porovnávané skupiny

Neparametrické testy – data nemají normální rozdělení

- výpočty testů vycházejí z pořadových čísel měření – redukce informační hodnoty primárních dat
- neparametrické testy mají při stejném počtu pozorování nižší sílu testu než parametrické testy
- lze je použít při hodnocení souborů s nízkým počtem pozorování

Neparametrické testy

- **Dvouvýběrové** – porovnání dvou výběrových souborů
 - a) párový – závislé výběry (1 skupina zvířat, před a po pokusném zásahu) – Wilcoxonův test
 - b) nepárový – nezávislé výběry (2 skupiny pokus x kontrola) – Mann-Whitney test
- **3 a více výběrových souborů**
 - a) Kruskal-Wallis ANOVA – neparametrická alternativa one way ANOVY, testy pro mnohonásobné porovnávání

Typ dat $\Rightarrow$ volba testu

2) Kvalitativní „kategoriální data“

- Binární: mohou nabývat pouze dvou hodnot, mortalita (úhyn – přežití), vyšetření na parazity (pozitivní – negativní)
- Nominální: obsahují více kategorií, které nelze vzájemně seřadit, krevní skupiny (A/B/AB/0)
- Ordinální: obsahují více kategorií, na rozdíl od nominálních dat je však lze vzájemně seřadit, stádium maligního onemocnění (I/II/III/IV)

χ^2 test, kontingenční tabulky

- u kvalitativních dat pracujeme s četností výskytu jednotlivých znaků
- k tomu využíváme kontingenční tabulky 2x2 nebo k x m
- tabulka 2x2 slouží k posouzení závislosti dvou binárních znaků (samci a samice, pozitivní a negativní
- tabulka k x m slouží k posouzení více znaků

Kontingenční tabulka $k \times m$

Při vyšetření masa zvěře na přítomnost parazitů (trichinella) byl sledován počet pozitivních a negativních vzorků ze 3 lokalit. Liší se četnost výskytu pozitivních a negativních vzorků mezi jednotlivými lokalitami?

	Negativní	Pozitivní
Lokalita A	96	25
Lokalita B	121	22
Lokalita C	89	16

Kontingenční tabulka k x m

Unistat:

	A	B	C	D	E	F	G	H	I	J
1		Negativní	Pozitivní							
2	A	96	25							
3	B	121	22							
4	C	89	16							
5										
6										
7										
8										
9										

	A	B	C	D
1	Kontingenční tabulka			
2	3 Řádky x 2 Sloupce			
3	Chi-kvadrát testy			
4		Statistika Chí-kvadrát	Stupně volnosti	Pravostr. pravděp.
5	Pearson	1.6379	2	0.4409
6	Věrohodnostní poměr	1.5983	2	0.4497
7	+ Yatesova korekce			
8	# Lineární /Lineární	1.2284	1	0.2677
9	~ McNemar-Bowker			
10				

$p > 0,05$ nevýznamný (-)

$p < 0,05$ významný (+)

$p < 0,01$ vysoce významný (++)

Korelační a regresní analýza

- popisují vztah mezi dvěma spojitými veličinami

Korelační analýza

- kvantifikuje míru vztahu dvou proměnných
- rozlišujeme parametrickou a neparametrickou korelaci

Regresní analýza

- vytváří model vztahu dvou proměnných, jakým způsobem jedna proměnná závisí na druhé

Korelační analýza

- kvantifikuje míru vztahu dvou spojitých proměnných
- výpočet korelačního koeficientu ($r = \langle -1; +1 \rangle$)

Parametrický (Pearsonův korelační koeficient)

- hodnotí míru lineární závislosti mezi dvěma spojitými proměnnými

Neparametrický (Spearmanův korelační koeficient)

- hodnotí míru lineární závislosti pořadí hodnot dvou spojitých proměnných

Korelační koeficient: MS Excel

Vložit funkci

Vyhledat funkci:

Zadejte stručný popis požadované činnosti a potom klikněte na tlačítko Přejít.

Přejít

Vybrat kategorii: Statistické

Vybrat funkci:

- BINOM.INV
- CONFIDENCE.NORM
- CONFIDENCE.T
- CORREL**
- COUNTBLANK
- COUNTIF
- COUNTIFS

CORREL(matice1;matice2)
Vrátí korelační koeficient mezi dvěma množinami dat.

[Nápověda k této funkci](#)

Argumenty funkce

CORREL

Matice1	B2:B11	= {2.2;1.8;2.1;1.7;2.4;2.2;1.9;2.3;1.9}
Matice2	C2:C11	= {41;36;40;36;42;39;40;37;41;38}

= 0.952664357

Vrátí korelační koeficient mezi dvěma množinami dat.

Matice2 je druhá oblast buněk s hodnotami. Hodnoty mohou být čísla, názvy, matice nebo odkazy obsahující čísla.

Výsledek = 0.952664357

[Nápověda k této funkci](#)

OK Storno



Financováno
Evropskou unií
NextGenerationEU



Popisné charakteristiky statistických souborů

Biochemické a hematologické parametry myší

	hct_samci	pH_samci	HCO3_samci	BMI_samci	hct_samice	pH_samice	HCO3_samice	BMI_samice
	47	7,37	24,7	0,426	48	7,313	29,2	0,407
	51	7,321	28,3	0,452	49	7,398	26,2	0,512
	49	7,33	24,3	0,481	44	7,318	17,6	0,469
	51	7,387	25,9	0,490	54	7,425	26,4	0,405
	51	7,322	26,3	0,495	52	7,388	27,8	0,465
	38	7,425	27	0,391	47	7,332	25,8	0,417
	48	7,316	27,5	0,428	49	7,382	26	0,418
	42	7,326	24,1	0,400	49	7,52	26,8	0,398
	42	7,396	24,4	0,385	51	7,357	27,4	0,440
	51	7,417	24	0,438	48	7,362	24,2	0,486
	48	7,31	25,4	0,444	48	7,364	20,7	0,480
	47	7,4	31,7	0,442	50	7,357	25,1	0,459
	51	7,34	28,7	0,450	47	7,417	27	0,400
	43	7,28	20,6	0,430	44	7,394	30,7	0,470
	47	7,311	17,1	0,435	51	7,378	22,3	0,453
průměr	47,0666667	7,3500667	25,33333333	0,439133333	48,73333333	7,380333333	25,54666667	0,445266667
medián	48	7,33	25,4	0,438	49	7,378	26,2	0,453
SD	4,07898213	0,0450214	3,435459351	0,032682163	2,71152742	0,050610087	3,302783819	0,03592744
min.	38	7,28	17,1	0,385	44	7,313	17,6	0,398
max.	51	7,425	31,7	0,495	54	7,52	30,7	0,512

Výběrové charakteristiky

Metoda kvantilů: Prostý průměr

	hct_samci	pH_samci	HCO3_samci	BMI_samci	hct_samice	pH_samice	HCO3_samice	BMI_samice
Průměr	47,0667	7,3501	25,3333	0,4391	48,7333	7,3803	25,5467	0,4453
Medián	48,0000	7,3300	25,4000	0,4380	49,0000	7,3780	26,2000	0,4530
Dolní kvartil	43,0000	7,3160	24,1000	0,4260	47,0000	7,3570	24,2000	0,4070
Horní kvartil	51,0000	7,3960	27,5000	0,4520	51,0000	7,3980	27,4000	0,4700
Minimum	38,0000	7,2800	17,1000	0,3850	44,0000	7,3130	17,6000	0,3980
Maximum	51,0000	7,4250	31,7000	0,4950	54,0000	7,5200	30,7000	0,5120
Nevychýlený rozptyl	16,6381	0,0020	11,8024	0,0011	7,3524	0,0026	10,9084	0,0013
Nevychýlená směr odchylka	4,0790	0,0450	3,4355	0,0327	2,7115	0,0506	3,3028	0,0359
Směrodatná chyba průměru	1,0532	0,0116	0,8870	0,0084	0,7001	0,0131	0,8528	0,0093



Financováno
Evropskou unií
NextGenerationEU



Neárový test-dle normality: Neárový t-test (data mají normální rozdělení, Mann-Whitney test (data nemají normální rozdělení).

Biochemické a hematologické parametry myši

hct_samci	pH_samci	HCO3_samci	BMI_samci	hct_samice	pH_samice	HCO3_samice	BMI_samice
47	7,37	24,7	0,426	48	7,313	29,2	0,407
51	7,321	28,3	0,452	49	7,398	26,2	0,512
49	7,33	24,3	0,481	44	7,318	17,6	0,469
51	7,387	25,9	0,490	54	7,425	26,4	0,405
51	7,322	26,3	0,495	52	7,388	27,8	0,465
38	7,425	27	0,391	47	7,332	25,8	0,417
48	7,316	27,5	0,428	49	7,382	26	0,418
42	7,326	24,1	0,400	49	7,52	26,8	0,398
42	7,396	24,4	0,385	51	7,357	27,4	0,440
51	7,417	24	0,438	48	7,362	24,2	0,486
48	7,31	25,4	0,444	48	7,364	20,7	0,480
47	7,4	31,7	0,442	50	7,357	25,1	0,459
51	7,34	28,7	0,450	47	7,417	27	0,400
43	7,28	20,6	0,430	44	7,394	30,7	0,470
47	7,311	17,1	0,435	51	7,378	22,3	0,453

normalita NE ANO ANO ANO ANO ANO ANO ANO

Shapiro-Wilk

Test normality

Malé pravděpodobnosti indikují nenormalitu.

	Pravděpodob.
hct_samci	0,0267 p<0,05
pH_samci	0,1561 p>0,05
HCO3_samci	0,4451 p>0,05
BMI_samci	0,4729 p>0,05
hct_samice	0,7138 p>0,05
pH_samice	0,0748 p>0,05
HCO3_samice	0,2382 p>0,05
BMI_samice	0,2799 p>0,05

Test normality

Malé pravděpodobnosti indikují nenormalitu.

* Lillieforsova pravděpodobnost = 0.2 znamená 0.2 nebo větší.

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka	Shapiro-Wilkův test	Pravděpodob.	Kolmogorov- Smirnovův test	* Pravděpodob.
hct_samci	15	0	47,0667	4,0790	0,8630	0,0267	0,2268	0,0366
pH_samci	15	0	7,3501	0,0450	0,9140	0,1561	0,2054	0,0885
HCO3_samci	15	0	25,3333	3,4355	0,9447	0,4451	0,2156	0,0589
BMI_samci	15	0	0,4391	0,0327	0,9466	0,4729	0,1469	0,2000
hct_samice	15	0	48,7333	2,7115	0,9612	0,7138	0,1280	0,2000
pH_samice	15	0	7,3803	0,0506	0,8931	0,0748	0,1635	0,2000
HCO3_samice	15	0	25,5467	3,3028	0,9261	0,2382	0,1972	0,1168
BMI_samice	15	0	0,4453	0,0359	0,9307	0,2799	0,1761	0,2000



Financováno
Evropskou unií
NextGenerationEU



Párový test-dle normality: Párový t-test (data mají normální rozdělení, Wilcoxonův test (data nemají normální rozdělení).

Biochemické a hematologické parametry myši

ID	hct_hib	hct_post	pH_hib	pH_post	HCO3_hib	HCO3_post	BMI_hib	BMI_post
1	47	37	7,37	7,247	24,7	25,3	0,426	0,407
2	51	51	7,321	7,277	28,3	17,2	0,452	0,424
3	51	53	7,387	7,258	25,9	19,2	0,490	0,426
4	51	53	7,322	7,286	26,3	24,6	0,495	0,420
5	48	51	7,316	7,276	27,5	25,3	0,428	0,390
6	51	53	7,417	7,254	24	22,8	0,438	0,401
7	48	45	7,31	7,196	25,4	15,5	0,444	0,426
8	47	47	7,4	7,328	31,7	26,3	0,442	0,414
9	51	49	7,34	7,306	28,7	23	0,450	0,432
10	43	41	7,28	7,343	20,6	21,8	0,430	0,402
11	47	51	7,311	7,286	17,1	21	0,435	0,396
12	49	51	7,398	7,357	26,2	22,8	0,512	0,476
13	44	37	7,318	7,296	17,6	13,7	0,469	0,450
14	54	55	7,425	7,344	26,4	25,7	0,405	0,387
15	52	46	7,388	7,305	27,8	25	0,465	0,456
16	47	45	7,332	7,299	25,8	25,7	0,417	0,398
17	49	53	7,382	7,191	26	17,8	0,418	0,372
18	49	42	7,52	7,371	26,8	17,3	0,398	0,446
19	51	49	7,357	7,287	27,4	24,6	0,440	0,404
20	48	48	7,462	7,241	24,2	18,2	0,486	0,439
21	48	40	7,364	7,202	20,7	15,5	0,480	0,462
22	47	42	7,517	7,345	27	20,2	0,400	0,418
23	44	49	7,394	7,339	30,7	28,7	0,470	0,432
24	51	46	7,478	7,201	22,3	23,1	0,453	0,445

normalita ANO ANO ANO ANO ANO ANO ANO ANO

Shapiro-Wilk

Test normality

Malé pravděpodobnosti indikují nenormalitu.

	Pravděpodo b.	
hct_hib	0,1661	p>0,05
hct_post	0,1364	p>0,05
pH_hib	0,1113	p>0,05
pH_post	0,2622	p>0,05
HCO3_hib	0,0997	p>0,05
HCO3_post	0,2673	p>0,05
BMI_hib	0,7915	p>0,05
BMI_post	0,9897	p>0,05

Test normality

Malé pravděpodobnosti indikují nenormalitu.

* Lillieforsova pravděpodobnost = 0.2 znamená 0.2 nebo větší.

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka	Shapiro-Wilkův test	Pravděpodob.	Kolmogorov- Smirnovův test	* Pravděpodob.
hct_hib	24	0	48,6667	2,7134	0,9404	0,1661	0,1801	0,0428
hct_post	24	0	47,2500	5,2274	0,9365	0,1364	0,1384	0,2000
pH_hib	24	0	7,3795	0,0654	0,9326	0,1113	0,1272	0,2000
pH_post	24	0	7,2848	0,0531	0,9493	0,2622	0,1072	0,2000
HCO3_hib	24	0	25,3792	3,5955	0,9304	0,0997	0,1716	0,0657
HCO3_post	24	0	21,6792	4,0583	0,9497	0,2673	0,1505	0,1577
BMI_hib	24	0	0,4476	0,0309	0,9751	0,7915	0,0975	0,2000
BMI_post	24	0	0,4218	0,0260	0,9881	0,9897	0,0902	0,2000



Financováno
Evropskou unií
NextGenerationEU



ANOVA

Biochemické a hematologické parametry myši

hct_neg	pH_neg	HCO3_neg	BMI_neg	hct_try	pH_try	HCO3_try	BMI_try	hct_bab	pH_bab	HCO3_bab	BMI_bab
47	7,37	24,7	0,426	49	7,33	24,3	0,481	47	7,311	17,1	0,435
51	7,321	28,3	0,452	51	7,387	25,9	0,490	44	7,318	17,6	0,469
48	7,316	27,5	0,428	51	7,322	26,3	0,495	52	7,388	27,8	0,465
42	7,326	24,1	0,400	47	7,4	31,7	0,442	49	7,382	26	0,418
42	7,396	24,4	0,385	49	7,398	26,2	0,512	46	7,317	17,3	0,439
51	7,417	24	0,438	48	7,462	24,2	0,486	45	7,314	17,8	0,461
48	7,31	25,4	0,444	48	7,364	20,7	0,480	47	7,345	24,1	0,452
51	7,34	28,7	0,450	51	7,478	22,3	0,453	48	7,371	21,1	0,402
48	7,313	29,2	0,407	50	7,31	24,2	0,482	46	7,315	18,3	0,432
54	7,425	26,4	0,405	51	7,381	25,7	0,487	49	7,326	19,2	0,429
47	7,332	25,8	0,417	50	7,328	26,5	0,497	47	7,415	18,7	0,433
49	7,52	26,8	0,398	48	7,42	31,8	0,441	45	7,335	17,9	0,441
51	7,357	27,4	0,440	50	7,392	26,7	0,518	49	7,341	19	0,423
50	7,357	25,1	0,459	49	7,467	24,5	0,481	46	7,314	18,2	0,421
47	7,517	27	0,400	47	7,36	20,8	0,482	49	7,305	16,9	0,420
44	7,394	30,7	0,470	50	7,475	22,4	0,456	49	7,305	16,9	0,420

normalita	ANO	NE	ANO	ANO	ANO	ANO	ANO	ANO	NE	NE	ANO
-----------	-----	----	-----	-----	-----	-----	-----	-----	----	----	-----

Shapiro-Wilk

Test normality

Malé pravděpodobnosti indikují nenormalitu.

	Pravděpodobnost
hct_neg	0,2692 $p>0,05$
pH_neg	0,0112 $p<0,05$
HCO3_neg	0,6183 $p>0,05$
BMI_neg	0,6175 $p>0,05$
hct_try	0,0786 $p>0,05$
pH_try	0,2263 $p>0,05$
HCO3_try	0,1010 $p>0,05$
BMI_try	0,1984 $p>0,05$
hct_bab	0,3635 $p>0,05$
pH_bab	0,0116 $p<0,05$
HCO3_bab	0,0008 $p<0,05$
BMI_bab	0,4346 $p>0,05$

Uspořádání dat pro potřeby ANOVY

hct	pH	HCO3	BMI	status
47	7,37	24,7	0,426	negativní
51	7,321	28,3	0,452	negativní
48	7,316	27,5	0,428	negativní
42	7,326	24,1	0,400	negativní
42	7,396	24,4	0,385	negativní
51	7,417	24	0,438	negativní
48	7,31	25,4	0,444	negativní
51	7,34	28,7	0,450	negativní
48	7,313	29,2	0,407	negativní
54	7,425	26,4	0,405	negativní
47	7,332	25,8	0,417	negativní
49	7,52	26,8	0,398	negativní
51	7,357	27,4	0,440	negativní
50	7,357	25,1	0,459	negativní
47	7,517	27	0,400	negativní
44	7,394	30,7	0,470	negativní
49	7,33	24,3	0,481	trypanos
51	7,387	25,9	0,490	trypanos
51	7,322	26,3	0,495	trypanos
47	7,4	31,7	0,442	trypanos
49	7,398	26,2	0,512	trypanos
48	7,462	24,2	0,486	trypanos
48	7,364	20,7	0,480	trypanos
51	7,478	22,3	0,453	trypanos
50	7,31	24,2	0,482	trypanos
51	7,381	25,7	0,487	trypanos
50	7,328	26,5	0,497	trypanos
48	7,42	31,8	0,441	trypanos
50	7,392	26,7	0,518	trypanos
49	7,467	24,5	0,481	trypanos
47	7,36	20,8	0,482	trypanos
50	7,475	22,4	0,456	trypanos
47	7,311	17,1	0,435	babesia
44	7,318	17,6	0,469	babesia
52	7,388	27,8	0,465	babesia
49	7,382	26	0,418	babesia
46	7,317	17,3	0,439	babesia
45	7,314	17,8	0,461	babesia
47	7,345	24,1	0,452	babesia
48	7,371	21,1	0,402	babesia
46	7,315	18,3	0,432	babesia
49	7,326	19,2	0,429	babesia
47	7,415	18,7	0,433	babesia
45	7,335	17,9	0,441	babesia
49	7,341	19	0,423	babesia
46	7,314	18,2	0,421	babesia
49	7,305	16,9	0,420	babesia
49	7,305	16,9	0,420	babesia

Test normality

Malé pravděpodobnosti indikují nenormalitu.

* Lillieforsova pravděpodobnost = 0.2 znamená 0.2 nebo větší.

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka	Shapiro-Wilkův test	Pravděpodob.	Kolmogorov- Smirnovův test	* Pravděpodob.
hct_neg	16	0	48,1250	3,3441	0,9327	0,2692	0,1808	0,1637
pH_neg	16	0	7,3757	0,0668	0,8441	0,0112	0,1727	0,2000
HCO3_neg	16	0	26,5938	1,9723	0,9576	0,6183	0,1025	0,2000
BMI_neg	16	0	0,4262	0,0254	0,9575	0,6175	0,1504	0,2000
hct_try	16	0	49,3125	1,4009	0,8994	0,0786	0,1882	0,1297
pH_try	16	0	7,3921	0,0561	0,9280	0,2263	0,1435	0,2000
HCO3_try	16	0	25,2625	3,1812	0,9062	0,1010	0,2007	0,0844
BMI_try	16	0	0,4802	0,0223	0,9244	0,1984	0,2466	0,0102
hct_bab	16	0	47,3750	2,0616	0,9412	0,3635	0,1597	0,2000
pH_bab	16	0	7,3376	0,0338	0,8453	0,0116	0,2195	0,0380
HCO3_bab	16	0	19,6188	3,3855	0,7570	0,0008	0,2992	0,0005
BMI_bab	16	0	0,4350	0,0188	0,9464	0,4346	0,1250	0,2000

Analýza rozptylu

Přístup: Klasický experiment

Závisle proměnná: hct

Zdroj variability	Součet čtverců	St. vol.	Průměrný čtverec	Stat F	Pravd
Hlavní efekty	30,542	2	15,271	2,634	0,0829
status	30,542	2	15,271	2,634	0,0829
Vysvětleno	30,542	2	15,271	2,634	0,0829
Chyba	260,938	45	5,799		
Celkem	291,479	47	6,202		

- p>0.05



Kvalitativní data-chí-kvadrát test metodou kontingenčních tabulek

1) Přítomnost parazitů v mase zvěře (trichinella)

	negativní	pozitivní
Lokalita A	96	25
Lokalita B	121	22
Lokalita C	89	16

Pearson $p=0.44088731579170$ $p>0.05$

procentuální vyjádření

negativní	pozitivní
79,33884	20,66116
84,61538	15,38462
84,7619	15,2381

	negativní	pozitivní
Lokalita A	96	25
Lokalita B	22	121
Lokalita C	89	16

Pearson $p=5.39185309717211E-35$ $p<0.01$

procentuální vyjádření

negativní	pozitivní
79,33884	20,66116
15,38462	84,61538
84,7619	15,2381

Yatesova korekce $p=7.84167558209036E-25$ $p<0.01$

	negativní	pozitivní
Lokalita A	96	25
Lokalita B	22	121

Yatesova korekce $p=0.377729594352029$ $p>0.05$

2) Porovnání četností aktivního a klidového chování u velkého a malého psa v různých obdobích roku

aktivní chování

	rozdíl mezi aktivitou	aktivita lét	aktivita podzim
pes velký	43	74	
pes malý	27	89	

Yatesova korekce $p=0.03567591640806$ $p<0.05$

procentuální vyjádření

aktivita lét	aktivita podzim
36,75214	63,24786
23,27586	76,72414

18,45494 31,75966 50,21459
11,58798 38,19742 49,78541
30,04292 69,95708

klidové chování

	klid léto	klid podzim
pes velký	199	151
pes malý	207	159

Yatesova korekce $p=0.99562831039408$ $p>0.05$

procentuální vyjádření

klid léto	klid podzim
56,85714	43,14286
56,55738	43,44262

27,7933 21,08939 48,88268
28,91061 22,2067 51,11732
56,70391 43,29609

			procentuální vyjádření	
	negativní	pozitivní	negativní	pozitivní
Lokalita B	22	121	15,38%	84,62%
Lokalita C	89	16	84,76%	15,24%

Yatesova korekce $p=7.59209764534036E-27$ $p<0.01$

Závěr: Mezi lokalitami A a B a mezi lokalitami B a C je statisticky vysoce významný rozdíl v četnostech výskytu pozitivních a negativních vzorků.

3) Porovnání četnosti výskytu škrkavek u psů mezi útluky

	škrkavky + škrkavky -	
útulek A	0	120
útulek B	63	24

Fischerův přesný test $p=1.62658666579811$ $p<0.01$

procentuální vyjádření

škrkavky + škrkavky -	
0	100
72,41379	27,58621

0 57,97101 57,97101
30,43478 11,5942 42,02899
30,43478 69,56522

Závěr: V útluku B je statisticky vysoce významně vyšší četnost výskytu škrkavek.

Kontingenční tabulka

Celkem % - procentuální vyjádření

		C1	C2	Součet řádku
	R1	26,02%	6,78%	32,79%
	R2	32,79%	5,96%	38,75%
	R3	24,12%	4,34%	28,46%
	Sl.souč.	82,93%	17,07%	100,00%

Sloupec % - procentuální vyjádření

		C1	C2
	R1	31,37%	39,68%
	R2	39,54%	34,92%
	R3	29,08%	25,40%

Řádek % - procentuální vyjádření

		C1	C2
	R1	79,34%	20,66%
	R2	84,62%	15,38%
	R3	84,76%	15,24%

Chi-kvadrát testy

	Statistika Chi-kvadrát	Stupně volnosti	Pravostranná pravděpodobnost
Pearson	1,6379	2	0,4409
Věrohodnostní poměr	1,5983	2	0,4497
+ Yatesova korekce			
# Lineární /Lineární	1,2284	1	0,2677
~ McNemar-Bowker			

pro kontingenční tabulku k x m

pro kontingenční tabulku 2 x 2

+ Vykazováno pro 2 x 2 tabulky.

Tabulka skóru

~ Vykazováno pro 3 x 3 nebo větší čtvercové tabulky.

Buňky s očekávaným počtem < 5 = 0 (0,00%)

Minimální očekávaný počet = 17,9268

Fí = 0,0666

Koeficient kontingence = 0,0665

Cramerovo V = 0,0666

Fisherův přesný test

	Dvoustranná pravděpodobnost	Tabulka pravděpodobnosti
Fisherův přesný	0,4738	0,0072

v případě, že je některá četnost <5

Regresní a korelační analýza

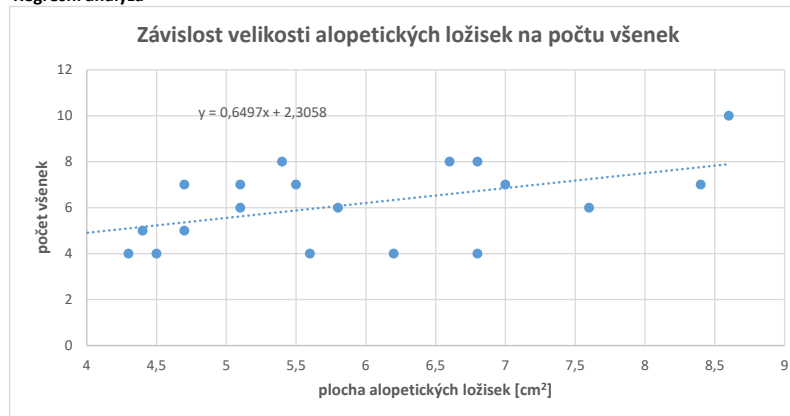
V chovatelské stanici morčat, ve které došlo k masivnímu rozšíření všenek, byla sledována závislost mezi plochou alopetických ložisek na kůži (cm²) a počtem všenek. Všeny byly posbírány z |

ID	cm ²	počet všenek
1	5,5	7
2	7,6	6
3	5,1	7
4	3,7	5
5	7	7
6	4,3	4
7	4,7	7
8	6,8	8
9	6,8	4
10	5,8	6
11	8,4	7
12	5,4	8
13	6,2	4
14	4,5	4
15	8,6	10
16	6,6	8
17	4,4	5
18	5,6	4
19	4,7	5
20	5,1	6
normalita	ANO	ANO

Pearsonův korelační koeficient-Excel
Pearsonův korelační koeficient-unistat
Ověření statistické významnosti korelačního koeficientu

r= 0,522191
r=0,522190785521878
p=0,0181837396049976
r>1, pozitivní/přímá korelace, středně silná korelace
p<0,05, statisticky významná korelace

Regresní analýza



Závěr: Mezi počtem všenek a velikostí alopetických ložisek je statisticky významná, pozitivní, středně silná korelace.

Test normality

Malé pravděpodobnosti indikují nenormalitu.

	Pravděpodo b.	
cm2	0,4146	p>0,05
počet všenek	0,0781	p>0,05

Korelační koeficienty

Pro cm2 a počet všenek

	dvoustranná Korelační koeficient	pravděpodo bnost
Pearson	0,5222	0,0182 p<0,05
Spearmanov o pořadí	0,4484	0,0474

Test normality

Malé pravděpodobnosti indikují nenormalitu.

* Lillieforsova pravděpodobnost = 0.2 znamená 0.2 nebo větší.

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka	Shapiro-Wilkův test	Pravděpodob.	Kolmogorov- Smirnovův test	* Pravděpodob.
cm2	20	0	5,8400	1,3774	0,9530	0,4146	0,1192	0,2000
počet všenek	20	0	6,1000	1,7137	0,9146	0,0781	0,1503	0,2000

Korelační koeficienty

Pro cm2 a počet všenek

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka
cm2	20	0	5,8400	1,3774
počet všenek	20	0	6,1000	1,7137
Párový	20	0		

	Korelační koeficient	Stupně volnosti	* Testovací statistika	jednostranná pravděp.	dvoustranná pravděpodobno st	Dolní 95%	Horní 95%
Pearson	0,5222	18	2,5978	0,0091	0,0182	0,1036	0,7836
					p<0,05		

* Z-statistika pro Kendalovo pořadí, t-statistika jinak



Financováno
Evropskou unií
NextGenerationEU

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**Národní
plán
obnovy**

Parametrické testy - dvouvýběrové



F-test

F-test

testování rozdílu 2 rozptylů

patří do skupiny **parametrických testů**
(vyžaduje normální rozdělení dat)

Použití v situacích, kdy potřebujeme zjistit zda:

- **experimentální zásah změnil rozptyl měřené veličiny**
- **2 populace mají stejnou variabilitu sledovaného znaku**
- **„předtest“ pro t-test**

t-test

t-test

testování rozdílu 2 středních hodnot

patří do skupiny **parametrických testů**
(vyžaduje normální rozdělení dat)

varianty t-testu

ZS x VS

ZS - μ (známá hodnota
např. z literatury)
VS - $n, \bar{x}, s$

VS x VS

VS₁ - $n, \bar{x}_1, s_1$
VS₂ - $n, \bar{x}_2, s_2$



PÁROVÝ

1 skupina

VS_{před} => VS_{po}

NEPÁROVÝ

2 skupiny

VS₁ VS₂

pravděpodobnost chyby α

Pokud je vypočítaná hodnota p

$p \geq 0,05 \Rightarrow$ nevýznamný rozdíl mezi testovanými parametry (-)

$p < 0,05 \Rightarrow$ významný rozdíl mezi testovanými parametry (+)

$p < 0,01 \Rightarrow$ vysoce významný rozdíl mezi testovanými parametry (++)



Financováno
Evropskou unií
NextGenerationEU

**MS
MT**
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Národní
plán
obnovy

Bakalářská práce I – Metodologie a biostatistika

Neparametrické testy



Charakteristika

- testování souborů s neznámým rozdělením (nelze popsat pomocí μ a σ)
- hypotéza se týká shody rozdělení četností (H_0 : 2 soubory mají stejný tvar křivky rozdělení)
- výpočty vycházejí z pořadových čísel naměřených hodnot souboru

Použití neparametrických testů

- mohou být použity i u dat, která nemají přesnou číselnou hodnotu (ve skutečnosti jsou jen pořadím - „ordinální znaky“)
- výpočty jsou obvykle jednodušší (používají se i u dat s normálním rozdělením pro orientační hodnocení)
- přesnost a rozlišovací schopnost (síla testu) je nižší než u parametrických testů (mezi testovanými soubory musí být velký rozdíl, aby byl statisticky průkazný)

Mann-Whitney pořadový test

- Nepárový test – na sobě navzájem nezávislé soubory
- Dvouvýběrový test – 2 výběrové soubory
- Jedná se o „neparametrickou alternativu“ dvouvýběrového nepárového t-testu

Interpretace výsledků

Výsledky testování je možno interpretovat na základě výpočtu pravděpodobnosti chyby alfa (vypočítaná dvoustranná pravděpodobnost p)

Pokud je vypočítaná hodnota pravděpodobnosti

- $p \geq 0,05 \Rightarrow$ nevýznamný rozdíl mezi testovanými parametry
- $p < 0,05 \Rightarrow$ významný rozdíl mezi testovanými parametry
- $p < 0,01 \Rightarrow$ vysoce významný rozdíl mezi testovanými parametry

Příklad:

Byl sledován vliv vitamínu B₁₂ přidávaného do krmiva na zvyšování hmotnostních přírůstků u selat, máte k dispozici dvě skupiny selat (kontrolní a pokusnou) – zjistěte jaký vliv má vit B12 na selata:

Hmotnost selat bez přídavku vit. B12: 27, 35, 38, 37.5, 29.5, 33.5, 37, 31.5, 34, 32, 33, 34.5, 37.5, 29.5, 33.5, 37, 31.5, 34, 32

Hmotnost selat s přídavkem vit. B12: 32.5, 30.5, 36, 38.5, 36, 43, 40.5, 36, 31, 42, 35.5, 40, 38.5, 36, 43, 40.5, 36, 31, 42

UNISTAT

Menu:

Unistat – Statistika1 – Neparametrické testy (1-2 výběrové) – Nepárové výběry

Nepárové výběry: Krok 1

Vyberte typ dat

- ☒ Vyberte data jako proměnné a volitelné kategoričké sloupce jako faktory
- ☐ Sloupec 1 jsou data, sloupec 2 spojitá prom. (vlozte dělicí hodn.)
- ☐ Jsou dány výběrové charakteristiky

Seznam proměnných

Vybrat / Vynechat

Proměnné j

C1 A
C2 B

Faktor

Nepárové výběry: Krok 2

Vše | Žádný | Vyberte volby pro výstup nebo stiskněte na tlač. 'Opt' pro další možnosti

- ☒ Mann-Whitneyův U test
- ☐ Hodges-Lehmann estimátor (nepárový)
- ☐ Wald-Wolfowitzův test iterací
- ☐ Mosesův test extrémní reakce
- ☐ Dvouvýběrový mediánový test

Opt

Opt

Nepárové výběry

Mann-Whitneyův U test

	Přip.	Souč.poř.	Průměrné pořadí	U
A	19	282.0000	14.8421	269.0000
B	19	459.0000	24.1579	92.0000
Celkem	38	741.0000	19.5000	

Korekce na shody = 5.5000

	U	Testovací statistika	jednostr. pravděp.	dvoustr. pravděp.
Asymptotické normální	92.0000	-2.5853	0.0049	0.0097
Asymptotické normální s CC		-2.5707	0.0051	0.0101
Asymptotické t		-2.5853	0.0069	0.0138
Asymptotické t s CC		-2.5707	0.0072	0.0143
Exaktní			0.0044	0.0088

Wilcoxonův test

- Párový test – na sobě závislé soubory (párové výsledky)
- Dvouvýběrový test – 2 výběrové soubory
- Jedná se o „neparametrickou alternativu“ dvouvýběrového párového t-testu

Příklad:

Vyhodnoťte výsledky testu bakteriální kontaminace po ošetření 2 preparáty. Byly odebrány vzorky pro kultivaci (1.polovina vzorku byla ošetřena preparátem A, 2.polovina preparátem B). Po kultivaci byly spočítány počty bakteriálních kolonií na kultivačních miskách.

Zjistěte rozdíl v baktericidní účinnosti preparátů A a B:

Preparát A: 35, 35, 38, 42, 35, 38, 39, 36, 34, 37, 36, 38, 37,
33, 29

Preparát B: 36, 38, 37, 40, 38, 42, 26, 38, 36, 39, 40, 38, 40,
42, 30

UNISTAT

Menu:

Unistat – Statistika1 – Neparametrické testy (1-2 výběrové) – Párové výběry

The screenshot shows the UNISTAT software interface. In the background, a spreadsheet has columns A and B with data. Two dialog boxes are open: 'Párové výběry: Krok 1' and 'Párové výběry: Krok 2'. The 'Krok 1' dialog shows a list of variables with 'C1 A' and 'C2 B' selected. The 'Krok 2' dialog shows the 'Wilcoxonův test' selected. Below these, the 'Párové výběry' output window is displayed, showing the results of the Wilcoxon test for variables A and B.

Párové výběry
Wilcoxonův test
Pro A a B

	Přip.	Souč.poř.	Průměrné pořadí
Záporné odchylky	1	2.0000	2.0000
Kladné odchylky	7	34.0000	4.8571
Shody	0	0.0000	
Celkem	8	36.0000	4.5000

	W	u-hodnota	jednostr. pravděp.	dvoustr. pravděp.
Asymptotický	2.0000	-2.1812	0.0146	0.0292
Asymptotické s CC		-2.2515	0.0122	0.0244
Exaktní			0.0156	0.0313

Znaménkový test

- Párový test – na sobě závislé soubory (párové výsledky)
- Pro veličiny, které nelze měřit – stačí rozhodnutí, zda jeden pokusný zásah působil více / méně než druhý
- K orientačnímu hodnocení předběžných pokusů např. v mikrobiologii

Příklad:

Pro mikrobiologické vyšetření bylo odebráno 15 vzorků ($n=15$). Každý vzorek byl rozdělen na dvě poloviny a na první polovinu každého vzorku byl aplikován FA a na druhou Penicilin. Po kultivaci byl sledován počet bakterií na 1. a na 2. polovině vzorku.

Zjistěte rozdíl v účinnosti těchto dvou antibiotik.

13x $FA > P$

1x $FA < P$

1x $FA = P$

Obecný postup při testování souborů

1) Testování normality

UNISTAT

Menu: Unistat – Statistika1 – Testy dobré shody – Testy normality: např. Shapiro-Wilk test

The screenshot displays the UNISTAT software interface. In the background, a spreadsheet shows data for variables A and B. Overlaid on this are two dialog boxes and a results table.

Test normality: Krok 1 dialog box:

- Wyberte typ dat:
 - ☒ Vyberte data jako proměnné a volitelné kategorické sloupce jako faktory
 - ☐ Jsou dány výběrové charakteristiky
- Seznam proměnných: (Empty)
- Vybrat / Vynechat: Proměnná i
- Vybrané proměnné: C1 A, C2 B

Test normality: Krok 2 dialog box:

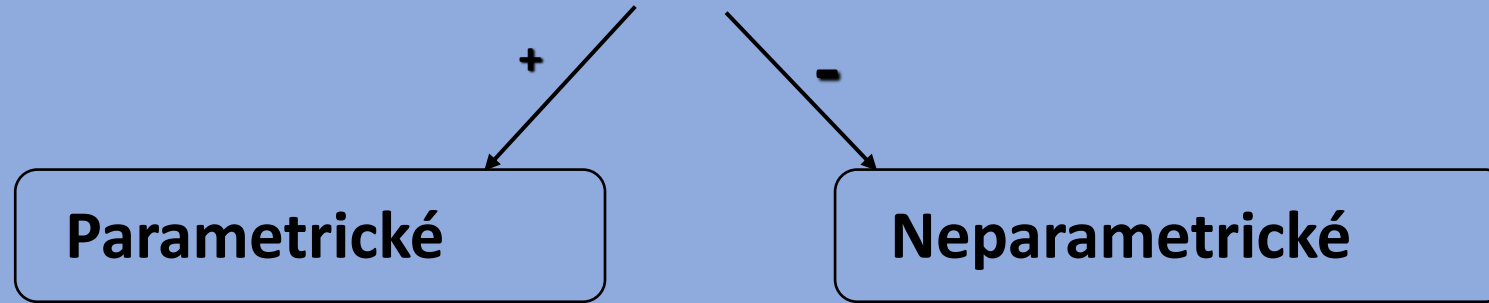
- Vše | Žádný | Vyberte volby pro výstup nebo stiskněte na tlač. 'Opt' pro další možnosti
- ☒ Shapiro-Wilkův test
- ☐ Kolmogorov-Smirnovův test
- ☐ Cramer-von Mises test
- ☐ Anderson-Darling Test
- Opt ☐ Normální pravděpodobnostní graf
- ☒ Přehled celkových statistik

Test normality results table:

		Platná pozorování	Chybějící	Průměr	Směrodatná odchylka	Shapiro-Wilkův test	Pravděp.
1	Test normality						
2	Malé pravděpodobnosti indikují nenormalitu.						
3							
4	A	19	0	33.4474	3.3078	0.9767	0.8981
5	B	19	0	36.5000	5.0607	0.8502	0.0067
6							

Obscný postup při testování souborů

Testování normality



(Vyloučení extrémních hodnot – Grubbsův t., Dixonův t.)

2 soubory



F- test ($\sigma_1^2 = \sigma_2^2$?)

t - test ($\mu_1 = \mu_2$?)

více souborů



ANOVA (analýza rozptylu)

($\mu_1 = \mu_2 = \mu_3 \dots = \mu_k$?)

Mann-Whitney test

(nepárový)

Wilcoxonův test

(párový)

Znaménkový test

(párový)



Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy

Mann-Whitney nepárový test

hmotnost selat v kg

selata bez B12	selata s B12
27	32,5
35	30,5
38	36
37,5	38,5
29,5	36
33,5	43
37	40,5
31,5	36
34	31
32	42
33	35,5
34,5	40
37,5	38,5
29,5	36
33,5	43
37	40,5
31,5	36
34	31
32	42
33,5	36

medián	33,5	36
Mann-Whitney	p=0.00537321966929744	++ p<0.01

Alternativní uspořádání dat-data uspořádána pro analýzu pomocí např. programu Statistica pro Windows

27	selata bez B12
35	selata bez B12
38	selata bez B12
37,5	selata bez B12
29,5	selata bez B12
33,5	selata bez B12
37	selata bez B12
31,5	selata bez B12
34	selata bez B12
32	selata bez B12
33	selata bez B12
34,5	selata bez B12
37,5	selata bez B12
29,5	selata bez B12
33,5	selata bez B12
37	selata bez B12
31,5	selata bez B12
34	selata bez B12
32	selata bez B12
32,5	selata s B12
30,5	selata s B12
36	selata s B12
38,5	selata s B12
36	selata s B12
43	selata s B12
40,5	selata s B12
36	selata s B12
31	selata s B12
42	selata s B12
35,5	selata s B12
40	selata s B12
38,5	selata s B12
36	selata s B12
43	selata s B12
40,5	selata s B12
36	selata s B12
31	selata s B12
42	selata s B12

Nepárové výběry

Mann-Whitneyův U test

	Příp.	Souč.poř.	Průměrné pořadí	U
selata bez B12	20	309,0000	15,4500	301,0000
selata s B12	20	511,0000	25,5500	99,0000
Celkem	40	820,0000	20,5000	

Korekce na shody = 25,0000

	U	Testovací statistika	jednostranná pravděp.	dvoustranná pravděpodobnost
Asymptotické normální	99,0000	-2,7385	0,0031	0,0062
Asymptotické normální s CC		-2,7249	0,0032	0,0064
Asymptotické t		-2,7385	0,0046	0,0093
Asymptotické t s CC		-2,7249	0,0048	0,0096
Exaktní			0,0027	0,0054

++ p<0.01



Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy

Wilcoxonův test, párový test

Počty bakteriálních kolonií na misce po kultivaci

Preparát A	Preparát B
35	36
35	38
38	37
42	40
35	38
38	42
39	26
36	38
34	36
37	39
36	40
38	38
37	40
33	42
29	30
medián	3638
Wilcoxonův test	p=0.0478515625 + p<0.05

Párové výběry

Wilcoxonův test

Pro Preparát A a Preparát B

	Příp.	Souč.poř.	Průměrné pořadí
Záporné odchylky	11	83,5000	7,5909
Kladné odchylky	3	21,5000	7,1667
Celkem	15	120,0000	8,0000

Korekce na shody = 2,3750

	W	u-hodnota	jednostranná pravděp.	dvoustranná pravděpodobnost	
Asymptotický	21,5000	-1,9237	0,0272	0,0544	
Asymptotické s CC		-1,9552	0,0253	0,0506	
Exaktní			0,0239	0,0479	+ p<0.05



Financováno
Evropskou unií
NextGenerationEU



Znaménkový test

Vyhodnocení nárůstu bakteriálních kolonií na miskách

vzorek č.	Formoalesten	Penicilin
1	1	0
2	1	0
3	1	0
4	0	1
5	1	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0
11	1	1
12	1	0
13	1	0
14	1	0
15	1	0
Znaménkový test		



Financováno
Evropskou unií
NextGenerationEU

**MS
MT**
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Národní
plán
obnovy

Bakalářská práce I – Metodologie a biostatistika Regresní a korelační analýza



Jednorozměrná statistika

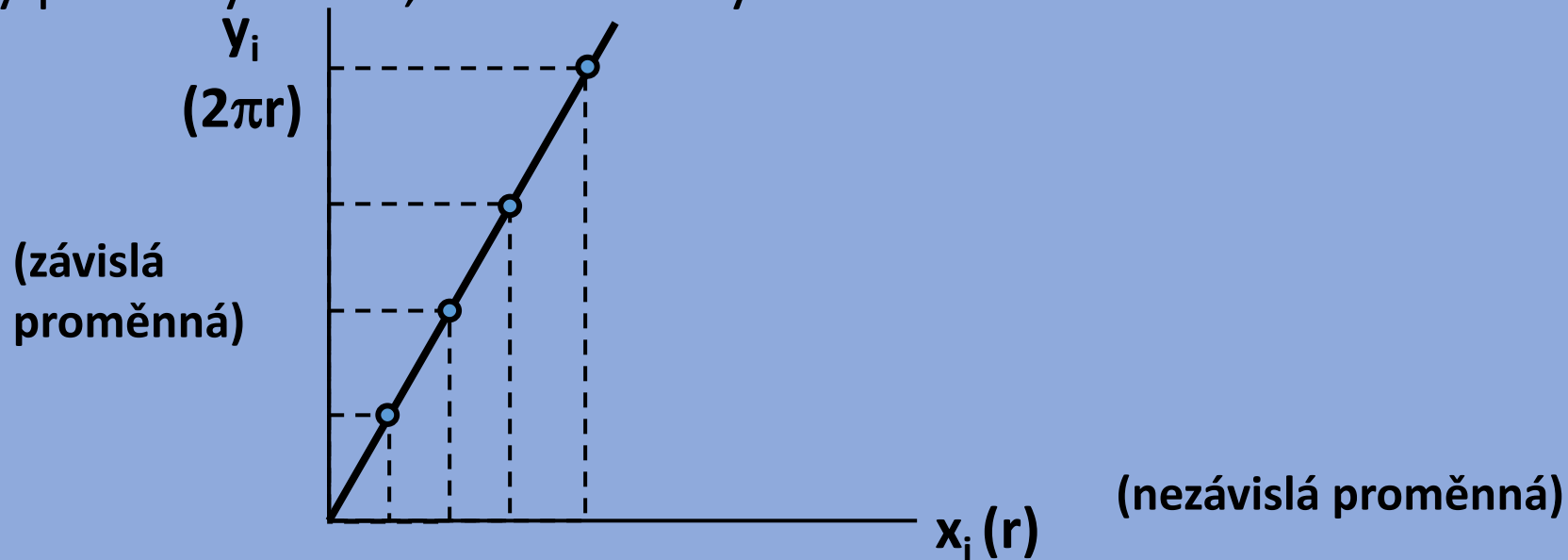
- Hodnocení jedné proměnné v různých souborech
- Vzájemné porovnávání souboru

Vícerozměrná statistika

- Závislost mezi dvěma a více proměnnými v jednom souboru
- Vyjádření a popis vzájemného vztahu mezi proměnnými

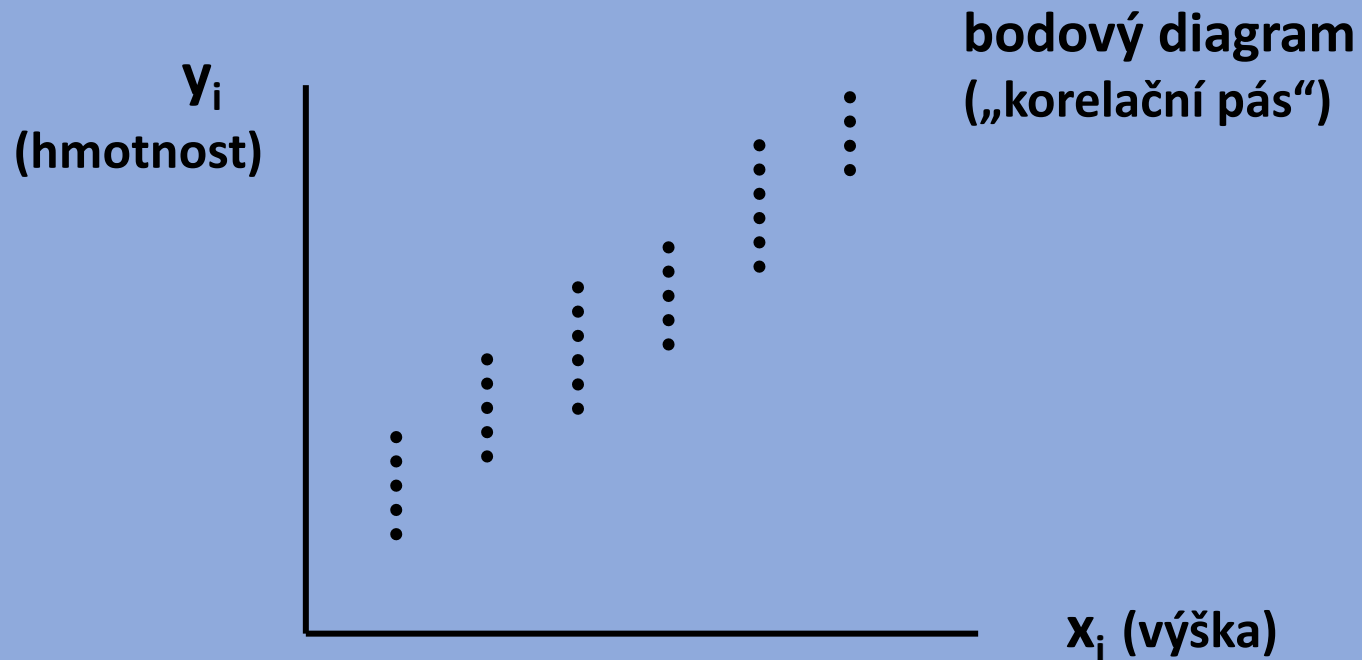
Funkční závislost

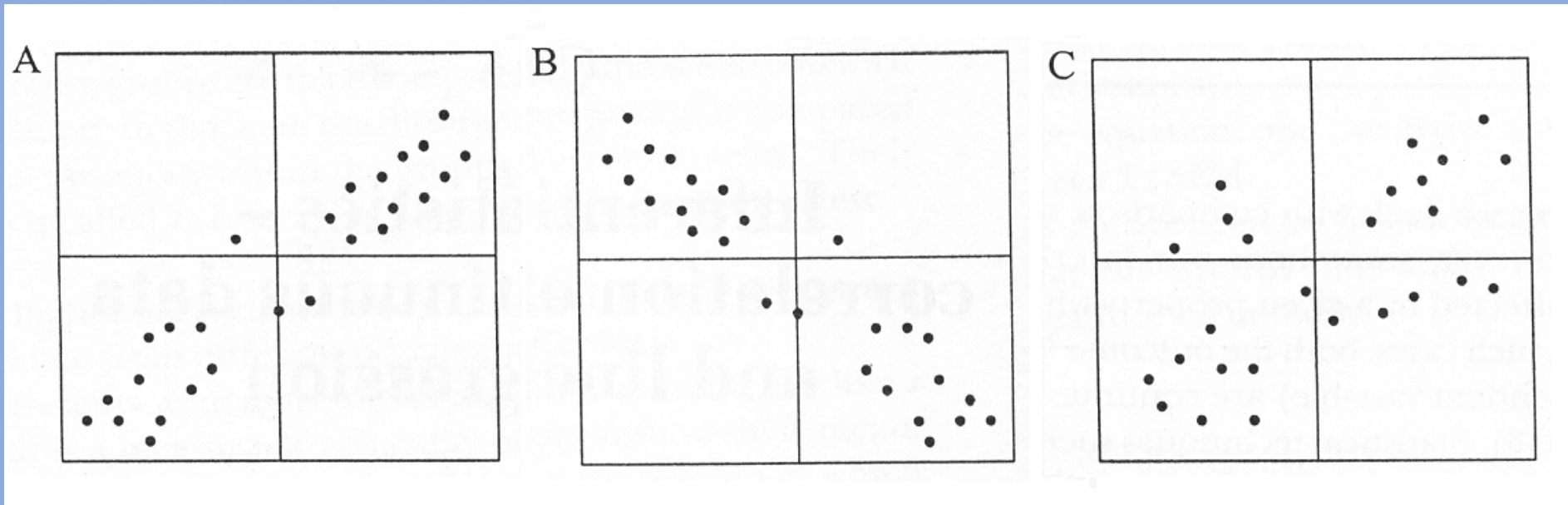
- Matematika, fyzika
- Každé číselné hodnotě jednoho znaku odpovídá jedna přesná hodnota druhého znaku
- Přesný popis rovnicí/vzorcem – např. vztah mezi poloměrem kruhu a jeho obvodem $y=2\pi r$
- Pevný příčinný vztah, neovlivněný náhodou



Korelační závislost

- Biologie, medicína
- Jedné číselné hodnotě prvního znaku odpovídá více náhodných hodnot druhého znaku
- Volná závislost – změna 1. znaku vyvolá změnu 2. znaku jen s určitou pravděpodobností





- Popis a charakteristika korelační závislosti v biologii
- Odhadování nejbližší funkční závislosti (ke které se korelační závislost blíží) - aproximace
- Funkční závislost vyjádříme rovnicí.

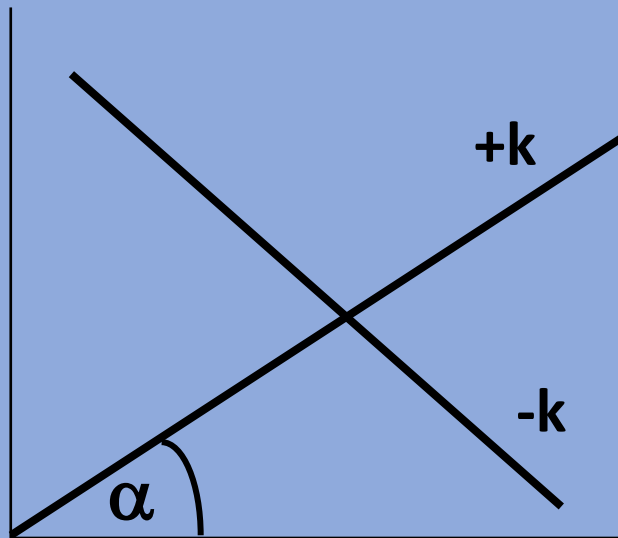
Typy funkčních závislostí:

1. Lineární závislost:

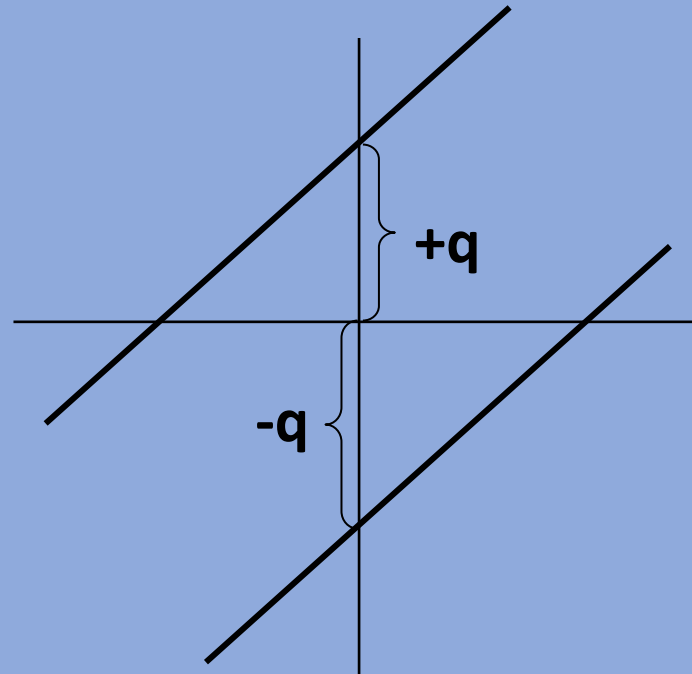
$$y = kx + q$$

k – směrnice přímky

($=\tan \alpha$; sklon přímky)

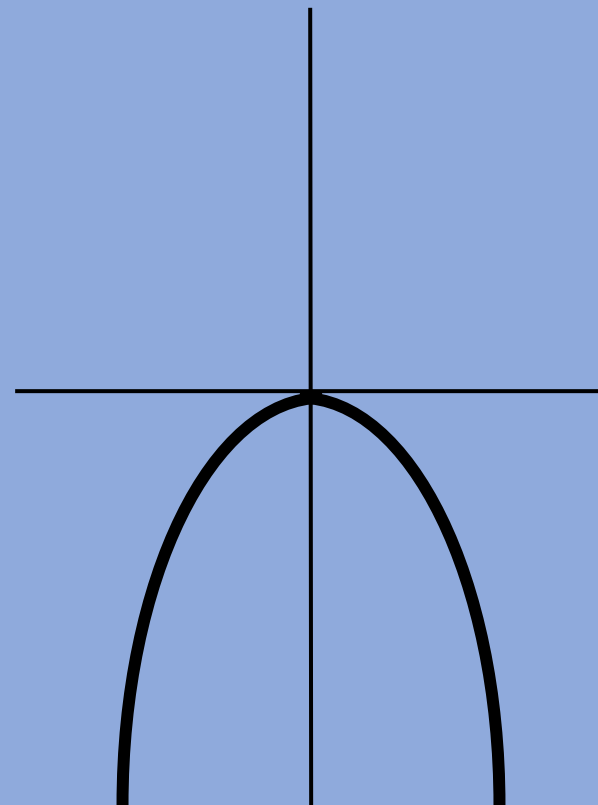
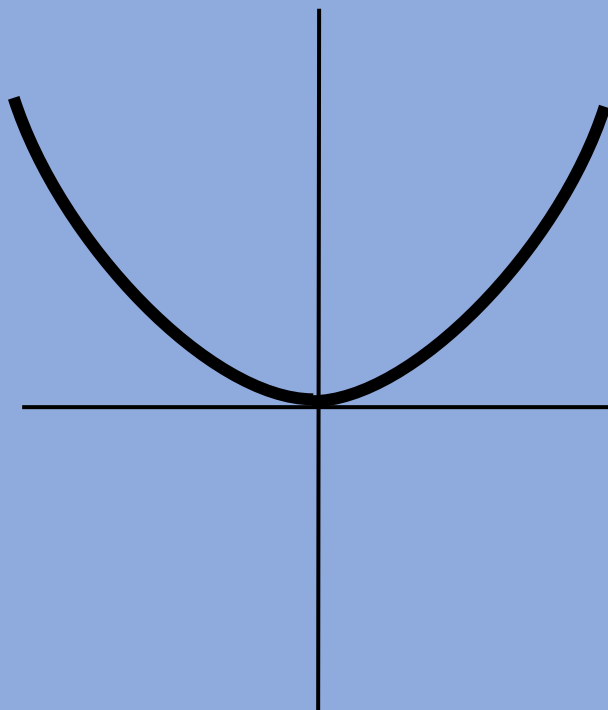


q – posun přímky na ose y



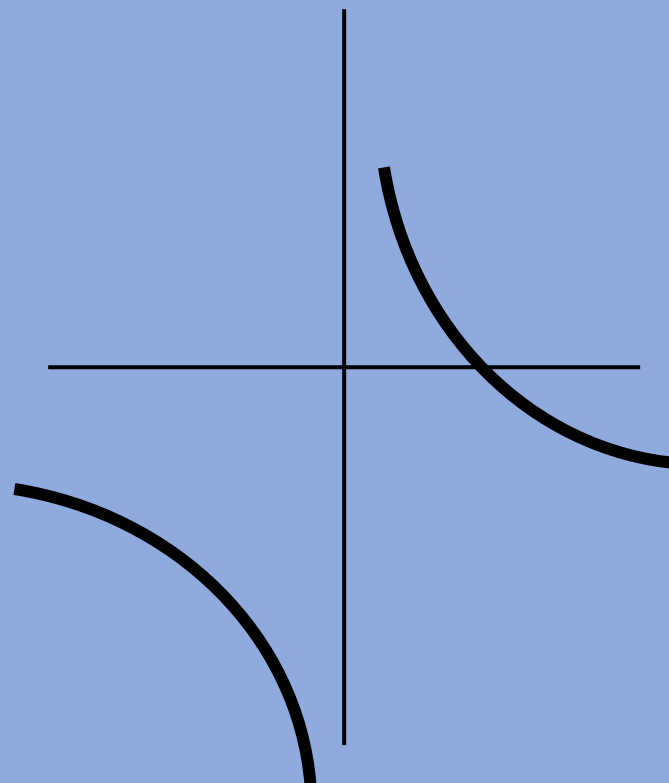
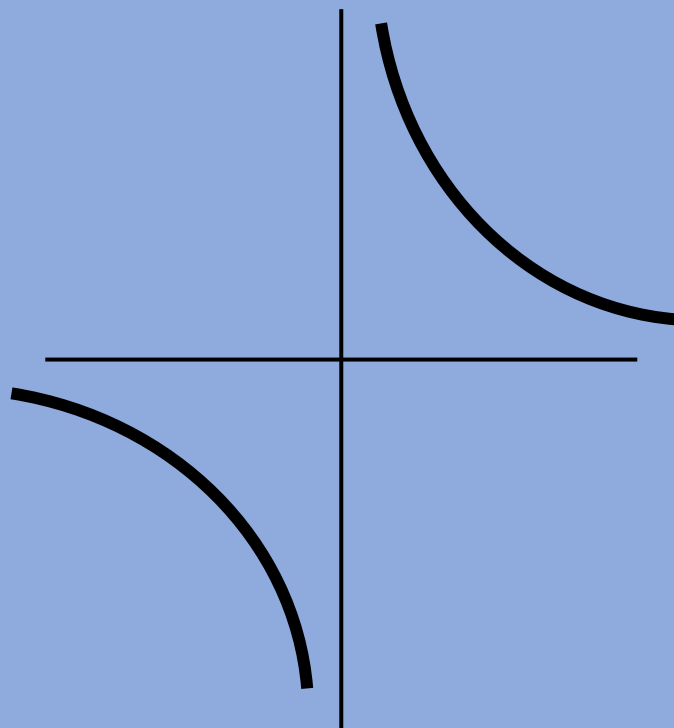
2. Kvadratická (parabolická) závislost:

$$y = ax^2 + bx + c$$



3. Hyperbolická závislost:

$$y = \frac{b}{x} + c$$



exponenciální ($y=a^x$), logaritmická ($y=\log x$)

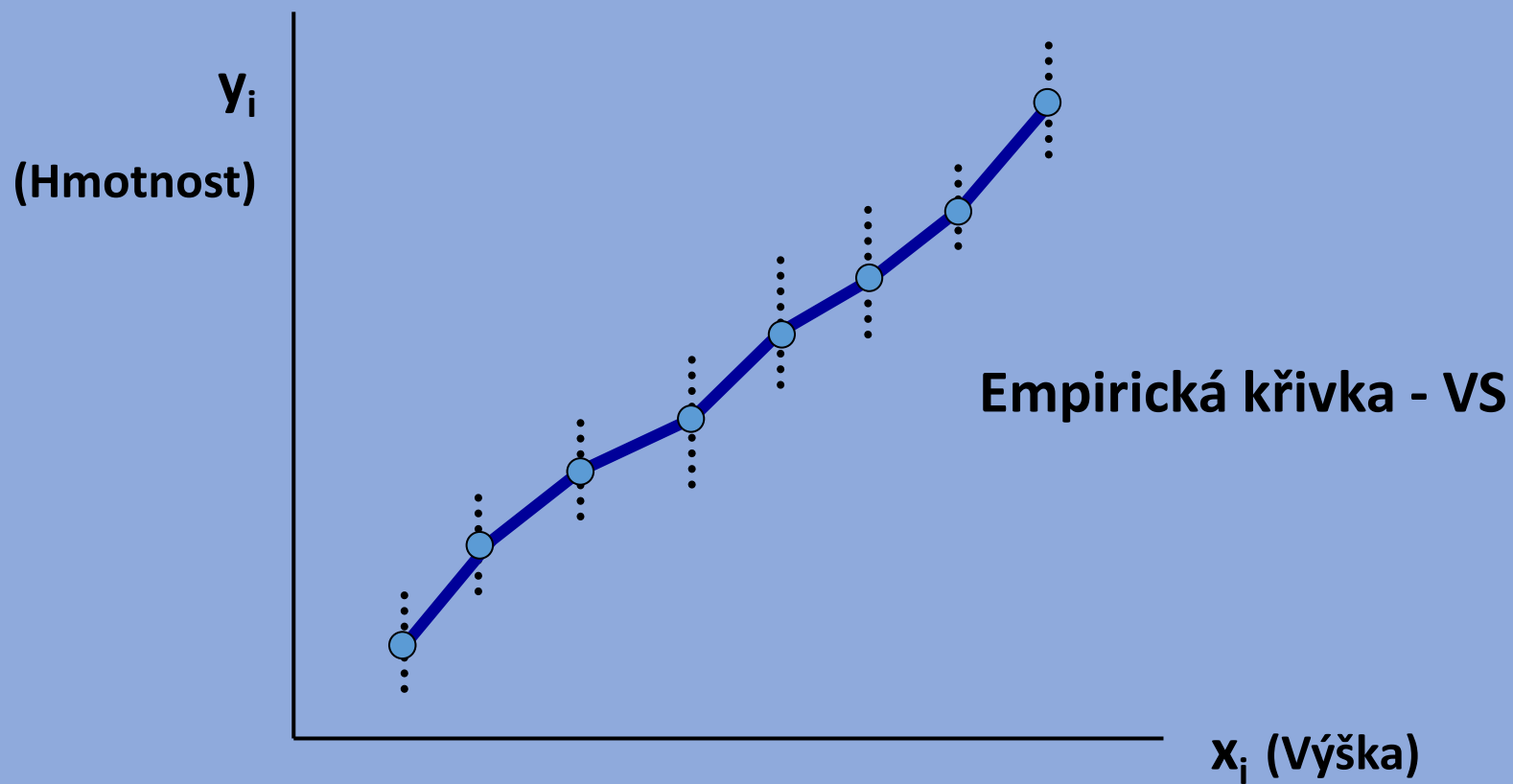
Odhadování funkční závislosti

Bodový diagram – podle rozložení bodů

- Lineární závislost
- Nelineární závislost

Lineární korelace

- Empirická křivka: pro opakované měření v jednom bodě získáme několik hodnot a vypočítáme jejich průměr



Empirická křivka – popisuje závislost na úrovni VS –
odhad skutečné závislosti v populaci)

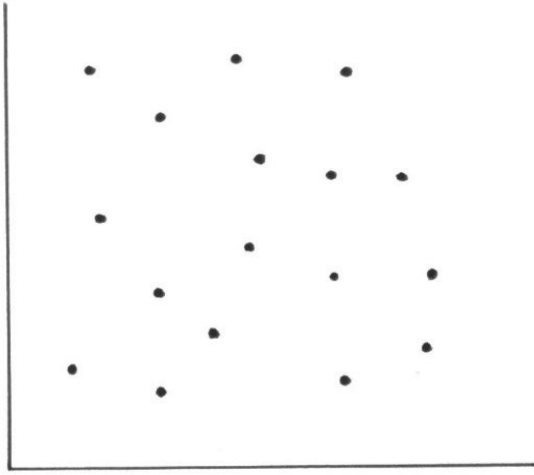
Korelační analýza

- Kvantifikace vztahu dvou proměnných – vyjádření síly závislosti

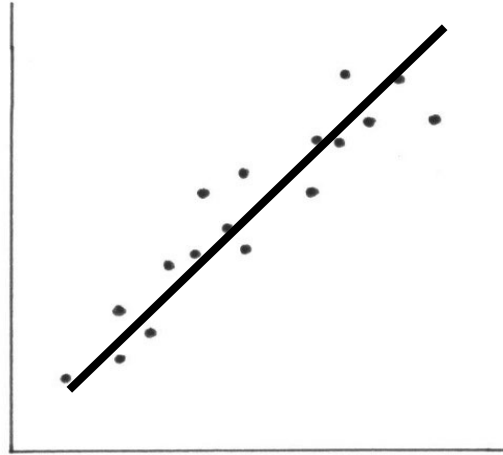
Výpočet korelačního koeficientu

- Pearsonův korelační koeficient – parametrický
- Spearmanův koeficient pořadové korelace - neparametrický

$$r = \langle -1; +1 \rangle$$

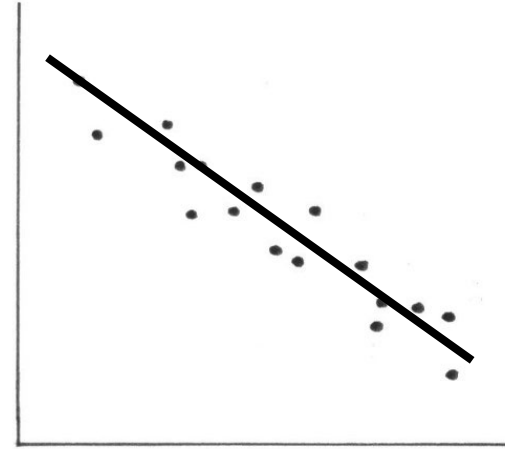


$r = 0$



$r > 0$

**přímá závislost
(pozitivní korelace)**



$r < 0$

**nepřímá závislost
(negativní korelace)**

$r = +1$

**závislost úplná
(funkční)**

$r = -1$

**závislost úplná
(funkční)**

Statistická významnost korelačního koeficientu

Pokud je vypočítaná hodnota pravděpodobnosti

- $p \geq 0,05 \Rightarrow$ statisticky nevýznamný vztah
- $p < 0,05 \Rightarrow$ statisticky významný vztah
- $p < 0,01 \Rightarrow$ statisticky vysoce významný vztah mezi proměnnými

Spearmanův koeficient pořadové korelace

- Neparametrická metoda – nevyžaduje normální rozdělení dat
- Místo naměřených hodnot využívá pro další analýzu jen jejich pořadová čísla
- Lze ho použít u jakékoliv korelace lineární i nelineární

Příklad:

Byl sledován vzájemný vztah mezi hmotností kuřat (kg) a hmotností Fabriciovy burzy (g) u výběrového souboru 10 kuřat.

Existuje vztah mezi těmito proměnnými?

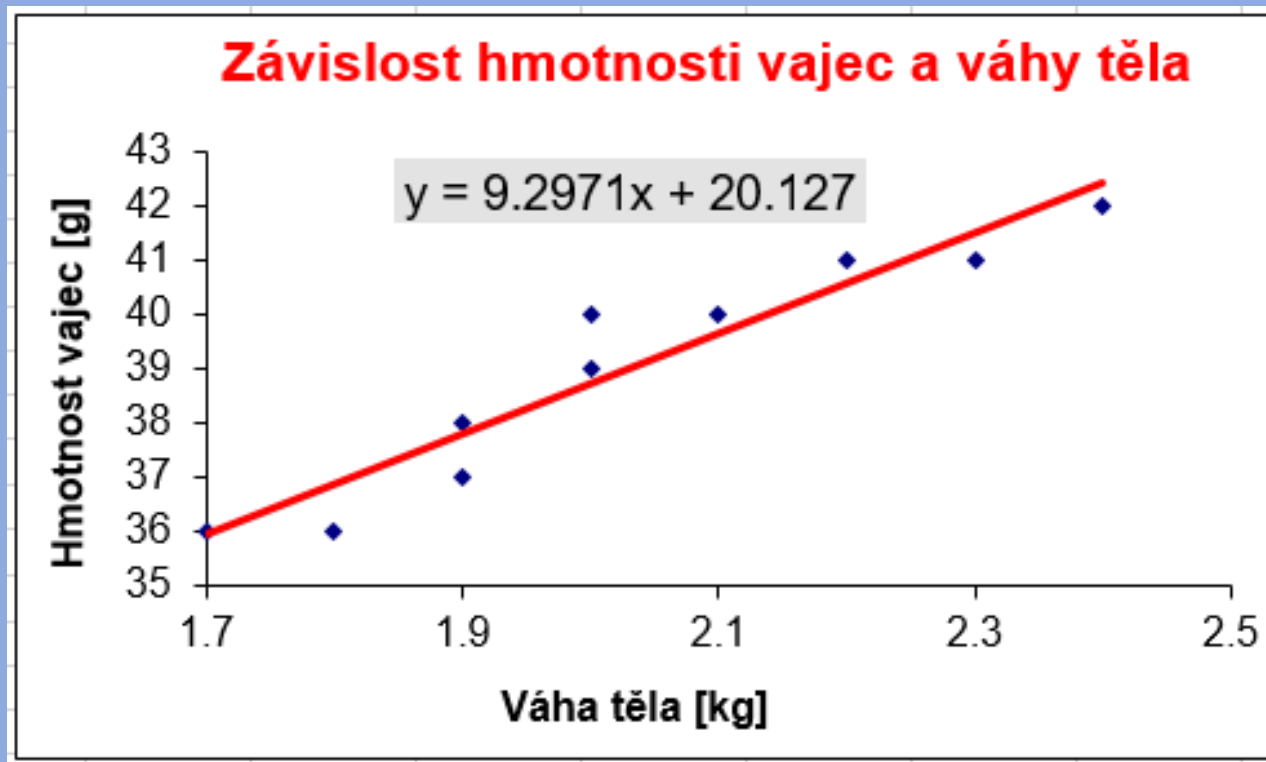
Hmotnost kuřat: 2,2; 1,8 ; 2,1 ; 1,7 ; 2,4 ; 2,0 ; 2,0 ; 1,9 ; 2,3 ; 1,9

Hmotnost Fabriciovy burzy: 41 ; 36 ; 40 ; 36 ; 42 ; 39 ; 40 ; 37 ; 41 ; 38 ; 41

Regresní analýza – Excel:

výběr nejvhodnější funkce pro grafické zobrazení vztahu

Menu: Vložení – Bodový graf (Přidat spojnici trendu)



Možnosti spojnice trendu

- ☐ Exponenciální
- ☒ Lineární
- ☐ Logaritmická
- ☐ Polynomičká Pořadí 2
- ☐ Mocninná
- ☐ Klouzavý průměr Perioda 2

Název spojnice trendu

- ☒ Automaticky Lineární (Řady1)
- ☐ Vlastní

Odhad

Dopředu 0.0 období

Dozadu 0.0 období

☐ Hodnota Y 0.0

☒ Zobrazit rovnici v grafu

☐ Zobrazit hodnotu spolehlivosti R

Korelační analýza – Excel:

Vložit funkci

Vyhledat funkci:

Zadejte stručný popis požadované činnosti a potom klikněte na tlačítko Přejít.

Vybrat kategorii: Statistické

Vybrat funkci:

- BINOM.INV
- CONFIDENCE.NORM
- CONFIDENCE.T
- CORREL**
- COUNTBLANK
- COUNTIF
- COUNTIFS

CORREL(matice1;matice2)
Vrátí korelační koeficient mezi dvěma množinami dat.

[Nápověda k této funkci](#)

Argumenty funkce

CORREL

Matice1	B2:B11	= {2.2;1.8;2.1;1.7;2.4;2.2;1.9;2.3;1.9}
Matice2	C2:C11	= {41;36;40;36;42;39;40;37;41;38}
		= 0.952664357

Vrátí korelační koeficient mezi dvěma množinami dat.

Matice2 je druhá oblast buněk s hodnotami. Hodnoty mohou být čísla, názvy, matice nebo odkazy obsahující čísla.

Výsledek = 0.952664357

[Nápověda k této funkci](#)

OK Storno

Unistat (stat. významnost korelace):

Menu: Unistat – Statistika 1 – Korelační koeficienty

Korelační koeficienty: Krok 2

Vše

Zádný

Zatrhnete možnosti výstupu

☒ Pearson

☒ Přehled celkových statistik

☒ Spearmanovo pořadí

☐ Kendallovo pořadí

☐ Kindalovo pořadí

☐ Spojitou a dichoto

☐ Point Biserial (pop

Nápověda

Korelační koeficienty

Pro Váha těla [kg] a Hmotnost vajec [g]

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka
Váha těla [kg]	10	0	2.0300	0.2214
Hmotnost vajec [g]	10	0	39.0000	2.1602
Párový	10	0		

	Korelační koeficient	Stupně volnosti	Testovací statistika	Pravděp.
Pearson	0.9527	8	8.8629	0.0000
Spearmanovo pořadí	0.9815	8	14.5169	0.0000

Regresní a korelační analýza

hmotnost kuřat v kg a hmotnost Fabriziovy burzy v g		
kuře č.	hmotnost kuřat	hmotnost Fabriziovy burzy
1	2,2	41
2	1,8	36
3	2,1	40
4	1,7	36
5	2,4	42
6	2	39
7	2	40
8	1,9	37
9	2,3	41
10	1,9	38
průměr	2,03	39
SD	0,221359436	2,160246899

Korelační analýza

Excel

Unistat

Pearsonův korelační koeficient-data mají normální rozdělení

$r = 0,952664357$ $r > 0$, blíží se 1-silná pozitivní přímá/pozitivní korelace

$r = 0,952664357$

Ověření statistické významnosti korelačního koeficientu

$p = 2,07418E-05$ $p < 0.01$ - statisticky vysoce významná korelace

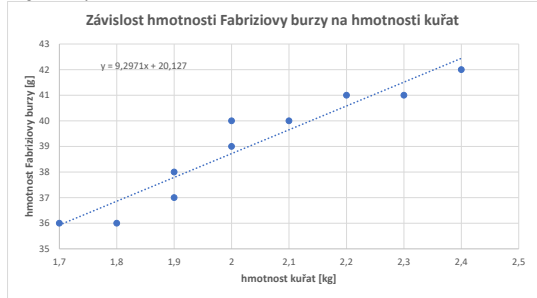
Spearmanův koeficient pořadové korelace-data nemají normální rozdělení

$r_s = 0,981543108$ $r > 0$, blíží se 1-silná pozitivní přímá/pozitivní korelace

Ověření statistické významnosti korelačního koeficientu

$p = 4,97E-07$ $p < 0.01$ - statisticky vysoce významná korelace

Regresní analýza



Korelační koeficienty

Pro hmotnost kuřat a hmotnost Fabriciovy burzy

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka
hmotnost kuřat	10	0	2,0300	0,2214
hmotnost Fabriciovy burzy	10	0	39,0000	2,1602
Párový	10	0		

	Korelační koeficient	Stupně volnosti	* Testovací statistika	jednostranná pravděp.	dvoustranná pravděpodobnost	Dolní 95%	Horní 95%
Pearson	0,9527	8	8,8629	0,0000	0,0000 p<0,01	0,8072	0,9890

* Z-statistika pro Kendallovo pořadí, t-statistika jinak

Korelační koeficienty

Pro hmotnost kuřat a hmotnost Fabriciovy burzy

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka
hmotnost kuřat	10	0	2,0300	0,2214
hmotnost Fabriciovy burzy	10	0	39,0000	2,1602
Párový	10	0		

	Korelační koeficient	Stupně volnosti	* Testovací statistika	jednostranná pravděp.	dvoustranná pravděpodobnost	Dolní 95%	Horní 95%
Spearmanovo pořadí	0,9815	8	14,5169	0,0000	0,0000	0,9213	0,9958
					p<0,01		

* Z-statistika pro Kendallovo pořadí, t-statistika jinak



Financováno
Evropskou unií
NextGenerationEU

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Národní
plán
obnovy

Bakalářská práce I – Metodologie a biostatistika

Kvalitativní data



Kvalitativní data

- Popisujeme je četností výskytu v souboru dat
- **Binární:** mohou nabývat pouze dvou hodnot, mortalita (úhyn – přežití), vyšetření na parazity (pozitivní – negativní)
- **Nominální:** obsahují více kategorií, které nelze vzájemně seřadit, krevní skupiny (A/B/AB/0)
- **Ordinální:** obsahují více kategorií, na rozdíl od nominálních dat je však lze vzájemně seřadit, stádium maligního onemocnění (I/II/III/IV)

χ^2 test - kontingenční tabulky

- U kvalitativních dat pracujeme s četností výskytu jednotlivých znaků
- K tomu využíváme kontingenční tabulky 2x2 nebo k x m
- Tabulka 2x2 slouží k posouzení závislosti dvou binárních znaků (samci a samice, pozitivní a negativní)
- Tabulka k x m slouží k posouzení více znaků

Kontingenční tabulka 2 x 2

A – aplikace vakcíny (onemocnění A)

B – výskyt onemocnění (onemocnění B)

	B	B'	celkem
A	a	b	a + b
A'	c	d	c + d
celkem	a + c	b + d	n = a + b + c + d

$$\chi^2 = \frac{n (ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$\underline{\nu} = (k - 1) (m - 1) = \underline{1}$$

Je-li $\chi^2 > \chi^2_{\text{krit.}} \Rightarrow$ zamítáme hypotézu nezávislosti

Kontingenční tabulky

2 skupiny– k (i) 2 třídy– m (j)

$\begin{matrix} m \\ k \end{matrix}$	Úhyn	Přežití	Skup Σ (s_i)
A	4 (10,16)	71 (64,84)	75
A'	17 (10,84)	63 (69,16)	80
Tř. Σ (t_j)	21	134	155(n)

n_e – empirické četn.

n_o – teoretické četn.

$$n_{oij} = \frac{s_i \cdot t_j}{n}$$

$$\chi^2 = \frac{155 \cdot (4 \cdot 63 - 71 \cdot 17)^2}{(4 + 71) \cdot (17 + 63) \cdot (4 + 17) \cdot (71 + 63)} = 8,373$$

$$\chi^2 = \sum_{i,j} \frac{(n_{ei} - n_{oi})^2}{n_{oi}} = \frac{(4 - 10,16)^2}{10,16} + \frac{(71 - 64,84)^2}{64,84} + \dots = \underline{8,373}$$

$$\chi^2_{0,05} = 3,841$$

$$\chi^2_{0,01} = 6,635 \Rightarrow \text{významnost: } p < 0,01$$

Příklad:

Byla sledována závislost výskytu některých chorob na krevních skupinách. Výsledky sledování u souboru 8766 osob jsou shrnuty v následující tabulce. Zjistěte, zda se liší frekvence výskytu sledovaných chorob u osob s různými krevními skupinami.

Kontingenční tabulka kxm

A_1-A_k
 B_1-B_m

$\begin{matrix} m \\ k \end{matrix}$	Žal.vředy	Rakovina žaludku	Kontrola	Celkem (a_i)
0	983 (872,39)	383 (428,91)	2892 (2956,7)	4258
A	679 (762,16)	416 (374,72)	2625 (2583,1)	3720
B	134 (161,45)	84 (79,38)	570 (574,18)	788
Celk. (b_j)	1796	833	6087	n=8766

n_e – empirické četn.

n_o – teoretické četn.

$$n_{oij} = \frac{a_i \cdot b_j}{n}$$

$$v=(k-1)(m-1)$$

$$\chi^2 = \sum_{i,j} \frac{(n_e - n_o)^2}{n_o} = \frac{(983 - 872,39)^2}{872,39} + \dots = \underline{40,54}$$

$$v = 4$$

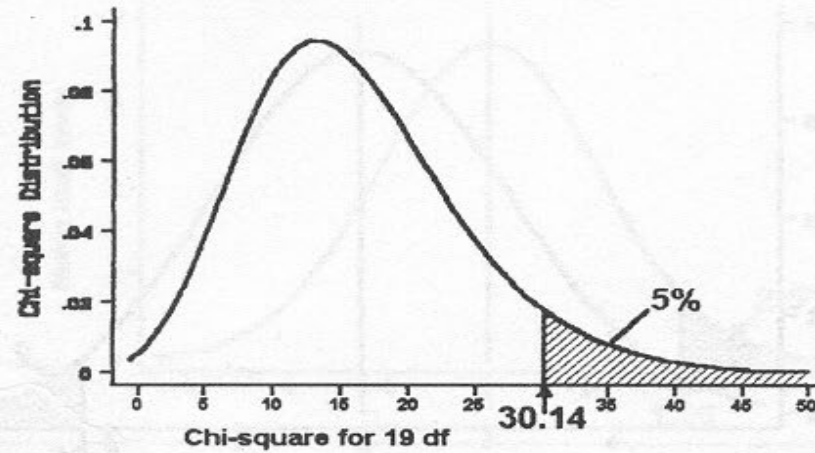
$$\chi^2_{0,01} = 13,28$$

++

	vředy	rakovina	kontrola
0	14,02	(-) 4,91	(-) 1,42
A	(-) 9,07	4,55	0,68
B	(-) 4,67	0,27	(-) 0,95

znam. + - : podle rozdílu $n_e - n_o$

Chi-Square Distribution, Right Tail^a



α (area in right tail)	.10	.05	.025	.01	.005	.001	.0005
$1-\alpha$ (except right tail)	.90	.95	.975	.99	.995	.999	.9995
df= 1	2.71	3.84	5.02	6.63	7.88	10.81	12.13
2	4.61	5.99	7.38	9.21	10.60	13.80	15.21
3	6.25	7.81	9.35	11.34	12.84	16.26	17.75
4	7.78	9.49	11.14	13.28	14.86	18.46	20.04
5	9.24	11.07	12.83	15.08	16.75	20.52	22.15
6	10.64	12.59	14.45	16.81	18.54	22.46	24.08
7	12.02	14.07	16.01	18.47	20.28	24.35	26.02
8	13.36	15.51	17.53	20.09	21.95	26.10	27.86
9	14.68	16.92	19.02	21.67	23.59	27.86	29.71
10	15.99	18.31	20.48	23.21	25.19	29.58	31.46
11	17.28	19.68	21.92	24.72	26.75	31.29	33.13
12	18.55	21.03	23.34	26.22	28.30	32.92	34.80
13	19.81	22.36	24.74	27.69	29.82	34.54	36.47
14	21.06	23.69	26.12	29.14	31.32	36.12	38.14
15	22.31	25.00	27.49	30.57	32.81	37.71	39.73
16	23.54	26.30	28.84	32.00	34.27	39.24	41.31

Kontingenční tabulka k x m

Při vyšetření masa zvěře na přítomnost parazitů (trichinella) byl sledován počet pozitivních a negativních vzorků ze 3 lokalit. Liší se četnost výskytu pozitivních a negativních vzorků mezi jednotlivými lokalitami?

	Negativní	Pozitivní
Lokalita A	96	25
Lokalita B	121	22
Lokalita C	89	16

UNISTAT - Kontingenční tabulky

Menu: Unistat – Statistika1 – Tabulky – Kontingenční tabulka

	A	B	C	D	E	F	G	H	I	J
1		Negativní	Pozitivní							
2	A	96	25							
3	B	121	22							
4	C	89	16							
5										
6										
7										
8										
9										
10										
11										
12										

Kontingenční tabulka: Krok 1

Seznam proměnných

Vybrat / Vynechat

Proměnná j

Vybrané proměnné

C1 Negativní
C2 Pozitivní

Kontingenční tabulka: Krok 2

Vše Zádný Vyberte volby pro výstup nebo stiskněte na tlač. 'Opt' pro další možnosti

0 Skóry: 0: Tabulka, 1: Pořadí, 2: Ridit, 3: Modifikovaný ridit

	A	B	C	D
1	Kontingenční tabulka			
2	3 Řádky x 2 Sloupce			
3	Chi-kvadrát testy			
4		Statistika Chi-kvadrát	Stupně volnosti	Pravostr. pravděp.
5	Pearson	1.6379	2	0.4409
6	Věrohodnostní poměr	1.5983	2	0.4497
7	+ Yatesova korekce			
8	# Lineární /Lineární	1.2284	1	0.2677
9	~ McNemar-Bowker			
10				

Tabulky

R x C tabulkové statistiky

2 x 2 tabulkové statistiky

Analýza rozvrstvenosti

da Storno ≤ Zpět Další ≥ Dokončit

$p > 0,05$ nevýznamný (-)

$p < 0,05$ významný (+)

$p < 0,01$ vysoce významný (++)

UNISTAT - Kontingenční tabulky

Menu: Unistat – Statistika1 – Tabulky – Kontingenční tabulka

The screenshot displays the UNISTAT software interface for contingency table analysis. It includes a data table, two dialog boxes for selecting variables and output options, and a results window with statistical tests and their p-values.

Data Table:

	Žal.vředy	Rakovina žal.	Kontrola
sk.O	983	383	2892
sk.A	679	416	2625
sk.B	134	84	570

Kontingenční tabulka: Krok 1

Seznam proměnných: []

Vybrat / Vynechat: [Proměnná j]

Vybrané proměnné: C1 Žal.vředy, C2 Rakovina žal., C3 Kontrola

Kontingenční tabulka: Krok 2

Vše | Zádný | Vyberte volby pro výstup nebo stiskněte na tlač. 'Opt' pro další možnosti

0 Skóry: 0: Tabulka, 1: Pořadí, 2: Ridit, 3: Modifikovaný ridit

Kontingenční tabulka

3 Řádky x 3 Sloupce

Chi-kvadrát testy

	Statistika Chi-kvadrát	Stupně volnosti	Pravost. pravděp.
Pearson	40.5434	4	0.0000
Věrohodnostní poměr	40.6401	4	0.0000
+ Yatesova korekce			
# Lineární /Lineární	21.0035	1	0.0000
McNemar-Bowker	4979.6552	3	0.0000

+ Vykazováno pro 2 x 2 tabulky.

Tabulky

R x C tabulkové statistiky

2 x 2 tabulkové statistiky

Analýza rozvrstvenosti

da | Storno | ≤ Zpět | Další ≥ | Dokončit