C V I Č E N Í ZE STATISTIKY PRO BIOLOGY:

S B Í R K A P Ř Í K L A D Ů (VERZE 1.3)

Martin Duchoslav

Olomouc 2004

- Předložený text reprezentuje výběr příkladů, které doplňují přednášky a cvičení kurzu *Základy biostatistiky* pro odborné a učitelské studium biologie a OTŽP na PřF UP Olomouc. Příklady mají sloužit studentům pro doplňkové procvičování základních metod vyučovaných v kurzu a zároveň slouží jako vzory pro příklady, které se objeví v testech. Část příkladů byla kompilována z různých zdrojů.
- Pro příslušné testy volte, prosím, hladinu významnosti 0,05.
- Všechny výsledky v tomto textu byly spočítány statistickým programem NCSS 2000, popř. Statgraphics for Win 4.0.
- Naleznete-li v textu zadání či ve výsledcích chyby a nesrovnalosti, prosím napište mi je a já to napravím. Díky i za konstruktivní připomínky.

Příklady:

1)
Ve vzorku náhodně vybraných zdravých mužů 20-30letých byly naměřeny tyto hodnoty hemoglobinu (mmol/l):

9,1 8,4 10,2 9,7 8,9 9,1 9,4 9,5 9,3 9,5 9,7 10,2

spočítejte výběrové odhady pro: a) 25. percentil, 75. percentil, medián; b) aritmetický průměr; c) směrodatnou odchylku; d) variační koeficient

2)
Chceme zjistit, zda-li se naše krátkodobé výsledky měření znečištění ovzduší liší od dlouhodobých údajů. Při našem měření o rozsahu n=57 byl průměr 36,6 a směrodatná odchylka 15,44. Dlouhodobá průměrná hodnota znečištění této oblasti byla 50,0.

Je námi zjištěná nižší hodnota pouze nahodilá, nebo opravdu došlo k významné změně průměrného znečištění? Předpokládáme normální rozložení dat. Otestujte.

3)
Byly změřeny následující hodnoty IQ u náhodně vybraných 10-letých chlapců a dívek:

chlapci: 113, 115, 103, 80, 92, 109, 109, 128, 117, 88, 103, 100

dívky: 94, 101, 109, 116, 128, 100, 75, 75, 123, 82, 123, 94, 92.

Zjistěte na základě těchto údajů, zda je rozdíl v inteligenci pohlaví.

4)
Ve studované populaci hraboše polního se za poslední měsíc narodilo 89 samců a 99 samic. Byl poměr průkazně odlišný od očekávaného 1:1?

5)
Na náhodně vybraných 10 rostlinách vstavače nachového jsme měřili výšku rostliny (cm) a průměr růžice (cm):

výška: 10, 10, 14, 15, 9, 11, 14, 18, 10, 11

průměr růžice: 10, 20, 22, 25, 26, 27, 20, 22, 23, 10

a) Spočítejte základní charakteristiky polohy a variability znaku pro obě charakteristiky (průměr, medián, standardní odchylku).
b) Existuje vztah mezi výškou rostliny a průměrem růžice? Otestujte.

6)
Homogenní pletivo bylo rozděleno na 20 vzorků a ty byly po desíti odeslány do dvou laboratoří ke stanovení obsahu dusíku. Výběrový rozptyl první laboratoře byl 3,82 a druhé 1,04. Jsou obě laboratoře stejně kvalitní? Otestujte.

7)
Byl zkoumán vztah mezi váhou těla a kapacitou plic (měřeno na spirometru). Studovali jsme skupinu 10 náhodně vybraných žen mezi 17 a 19 lety:

Subjekt:	1	2	3	4	5	6	7	8	9	10
Hmotnost (v kg):	54,4	56,2	49,0	63,5	60,8	59,9	62,6	62,1	52,2	50,8
Objem (v l):	3,87	3,26	2,14	4,13	3,44	2,78	2,91	3,33	3,2	2,17

Existuje závislost mezi proměnnými? Vyneste tyto údaje do diagramu, zjistěte typ závislosti, zvolte příslušnou metodu a otestujte včetně rozhodnutí o H_0.

8)
Včely jsou postupně vypouštěny do pokusného prostoru se žlutými, červenými a modrými terči. Sledujeme barvu terče, na který včela poprvé usedne. Vypustili jsme 100 včel. Získali jsme tato data - četnosti barev: žlutá 47, červená 38, modrá 15. Lze z těchto dat usoudit, že včely některou barvu preferovaly? Otestujte.

9)
V určité definované populaci dětí byla zjištěna průměrná výška 123 cm a směrodatná odchylka 4 cm. Předpokládejme, že rozdělení výšek v této populaci je normální. Jaká část populace dětí má výšky v rozmezí od 119 do 127; menší než 119 cm a větší než 127 cm? Uveďte v procentech.

10)
Studovali jsme vliv léku na tepovou frekvenci 10 zdravých osob měřených v klidu před a po podání léku.

Počty tepů/min. byly:

před podáním: 68 65 76 70 79 69 77 80 70 72

po podání: 69 69 79 69 83 69 80 83 70 73

Liší se tepová frekvence před a po podání léku? Otestujte.

11)
Byl studován vztah mezi znečištěním řeky způsobených papírnami a přítomností mihulí. Z celkem zkoumaných náhodně vybraných 166 toků byly mihule přítomné ve 54 řekách bez znečištění a v 82 řekách se znečištěním (= přítomnost papírny). Ve 20 znečištěných řekách nebyly mihule zaznamenány. Existuje souvislost mezi znečištěním a výskytem mihulí? Otestujte.

12)
Při antropologickém měření obyvatelstva byla mimo jiné studována šířka nosu u 20-letých mužů dané oblasti. U náhodně vybraných 10 mužů byly zjištěny tyto hodnoty:

3,6; 4,1; 3,3; 3,4; 3,7; 3,1; 4,0; 4,0; 3,6; 3,0

Stanovte 95% interval spolehlivosti (konfidenční interval) průměru.

13)
Zaokrouhlete na 2 platné číslice:

1,250

1,15

1,151

0,005550

15213

14)
Byl vyšetřován vliv 3 druhů penicilínů na růst kolonií *Baccilus subtilis*. Jednotlivé hodnoty uvádějí průměrnou velikost kolonií na příslušné plotně. Na 5 náhodně vybraných ploten byl aplikován penicilin 1, podobně i pro peniciliny 2 a 3. Plotny byly na začátku pokusu umístěny náhodně do růstové komory a testovány po uplynutí 1 týdne. Výsledek je v následující tabulce.

Druh penicilinu	Měření				
1	10.6	8.5	9.8	8.3	8.1
2	7.3	9.1	8.4	8.8	7.6
3	8.2	7.7	8.0	7.2	6.4

Liší se účinky různých druhů penicilínu na růst kolonií? Otestujte.

15)
Bylo vybráno 8 hospodářství, na každém hospodářství vždy bylo jedno náhodně vybrané pole rozděleno na 2 poloviny a na příslušné (náhodně vybrané) poloviny vysety po jedné ze dvou odrůd pšenice ve stejných hustotách. Úlohou je zjistit, zda se výnosy pšenice liší na konci pokusu (sklizeň). Otestujte. Pomoc: *předběžné testy ukazují, že data nemají normální rozdělení!*

Hospodářství	Výnosy odrůdy A (q per ha)	Výnosy odrůdy B (q per ha)
1	35	33
2	46	44
3	50	53
4	40	43
5	55	57

6	38	36
7	43	46
8	52	50

16)
Domníváme se, že myšice na dvou blízkých ostrovech patří k různým rasám. Byl chycen jistý počet jedinců a změřili jsme vybrané biometrické charakteristiky. Délka ocasu (v mm) byla následující:

Ostrov J 101,111,105,121,107,99,103,117,123,100,109,96,106,98,115

Ostrov K 101,106,107,96,97,100,103,100,101,95,102,104,109,93,99,102,101,99,96,98

(a) Spočítejte aritmetický průměr, medián, směrodatnou odchylku a variační koeficient pro oba výběry. (b) Spočítejte 95 % interval spolehlivosti (konfidenční interval) pro rozdíl průměrů. (c) Porovnejte tyto dva výběry. Liší se obě populace v délce ocasu? Otestujte.

17)
Studujeme populaci r. *Drosophila*, která vznikla křížením heterozygotních samic (w/+) s w/y samci a předpokládáme, že recesivní mutovaný gen w (albín) je vázaný na pohlaví. Poměr červenookých a albínů by tedy měl být 1:1. Získali jsme tento výsledek: 768 červenookých a 818 albínů. Je náš předpoklad (hypotéza) správný? Otestujte.

18)
Při studiu vztahu mezi hmotností vegetativní částí rostliny a hmotností sexuálních struktur (květy + semena) jsme získali výsledky z náhodně vybrané skupiny 10 jedinců:

Jedinec	1	2	3	4	5	6	7	8	9	10
Hmotnost veget. č.(*g*)	54	56	49	60	61	58	63	62	52	50
Hmotnost sex. č. (*g*)	3	4	2	7	8	6	10	9	3	2

Závisí hmotnost sexuálních struktur na hmotnosti vegetativních částí rostliny? Otestujte.

19)
Během studia prostorové heterogenity na louce bylo odpozorováno, že přítomnost jetele *Trifolium repens* je pravděpodobně spojena s kypřící aktivitou dešťovek. Pro testování této hypotézy byla umístěna na louku mřížka 10x40 čtverců, každý o ploše 25 cm^2. V každém čtverci byla zaznamenána přítomnost/nepřítomnost jetele a trusu dešťovek. V celkem 400 čtvercích byl jetel přítomen v 260, trus ve 115 a oba (dešťovky a jetel) společně v 90. Otestujte hypotézu náhodnosti výskytu obou objektů.

20)
V experimentu, který zkoumal vliv koncentrace glukózy na lineární růst kolonií *Geotrichum candidum* byly získány výsledky z média obsahujícího glukózu 50 mg/l:

Den od inokulace	3	5	7	9	11	13
Průměr kolonie (mm)	7	13	17	23	26	29

Jaký je vztah (závisí na sobě?) těchto dvou proměnných? Otestujte.

21)
Při experimentu, který měl studovat vliv kvality světla na rychlost fotosyntézy jednoho druhu řasy jsme rozdělili 25 vzorků řasy náhodně do 5 skupin po 5 vzorcích a v rámci příslušné skupiny testovali vždy jednu světelnou délku světla. Získali jsme tyto výsledky (produkce kyslíku v mikrolitrech):

	Opakování (vzorky)				
	1	2	3	4	5
Část spektra					
Modrá	5	5	6	5	5
Zelená	6	5	6	6	6
Žlutá	15	16	16	17	17
Červená	19	19	19	18	17
Bílá	21	21	22	21	20

Otestujte nulovou hypotézu o shodnosti účinku různých částí spektra na intenzitu fotosyntézy.

22)
Předpokládejme, že houbovou chorobou je napadáno 10 jedinců pryšce chvojky z každých 100 jedinců. V naší populaci jsme náhodně prozkoumali 20 jedinců. S jakou pravděpodobností budou mezi prozkoumanými jedinci v naší populaci přítomni (a) žádný, (b) právě 3 a (c) 6 napadených jedinců? (d) Jaká je střední (s nejvyšší pravděpodobností očekávaná) hodnota?

23)
Na 12 ze 24 náhodně vybraných býků byl aplikován přídavek vitamínu B_{12} v krmné směsi, čímž jsme získali pokusný zásah B.

Váhové přírůstky v kg pro standardní směs (A) byly:

27	35	38	37	29	33	37	31	34	32	33	34

Váhové přírůstky po pokusném zásahu B byly:

32	30	36	38	36	43	31	40	36	42	35	40

Liší se způsoby výkrmu býků ve vztahu k jejich přírůstkům? Otestujte.

24)
Průměrná hmotnost dívek narozených v regionu byla dle dlouhodobých měření 3100 g. Ve vzorku 120 dívek, které se narodily matkám jež v průběhu těhotenství kouřily, byla průměrná hmotnost 2900 g a směrodatná odchylka 360 g. Je nižší průměrná hmotnost ve vzorku dívek, které se narodily kuřačkám pouze nahodilá nebo lze očekávat nižší porodní hmotnost v celé populaci dívek, které se rodí kuřačkám? Otestujte.

25)
Na frekventovaném místě ve městě byl sledován ve stejné denní době opakovaně vztah mezi počtem projíždějících aut za hodinu a objem CO v ovzduší.

Počet aut (tisíce /hod)										
1	1,2	1,4	1,5	1,5	2,2	2,4	2,9	3	3,1	3,1

CO (x 10^6) ve stejném měření:

6,5	8,7	7,7	7	11,2	12,2	13,2	20,5	19,2	21,6	20,4

(a) Která ze dvou veličin je závislá, která nezávislá?; (b) Spočítejte rovnici regresní přímky a otestujte model; (c) vypočítejte očekávanou koncentraci CO v ovzduší, když místem projede 2500 aut/hod.

26)
Poměr pohlaví u narozených dětí se dle dlouhodobých výzkumů pohybuje 100 žen:105 mužů. (a) S jakou pravděpodobností bude v našem výběrovém vzorku 30 narozených dětí 10 chlapců? (b) Jaký je očekávaný počet chlapců v našem výběru a jakou má pravděpodobnost?

27)
Předpokládejme, že délka korunních lístků u studovaného druhu rostliny má normální distribuci s aritmetickým průměrem=3,2 cm a s=0,8 cm. Jaká část populace bude mít délku okvětních lístků (a) větší než 4,5 cm?, (b) větší než 1,78 cm?, (c) mezi 1,78 a 4,5 cm?

28)
Při experimentu byly kříženy dvě plemena králíků, čímž jsme získali 27 F_1 hybridů. Provedli jsme inbreeding, čímž jsme získali 112 F_2 králíků. Získali jsme tyto údaje o délce femuru těchto králíků:

generace	n	průměr	standardní odchylka
F1	27	83.39	1.65
F2	112	80.5	3.81

(a) Existuje signifikantně větší obsah variability v délce femuru mezi F_2 hybridy než mezi F_1 hybridy? (b) Jaký dobře známý genetický fenomén ilustrují tyto data? (mimo soutěž).

29)
V našem pokusu jsme sledovali vliv infekce virem na teplotu králíků v různých časech od podání viru (viz tabulka).

čas od injekce viru (hodiny)	teplota (F)
24	102.8
32	104.5
48	106.5
56	107
72	103.9
80	103.2
96	103.1

(a) spočítejte průměr, medián, standardní odchylku, variační koeficient a mezikvartilové rozpětí pro teplotu; (b) vyneste data do grafu (stačí od ruky); (c) spočítejte regresní rovnici přímky a otestujte H_0: b=0; (e) je v datech nějaký problém a pokud ano, co byste doporučovali?

30)
Spočítejte 95 % konfidenční interval pro střední hodnotu (= aritmetický průměr). Je dáno: n=2666, aritmetický průměr=79,73; s=10,94.

31)
Rozložení krevních buněk v komůrkách hematocytometru dosahuje průměrné hodnoty 1,8 buňky na komůrku. Spočítejte relativní očekávané frekvence pro komůrky (a) bez žádné buňky, (b) pro vzorky s počtem buněk ≤ 2, (c) pro vzorky s počtem buněk > 2. Jaké jsou absolutní očekávané četnosti v případě výběrového vzorku n = 400 komůrek a výběrového průměru 1,8 buňky na komůrku pro komůrky (d) bez žádné buňky, (e) pro vzorky s počtem buněk ≤ 2, (f) pro vzorky s počtem buněk > 2.

32)
Při pokusu, který hodnotil typ dědičnosti mutantů, bylo získáno 146 divokých a 30 mutantních potomků při křížení F_1 generace mouchy domácí. Otestujte, zda-li získaná data souhlasí s hypotézou, že poměr divokých jedinců k mutantním je 3:1.

33)
Byl studován vztah mezi váhou žáber a váhou těla jednoho druhu kraba (*n*=12) - viz tabulka:

Váha žáber v miligramech	Váha těla v gramech
159	14.4
179	15.2
100	11.3
45	2.5
384	22.7
230	14.9
100	1.4
320	15.8
80	4.1
220	15.3
320	17.2
210	9.2

Předpokládejme normalitu rozdělení obou charakteristik. Jak těsný je vztah mezi těmito charakteristikami? Otestujte.

34)
Teoreticky by měla být variabilita ve velikostech pohlavních orgánů (částí) menší než orgánů (částí) somatických, protože pohlavní orgány jsou více kontrolovány genotypem. Při studiu morfologických vlastností trávy pěchavy vápnomilné (*Sesleria varia*) byly na vzorku n=30 zjištěny tyto vlastnosti: výška rostliny (n=30): průměr=20,5 cm; s=10,0 cm; délka plušky (n=30): průměr=0,44 cm; s=0,005 cm. Která charakteristika pěchavy je více variabilní?

35)
Bobr je dle předchozích výzkumů teritoriální zvíře. Studovali jsme, zda-li je jedním z hlavních důvodů značkování u bobra obrana teritoria. Odchyceným bobrům jsme připevnili vysílačky abychom věděli kde jsou a kolik jich je. Zajímal nás vliv počtu okolních jedinců na chování vždy náhodně vybraného jedince bobra. Získali jsme údaje: počet sousedů, průměrná vzdálenost k dalším teritoriím, počet pachových značek během jara. Ovlivňuje počet sousedů a vzdálenost k dalším teritoriím počet pachových značek vytvořených vybranými jedinci? [další čtení: Rosell F. a Nolet B. (1997): Factors affecting scent-marking behavior in Eurasian beaver.- J. Chem. Ecol., 23: 673-689.]

Bobr	Počet sousedů	Vzdálenost	Počet značek
1	1	4	2
2	2	3.5	5
3	3	3	7
4	4	2.5	12
5	3	2.5	10
6	5	1.5	19
7	6	0.5	25
8	5	1	16
9	1	4.5	3
10	2	3.5	8
11	6	0.5	27
12	7	0.1	31

36)
U člověka je poměr pohlaví narozených dětí 100 samic: 105 samcům. Pokud provedeme 10 000 náhodných výběrů o velikosti 6 novorozenců z celkové populace těchto dětí za rok, jaká bude očekávaná frekvence skupin 6, 5 a 4 chlapců z těchto 10 000 výběrů?

37)
Předpokládejme, že délka kališního lístku v populaci rostlin druhu X je normálně rozdělená s průměrem 3,2 cm a standardní odchylkou 0,8 cm. Jaká část populace bude mít délku kališních lístků (a) větší než 4,5 cm? (b) větší než 1,78 cm? (c) mezi 2,9 a 3,6 cm?

38)
Je předpokládáno, že populace (zvířata) rozšířené v severnějších oblastech budou mít kratší končetiny (přívěsky) než populace (zvířata) v jižnějších částech areálu. Testuj tuto hypotézu za využití délky křídla u jednoho druhu ptáka (údaje v mm):
a) severní arela: 120;113;125;118;116;114;119; b) jižní arela: 116;117;121;114;116;118;123;120.

39)
Data uvádějí spotřebu kyslíku u jednoho druhu ptáka měřenou za různých teplot prostředí

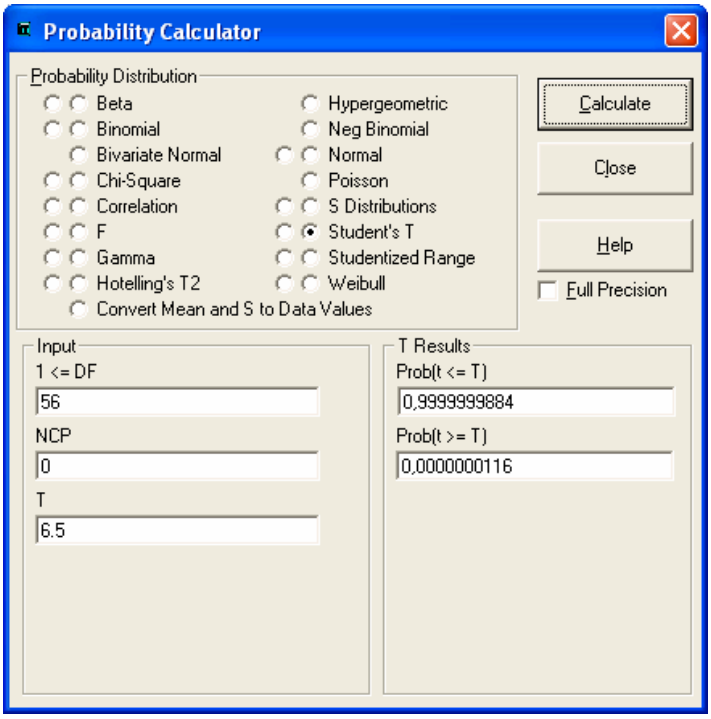
teplota	-18	-15	-10	-5	0	5	10	19
spotřeba O_2	5,2	4,7	4,5	3,6	3,4	3,1	2,7	1,8

a) pokud chceme spočítat lineární regresi, která proměnná je závislá a která nezávislá?
b) spočítejte parametry lineární regrese *a, b.*
c) otestujte ANOVou hypotézu H0: b=0.
d) spočítejte koeficient determinace.

VÝSLEDKY

1) *a) 9,1; 9,7; 9,45; b) 9,42; c) 0,51; d) 0,0547*

2) *jednovýběrový t-test, DF=56, t=-6,55232, P<<0.001, zamítáme H_0, došlo k významné změně. V NCSS nelze jednoduše spočítat (není vhodný modul). Je třeba spočítat t-test v ruce a pak např. (nechci-li brát tabulky) si ve volbě Probability Calculator nasázet příslušné hodnoty a zjistit statistickou významnost t (viz obrázek níže).*



3) *Dvě možnosti: buď parametrický t-test nebo Mann-Whitney U-test, lépe U-test – nelze zaručit normalitu dat:*
a) F-test: F=1.7920, P=0.336931; nezamítáme H_0 o rovnosti variancí, pak mohu užít „klasický“ t-test dvouvýběrový t-test, oboustranná alternativa, t=0,5987, P=0,555, nezamítáme H_0 o rovnosti průměrů (průměr1=104,75; průměr2=100,92)

b) U-test: U=67,5, P = 0,5857, nezamítám H_0 o rovnosti mediánů (medián1=106, medián2=100)

4) *test dobré shody, 2 kategorie; Chi-Square = 0.5319; df = 1; P = 0.465803, nezamítáme H_0 o poměru pohlaví 1:1.*

5) *a)*

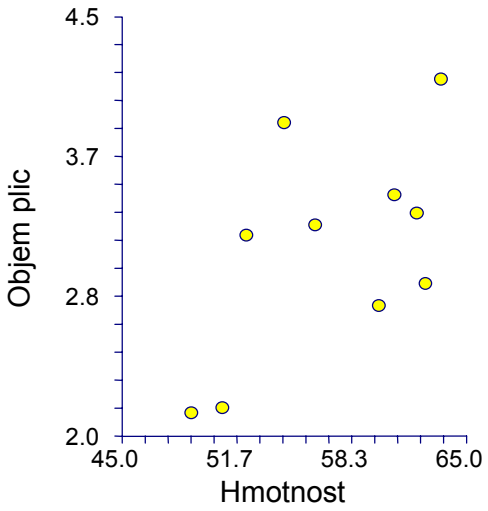
výška:						
Mean	*Standard Deviation*	*Standard Error*	*Minimum*	*Maximum*	*Range*	*Median*
12.2	*2.898275*	*0.9165151*	*9*	*18*	*9*	*11*
průměr růžice:						
Mean	*Standard Deviation*	*Standard Error*	*Minimum*	*Maximum*	*Range*	*Median*
20.5	*6.004628*	*1.89883*	*10*	*27*	*17*	*22*

b) správně Spearmanův korelační koeficient r=0,003, P=0,99, nelze zamítnout H_0 o absenci asociace(korelace) těchto dvou proměnných (v případě Pearsonova r=0,185, P=0,608).

6) *F-test, F=3,82/1,04=3,67, Fkrit(9,9)=3,2, zamítáme H_0 o rovností variancí – laboratoře se odlišují ve variabilitě, s jakou stanovují obsah dusíku*

.

7) diagram:



V tomto případě lze užít obě metody – regresi i korelaci, vzhledem k malému počtu dat je správnější užít neparametrickou korelaci (= Spearmanův korelační koeficient). Lineární regresi by bylo vhodné užít v případě, kdy máme důvody předpokládat, že např. objem plic závisí na hmotnosti člověka a chceme predikovat změny objemu plic v závislosti na hmotnosti lidí.

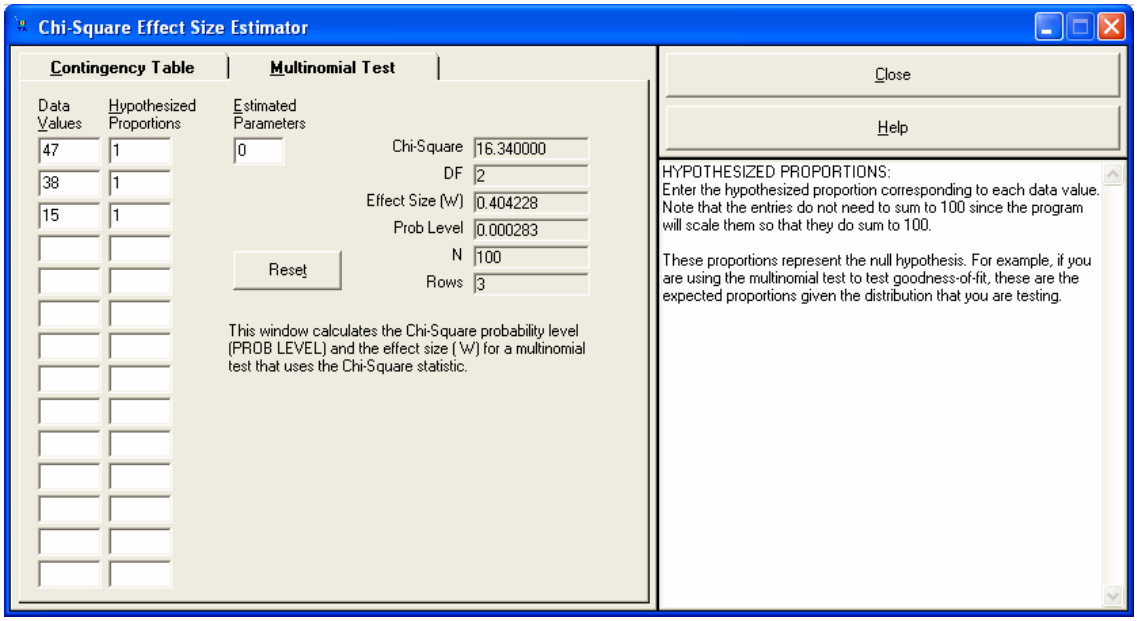
Výsledky:

korelace: Pearsonův korelační koeficient r=0,58, P>0,05, Spearmanův korelační koeficient r=0,60, P>0,05, nezamítáme H_0 o nekorelovanosti mezi oběma proměnnými (tj. r=0,0);

lineární regrese: rovnice y=-0,90+0,0704x; t=test testující H_0: b=0, t=2,01, P>0,05, nezamítáme H_0; ANOVA: Ssreg=1,266, Sse=2,512, Msreg=1,266, Mse=0,314, F=4,03, DF=1;8, P>0,05, nezamítáme H_0, není průkazná lineární regrese

***8)** test dobré shody, chi^2=16,34, DF=2, P<0,001, zamítáme H_0 s shodné preferenci barev u včel.*

v NCSS volba Multinomial test - viz níže:



***9)** řešíme pomocí Z-transformace, 15,9% populace má výšku nižší než 119 a doplněk do 100%, tj. 84,4% má výšku vyšší než 119 cm; 84,1% populace má výšku nižší než 127 cm a doplněk do 100%, tj. 15,9% populace má*

výšku vyšší než 127 cm. 100%-15,9%-15,9%=68,2% populace má výšku mezi 119 a 127 cm (všimněte si, že hodnoty 119 a 127 cm jsou vzdáleny od průměru –1 a +1 standardní odchylka).

10) *jedná se o párové uspořádání dat, lze užít vzhledem k malému počtu pozorování neparametrický Wilcoxonův test, popř. znaménkový test, či párový t-test (protože jsou ale splněny požadavky testu – viz tabulka níže, lze užít párový t-test)*

základní statistika:

Variable	Count	Mean	Standard Deviation	Standard Error
pred	10	72.6	5.081557	1.606929
po	10	74.4	6.131884	1.939072
Difference	10	-1.8	1.813529	0.5734884

Tests of Assumptions about Differences Section

Assumption	Value	Probability	Decision(5%)
Skewness Normality	0.2865	0.774489	Cannot reject normality
Kurtosis Normality	-1.5165	0.129401	Cannot reject normality
Omnibus Normality	2.3818	0.303952	Cannot reject normality
Correlation Coefficient	0.964927		

T-Test For Difference Between Means Section

Alternative Hypothesis	T-Value	Prob Level	Decision (5%)
pred-po<>0	-3.1387	0.011953	Reject Ho
pred-po<0	-3.1387	0.005977	Reject Ho
pred-po>0	-3.1387	0.994023	Accept Ho

Quantile (Sign) Test – znaménkový test

Hypothesized Value	Quantile	Number Lower	Number Higher	Prob Lower	Prob Higher	Prob Both
0	0.5	7	1	0.996094	0.035156	0.070313

Wilcoxon Signed-Rank Test for Difference in Medians

W Sum Ranks	Mean of W	Std Dev of W	Number of Zeros	Number Sets of Ties	Multiplicity Factor
4	26	9.688911	2	3	54

Alternative Hypothesis	Approximation Without Continuity Correction Z-Value	Prob Level	Decision (5%)
X1-X2<>0	2.2706	0.023169	Reject Ho
X1-X2<0	-2.2706	0.011584	Reject Ho
X1-X2>0	-2.2706	0.988416	Accept Ho

Závěr: oboustranný párový t-test testuje nulovou hypotézu o průměru rozdílů=0; t=-3,14,df=9, P=0,01, zamítáme H_0 o stejné tepové frekvenci pacientů před a po podání léku ; Wilcoxonův test: pozor – testuje, že medián(!!!) rozdílů je 0;klasický postup: Wmin= 4, n=10, P=0,01, obdobně i při postupu s normální aproximací – viz tabulka - zamítáme H_0 o stejné tepové frekvenci pacientů před a po podání léku. (v obou případech by šlo testovat i jednostrannou hypotézu, že lék ovlivňuje tepovou frekvenci jedním směrem, ale zadání nám neříká (logika věci ano), jakou stranu zvolit…)

11) *čtyřpolní tabulka = test dobré shody, chi^2=0,421, DF =1, P >0,05, nezamítáme H_0 o nezávislosti mezi přítomností papírny a mihulí v toku*

12) *<3,31;3,85>, průměr je 3,58*

13)

1,2

1,2

1,2

0,0056

15000

14) *jedná o znáhodněné uspořádání jednoho faktoru se 3 hladinami – jednocestná ANOVA. Lze užít jak parametrickou, tak neparametrickou ANOVu (nápověda: Bartletův test nezamítnul nulovou hypotézu o rovnosti variancí)*

Group Detail

Group	*Median*
1	*8.5*
2	*8.4*
3	*7.7*

Means and Effects Section

Term	*Count*	*Mean*
All	*15*	*8.266666*
A: peniciliny		
1	*5*	*9.06*
2	*5*	*8.24*
3	*5*	*7.5*

ANOVA:
Analysis of Variance Table

Source Term	*DF*	*Sum of Squares*	*Mean Square*	*F-Ratio*	*Prob Level*
Faktor A(peniciliny)	*2*	*6.089334*	*3.044667*	*3.98*	*0.047270*
S(A)	*12*	*9.184*	*0.7653334*		
Total (Adjusted)	*14*	*15.27333*			
Total	*15*				

Závěr: zamítám H_0 o stejné účinnosti různých penicilínů. Dílčí průměry se liší.

Kruskal-Wallis One-Way ANOVA on Ranks
Hypotheses
Ho: All medians are equal.
Ha: At least two medians are different.
Test Results

DF	*Chi-Square (H)*	*Prob Level*
2	*5.84*	*0.053934*

Závěr: nezamítám H_0 o stejné účinnosti různých penicilínů. Dílčí mediány se neliší – zde ale velmi těsně.

15) *párové uspořádání dat, buď párový t-test (jsou-li splněny podmínky testu, zde však nejsou, takže ho neužijeme), nebo Wilcoxonův test (ten !!!) či znaménkový test*

Descriptive Statistics Section

Variable	*Count*	*Mean*	*Standard Deviation*	*Standard Error*
A	*8*	*44.875*	*7.10005*	*2.510247*
B	*8*	*45.25*	*8.137217*	*2.87694*
Diference	*8*	*-0.375*	*2.559994*	*0.9050947*

Wilcoxonův test: Wmin=12, P>0,05, nezamítáme H_0 o stejném výnosu obou odrůd (tj. medián rozdílu = 0).

16) *a) ostrov J: průměr 107,4; medián 106; standardní odchylka 8,48; variační koeficient 7,89%; ostrov K: průměr 100,5; medián 100,5; standardní odchylka 4,10; variační koeficient 4,07%;*

b, c) testujeme nulovou hypotézu, že konfidenční interval pro rozdíl mezi průměry bude zahrnovat nulu – tj. případ, že nebude rozdíl mezi průměry

Confidence-Limits of Difference Section

Variance Assumption	DF	Mean Difference	Standard Deviation	Standard Error	95% LCL of Mean	95% UCL of Mean
Equal	33	6.95	6.335446	2.163967	2.547376	11.35262

Protože 95% konfidenční interval pro diferenci leží v intervalu <2,55;11,35>, tedy neobsahuje nulu, zamítáme H_0: neexistuje rozdíl mezi průměry. Závěr: průměry souborů se liší.

17) *test dobré shody, 2 kategorie; Chi-Square = 1,5763, df = 1, P = 0,209296; nezamítáme H_0 o poměru červenookých a albínů v F_1 generaci 1:1.*

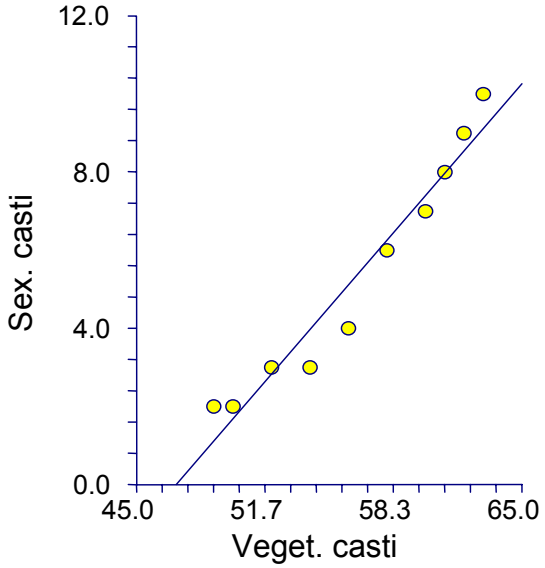
18) *jednoduchá lineární regrese, závislá proměnná – hmotnost sex. částí, nezávisle proměnná – h. veg. částí*

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)
absolutni clen	-26.92043	2.732267	-9.8528	0.000009	Reject Ho
veget.casti	0.572043	4.818356E-02	11.8722	0.000002	Reject Ho
R-Squared	0.946290				

Analysis of Variance Section

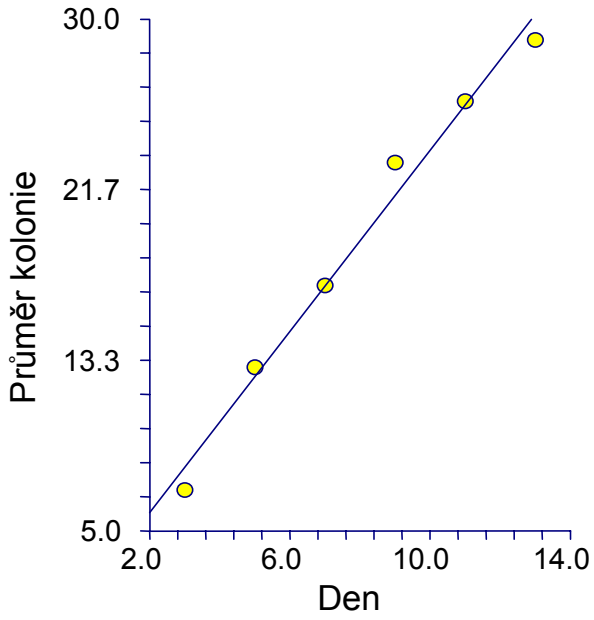
Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	1	291.6	291.6		
Model	1	76.08172	76.08172	140.9482	0.000002
Error	8	4.31828	0.539785		
Total(Adjusted)	9	80.4	8.933333		



Závěr: zamítáme H_0 o tom, že b=0. Lineární regrese je průkazná. Všimněte si pozice jednotlivých reziduí – zdá se, že model není zcela vhodný (proč?, ale máme málo dat…)

19) *čtyřpolní tabulka, test dobré shody, chi^2=12,47, DF=1, P<0,001, zamítáme H_0 o vzájemně náhodném výskytu dešťovek a jetele. Oba subjekty se vyskytují vzájemně nenáhodně.*

20) *jednoduchá lineární regrese, den = nezávislá proměnná, průměr kolonie = závislá proměnná*



Regression Equation Section

Independent Variable	*Regression Coefficient*	*Standard Error*	*T-Value (Ho: B=0)*	*Prob Level*	*Decision (5%)*
Intercept	*1.452381*	*1.232267*	*1.1786*	*0.303871*	*Accept Ho*
cas	*2.214286*	*0.1416617*	*15.6308*	*0.000098*	*Reject Ho*
R-Squared	*0.983892*				

Analysis of Variance Section

Source	*DF*	*Sum of Squares*	*Mean Square*	*F-Ratio*	*Prob Level*
Intercept	*1*	*2204.167*	*2204.167*		
Model	*1*	*343.2143*	*343.2143*	*244.3220*	*0.000098*
Error	*4*	*5.619048*	*1.404762*		
Total(Adjusted)	*5*	*348.8333*	*69.76667*		

Závěr: zamítáme nulovou hypotézu b=0. Lineární regrese je průkazná – existuje lineární závislost průměru kolonie na čase.

21) *jednocestná analýza variance – faktor spektrum má 5 hladin (5 částí spektra), znáhodněné uspořádání pokusu, dat málo, lépe užít Kruskal-Wallisův test, dále výsledky jak pro ANOVu, tak pro K-W test, faktor fixní:*

a) ANOVA:

Means and Effects Section

Term	*Count*	*Mean*
All	*25*	*13.32*
A: spektrum		
bila	*5*	*21*
cervena	*5*	*18.4*
modra	*5*	*5.2*
zelena	*5*	*5.8*
zluta	*5*	*16.2*

Analysis of Variance Table

Source	*DF*	*Sum of Squares*	*Mean Square*	*F-Ratio*	*Prob Level*
A: C4	*4*	*1077.84*	*269.46*	*561.38*	*0.000000**
S(A)	*20*	*9.6*	*0.48*		
Total (Adjusted)	*24*	*1087.44*			
Total	*25*				

Závěr: zamítáme H_0 o tom, že různá spektra mají stejný vliv na rychlost fotosyntézy (tj. že dílčí průměry jsou shodné). Spektra se ve svém účinku liší.

b) Kruskal-Wallis One-Way ANOVA on Ranks

Hypotheses
Ho: All medians are equal.
Ha: At least two medians are different.

Test Results

Method	*DF*	*Chi-Square (H)*	*Prob Level*	*Decision*
Not Corrected for Ties	*4*	*22.16123*	*0.000186*	*Reject Ho*

Group Detail

Group	*Count*	*Sum of Ranks*	*Median*
bila	*5*	*115.00*	*21*
cervena	*5*	*89.00*	*19*
modra	*5*	*20.00*	*5*
zelena	*5*	*35.00*	*6*
zluta	*5*	*66.00*	*16*

Závěr: zamítáme H_0 o tom, že různá spektra mají stejný vliv na rychlost fotosyntézy (tj. že dílčí mediány !!! jsou shodné). Spektra se ve svém účinku liší.

***22)** binomické rozdělení s parametry: n = 20, p = 0,1;*
*a) 12,2%; b) 19,0%; c) 0,89%; d) 2 (n*p=20*0,1=2)*

***23)** znáhodněné uspořádání, 2 hladiny 1 faktoru: t-test nebo (lépe, málo dat) Mann-Whitney U test*
Descriptive Statistics Section

Variable	*Count*	*Mean*	*Standard Deviation*	*Standard Error*
A	*12*	*33.33333*	*3.284491*	*.9481508*
B	*12*	*36.58333*	*4.209477*	*1.215171*

a) a) parametrický test
Variance-Ratio Equal-Variance Test (= F-test) F=1.6426 P=0.423420 Cannot reject equal variances
Equal-Variance T-Test Section

Alternative Hypothesis	*T-Value*	*Prob Level*	*Decision (5%)*
Difference <> 0	*-2.1086*	*0.046594*	*Reject Ho*
Difference < 0	*-2.1086*	*0.023297*	*Reject Ho*
Difference > 0	*-2.1086*	*0.976703*	*Accept Ho*

Diference: (A)-(B)
Závěr: Oboustranný test: P=0,047 – zamítáme nulovou hypotézu o rovnosti průměrů. Je rozdíl v účinku stravy.
b) neparametrický test:
Mann-Whitney U for Difference in Medians

Variable	*Mann Whitney U*
A	*41*
B	*103*

Závěr: Oboustranný test: P=0,07 – nezamítáme nulovou hypotézu o rovnosti mediánů. Není rozdíl v účinku stravy.

24) *jednovýběrový t-test. V NCSS nelze jednoduše spočítat (není vhodný modul). Je třeba spočítat jednovýběrový t-test v ruce a pak např. (nechci-li brát tabulky) si ve volbě Probability Calculator nasázet příslušné hodnoty a zjistit statistickou významnost t (viz obrázek níže).*

Níže výsledek z programu STATGRAPHICS:

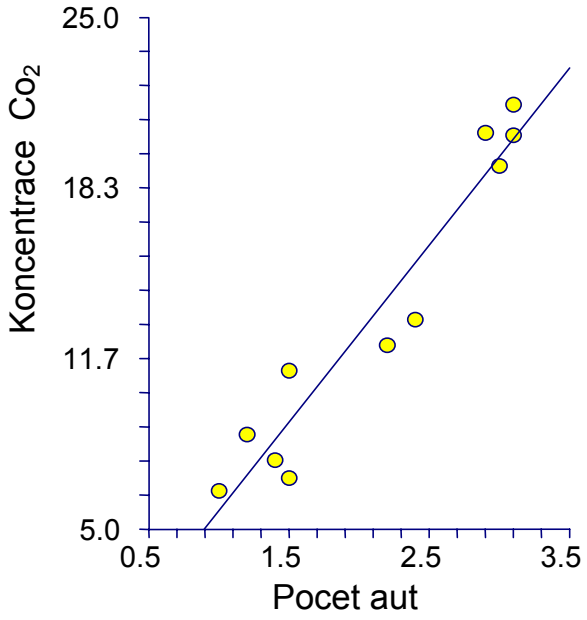
```
Hypothesis Tests
----------------
Sample mean = 2900.0
Sample standard deviation = 360.0
Sample size = 120

95.0% confidence interval for mean: 2900.0 +/- 65.0728   [2834.93;2965.07]

Null Hypothesis: mean = 3100.0
Alternative: not equal
Computed t statistic = -6.08581
P-Value = 2.18771E-7
Reject the null hypothesis for alpha = 0.05.
```

Tj. zamítáme H_0 o stejné hmotnosti dětí kuřaček a obecně dětí v daném regionu, hmotnost dětí kuřaček je nižší.

25) *jednoduchá lineární regrese, nezávislá proměnná – počet aut, závislá proměnná – koncentrace CO_2*



a) lineární regrese

Regression Equation Section

Independent Variable	*Regression Coefficient*	*Standard Error*	*T-Value (Ho: B=0)*	*Prob Level*	*Decision (5%)*
Intercept	*-1.179783*	*1.458982*	*-0.8086*	*0.439577*	*Accept Ho*
Pocet aut	*6.917494*	*.6458753*	*10.7103*	*0.000002*	*Reject Ho*
R-Squared	*0.927249*				

Analysis of Variance Section

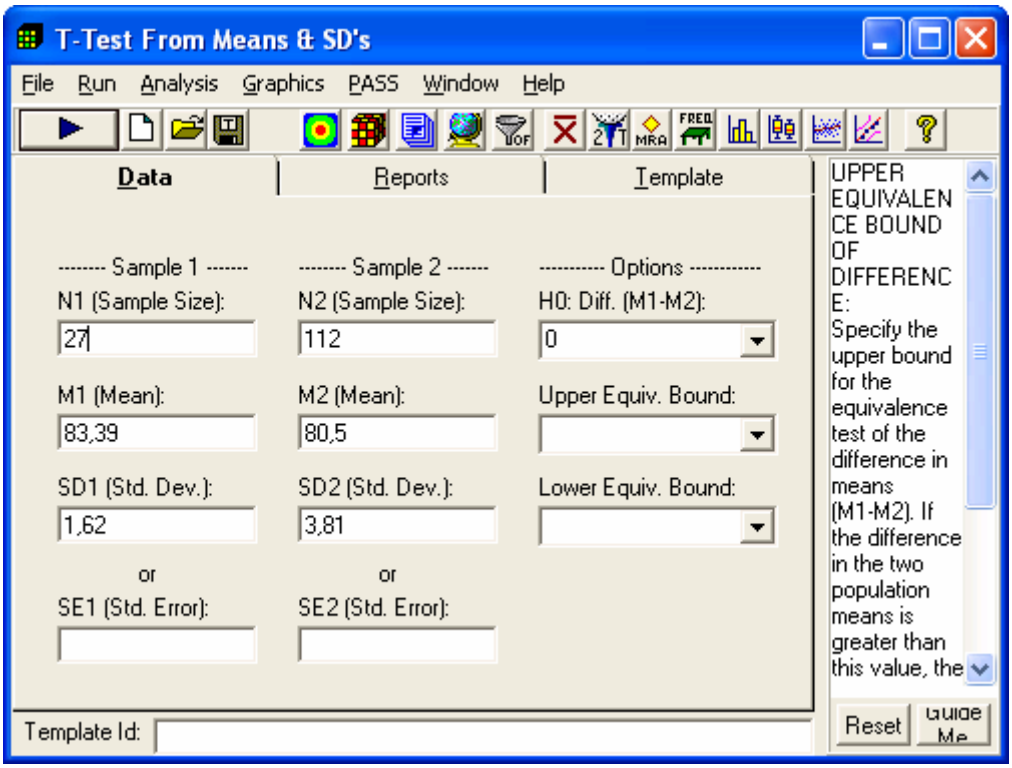
Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	*1*	*1996.658*	*1996.658*		
Model	*1*	*324.2607*	*324.2607*	*114.7097*	*0.000002*
Error	*9*	*25.44114*	*2.826794*		
Total(Adjusted)	*10*	*349.7018*	*34.97018*		

Závěr: existuje signifikantní lineární závislost koncentrace CO_2 na počtu aut. Rovnice: y=-1,1798+6,917x, kde x = počet aut v jednotkách tisíců (!!!).
*b) y=-1,1798+6,917*2,5=16,1127 (x 10^6)*

26) *binomické rozdělení s parametry n=30, p=0,512195 (jak?: 100+105=205, pak p(chlapec)=105/205=0,512195;*
a) 2,17%; b) 15,37 chlapců, tito s 14,45% pravděpodobností

27) *normální rozdělení, nutná Z-transformace;*
a) 5,2%; b) 96,2%; c) 96,2-5,2=91,0%

28) *F-test (v NCSS volba viz obrázek); hledejte output „Equal variance test“.*



Zde výstup ze Statgraphics...

```
Hypothesis Tests
----------------
Sample standard deviations = 3.81 and 1.65
Sample sizes = 112 and 27

95.0% confidence interval for ratio of variances: [2.71976;9.29556]

Null Hypothesis: ratio of variances = 1.0
Alternative: not equal
Computed F statistic = 5.3319
P-Value = 0.00000845845
Reject the null hypothesis for alpha = 0.05.
```

Zamítáme H_0 o stejném rozptylu délky femuru u F1 a F2 generace.

29) *a)* *Summary Section of Čas*

Mean	Deviation	Standard Error	Standard Minimum	Maximum	Range
58.28571	26.01099	9.831229	24	96	72

Means Section of čas

Parameter	Mean	Median
Value	58.28571	56

Skewness and Kurtosis Section of čas

Parameter	Skewness	Kurtosis	Fisher's g1	Fisher's g2	Coefficient of Variation
Value	7.054552E-02	1.763392	9.143745E-02	-1.16786	0.4462669

Quartile Section of čas

Parameter	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Value	24	32	56	80	96

Summary Section of teplota

Mean	Deviation	Standard Error	Standard Minimum	Maximum	Range
104.4286	1.688899	0.6383439	102.8	107	4.2

Means Section of teplota

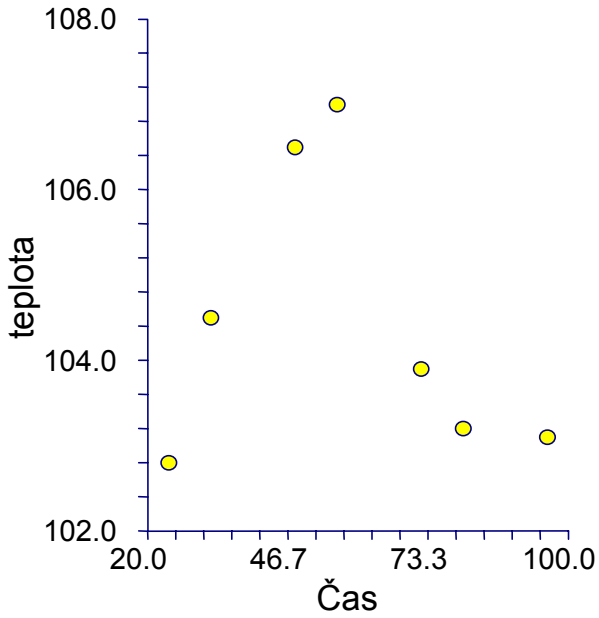
Parameter	Mean	Median
Value	104.4286	103.9

Parameter	Skewness	Kurtosis	Fisher's g1	Fisher's g2	Coefficient of Variation
Value	0.6436763	1.783806	0.8342998	-1.118867	1.617277E-02

Quartile Section of teplota

Parameter	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
	102.8	103.1	103.9	106.5	107

b) z obrázku je vidět, že lineární regrese není nejvhodnějším modelem popisujícím chování teploty, prakticky se ukazuje, že po zvýšení teploty jako reakce na infekci virem dochází k prudkému poklesu v důsledku úspěšné imunitní reakce.V daném případě by bylo vhodné časový interval rozdělit na část do maxima infekce a na část úspěšné imunitní reakce, popř. užít jiný než lineární model



c) lineární jednoduchá regrese, závisle proměnná – teplota, nezávislá proměnná - čas

Regression Equation Section

Independent Variable	*Regression Coefficient*	*Standard Error*	*T-Value (Ho: B=0)*	*Prob Level*	*Decision (5%)*
Intercept	*105.2507*	*1.787521*	*58.8808*	*0.000000*	*Reject Ho*
cas	*-1.410E-02*	*2.8344E-02*	*-0.4976*	*0.639860*	*Accept Ho*
R-Squared	*0.047188*				

Analysis of Variance Section

Source	***DF***	***Sum of Squares***	***Mean Square***	***F-Ratio***	***Prob Level***
Intercept	*1*	*76337.29*	*76337.29*		
Model	*1*	*.8075965*	*.8075965*	*0.2476*	*0.639860*
Error	*5*	*16.30669*	*3.261338*		
Total(Adjusted)	*6*	*17.11429*	*2.852381*		

Závěr: lineární model není signifikantní.

30) *V NCSS nelze jednoduše spočítat (není vhodný modul). Je třeba spočítat jednovýběrový t-test v ruce a pak např. (nechci-li brát tabulky) si ve volbě Probability Calculator nasázet příslušné hodnoty a zjistit statistickou významnost t (viz obrázek níže). Zde výsledek ze Statgraphics:*

```
Normal Tolerance Limits
-----------------------
Sample size = 2666
Sample mean = 79.73
Sample standard deviation = 10.94

95.0% tolerance interval for 99.0% of the population
     Xbar +/- 2.63575 sigma
     Upper: 108.565
     Lower: 50.8949
```

31) *Poissonovo rozdělení s parametrem $\mu=\sigma=1,8$ (v NCSS volba Probability Calculator)*

a) 16,53%; b) 73,06%; c) 26,94%; d) 66,12; e) 292,2; f) 107,6

32) *test dobré shody, dvě kategorie, H_0: divoké : mutantní = 3:1*

	Expected	*Actual*
Skupina	*Count*	*Count*
divoké	*146*	*132.00*
mutantní	*30*	*44.00*

Chi-Square = 5,9394; df = 1; P = 0,0148;
Závěr: zamítáme nulovou hypotézu o poměru 3:1 divokých ku mutantním potomkům v F1 generaci mouchy domácí.

33) *je-li splněna podmínka dvourozměrného normálního rozdělení, pak lze užít pro měření těsnosti vztahu Pearsonův korelační koeficient: r = 0,862805; P < 0,001, n = 12.*
Závěr: pozitivní korelace mezi oběma proměnnými je signifikantní.

34) *užiji variační koeficient; CV(výška)=48,78%; CV(délka plušky)=1,13%;*
Závěr: výška rostliny vykazuje vyšší variabilitu než délka plušky (pozor, bylo by nutné dále otestovat, viz např. Zar (1996); ale pro naše účely není nutné).

35) *mnohonásobná (lineární) regrese, grafické zobrazení - viz obrázek - ukazuje na linearitu vztahu.*

Analysis of Variance Section

Source	DF	R2	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	1		2268.75	2268.75		
Model	2	0.9470	1002.114	501.0569	80.332	0.0000
Error	9	0.0530	56.13615	6.23735		
Total(Adjusted)	11	1.0000	1058.25	96.20454		

Estimated Model
Počet značek= - 0.177+ 4.14 * Počet_sousedů - 0.70959 * Vzdálenost

Zamítáme H0 o neexistenci závislosti počtu značek na počtu sousedů a vzdálenosti. Počet značek stoupá s narůstajícím počtem sousedů a klesá s narůstající vzdáleností.

Pohled na t-testy, které testují parciální regresní koeficienty však ukazuje, že ani jeden z koeficientů není signifikantně odlišný od nuly. Máme tedy paradoxní situaci. Ta je důsledkem existence tzv. ***multikolinearity****, tj. silné (lineární) korelace mezi nezávislými proměnnými, což nám potvrzují další dvě tabulky a poslední obrázek. Řešením je tedy vybrat pouze jeden z nezávislých faktorů (v našem případě lépe pracuje Počet sousedů) a testovat pouze jeho vliv na počet značek.*

Regression Equation Section

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Prob Level	Reject H0 at 5%?
Intercept	-0.1771	15.4361	-0.011	0.9911	No
Počet_sousedů	4.1412	2.2639	1.829	0.1006	No
Vzdálenost	-0.7096	3.0910	-0.230	0.8236	No

Multicollinearity Section

	Variance	R2	Diagonal

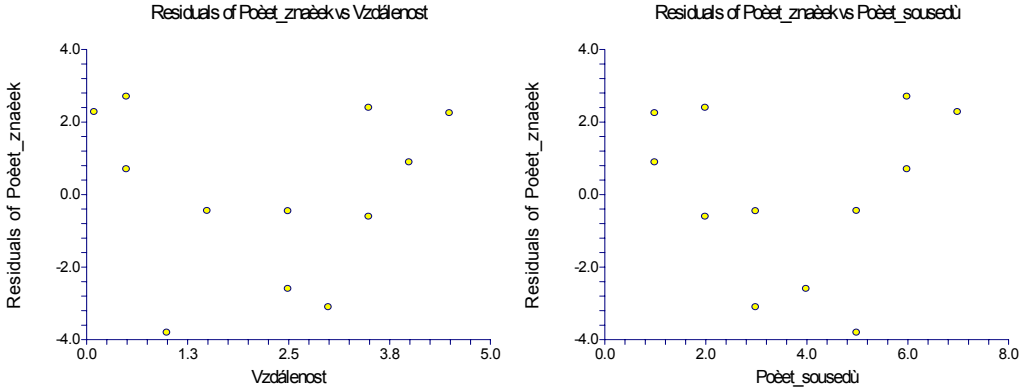
Independent Variable	Inflation Factor	Versus Other I.V.'s	Tolerance	of X'X Inverse
Počet_sousedů	38.0026	0.9737	0.0263	0.821677
Vzdálenost	38.0026	0.9737	0.0263	1.531795

Eigenvalues of Centered Correlations

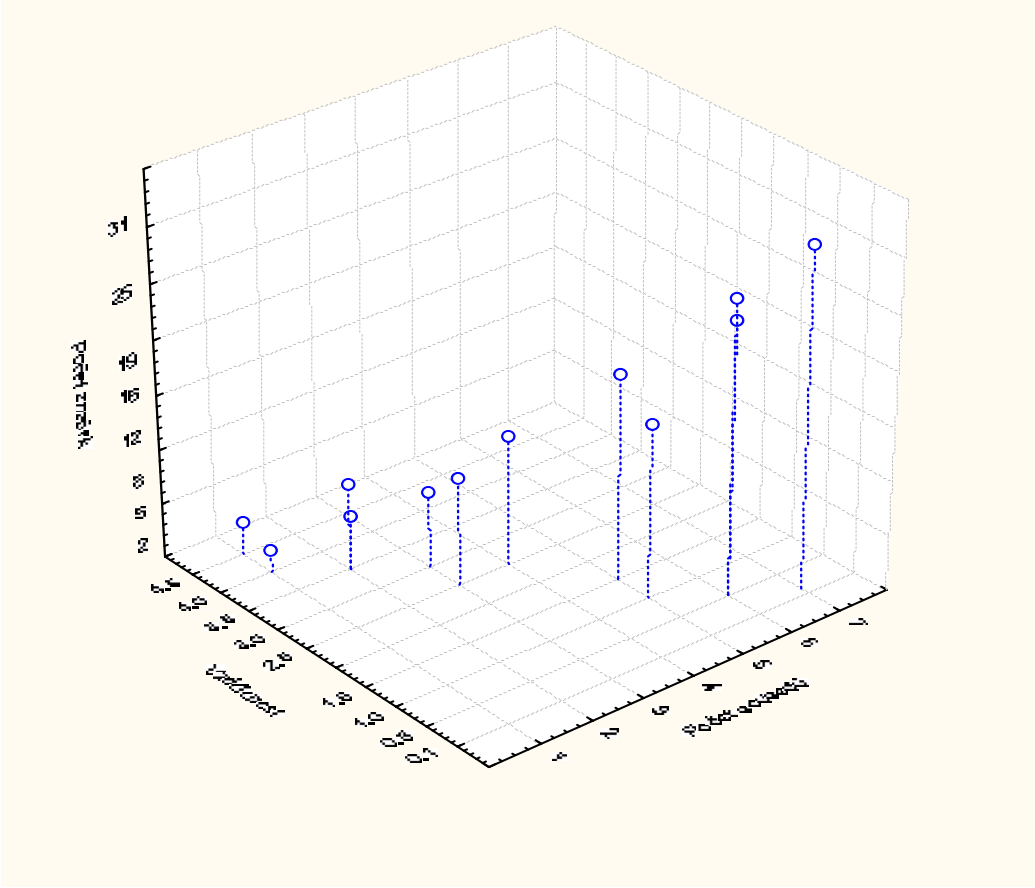
No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	1.9868	99.338	99.338	1.000
2	0.0132	0.662	100.000	150.004

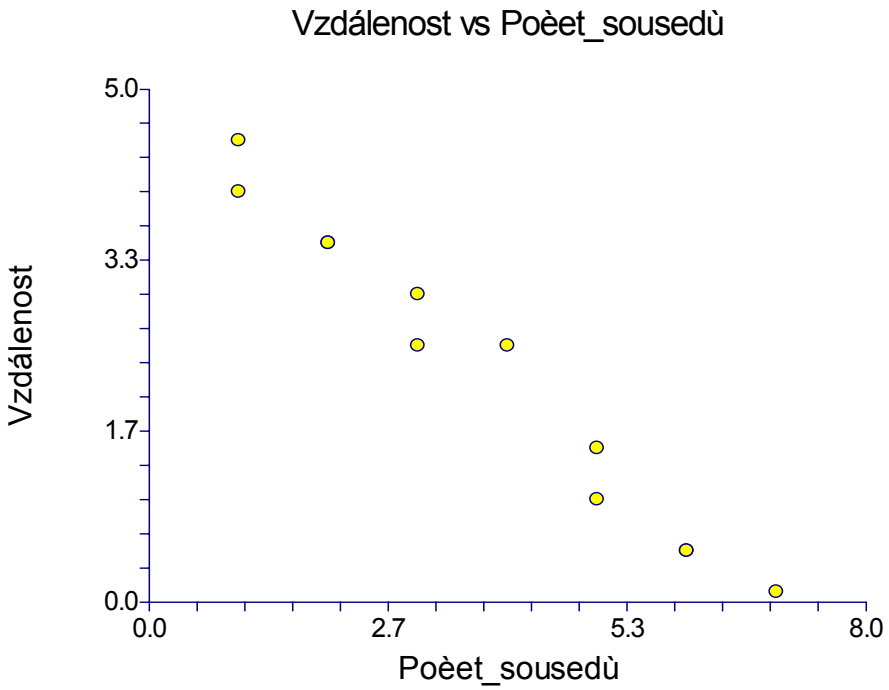
Some Condition Numbers greater than 100. Multicollinearity is a MILD problem.

Obr: Reziduály Početu značek na Vzdálenosti (vlevo) a Počtu sousedů (vpravo).

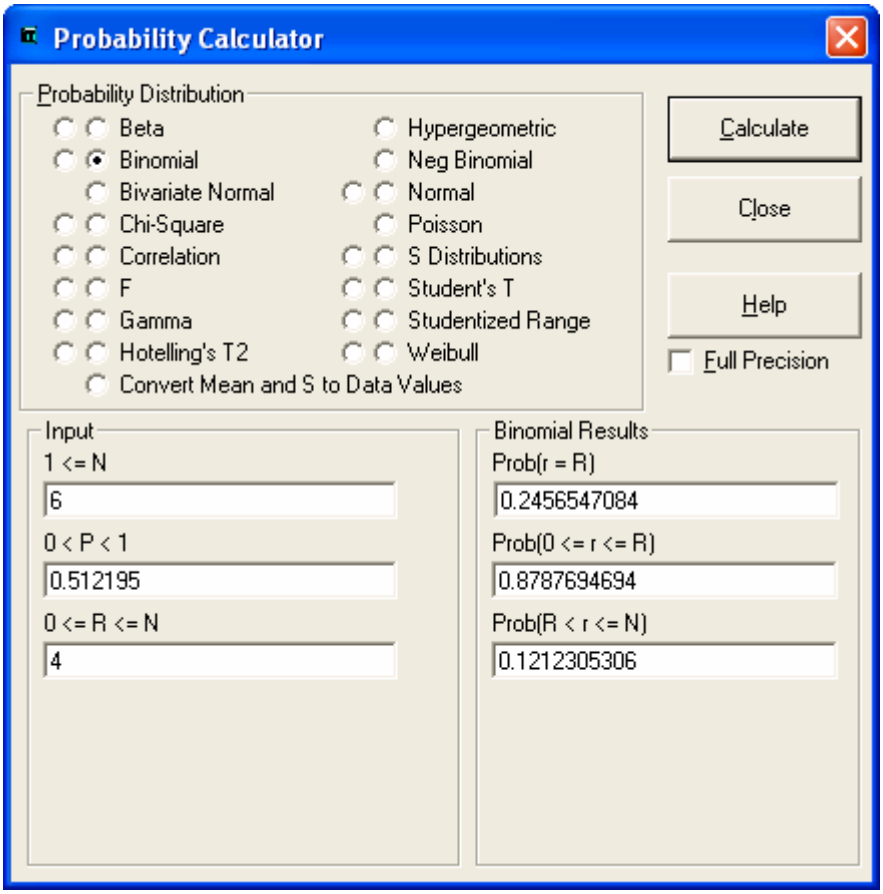


Obr.Vztah mezi vzdáleností okolních teritorií, počtem sousedů a počtem pachových značek vyprodukovaných bobrem za jarní sezónu.

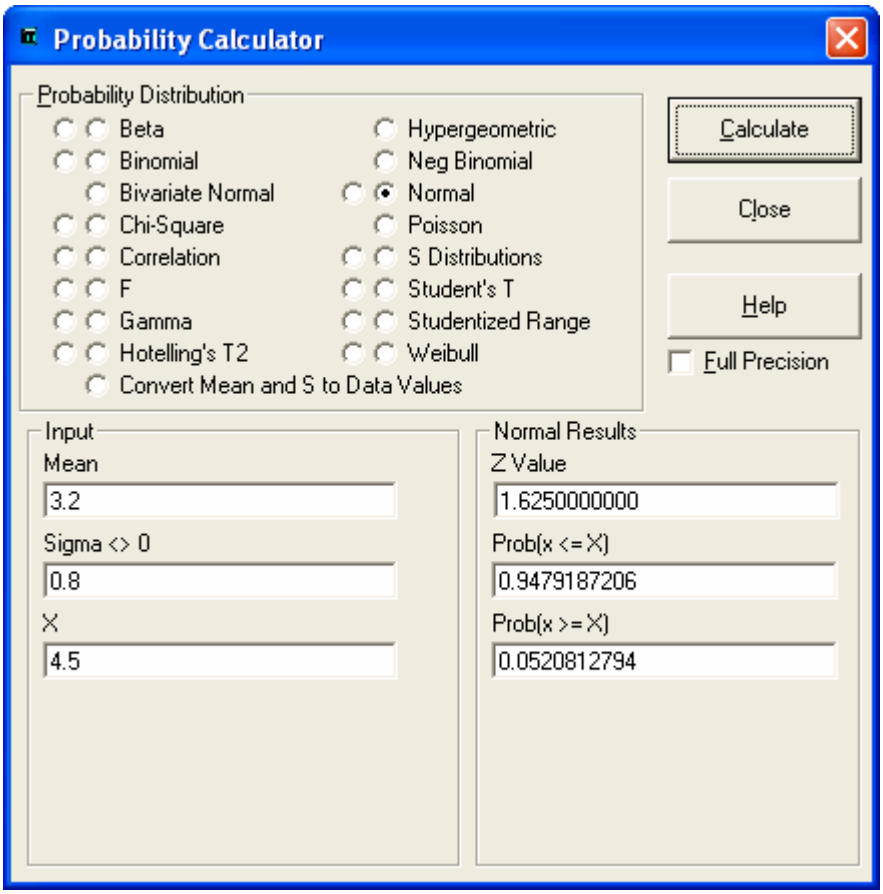




36) *v NCSS volte Probability Calculator, zde příklad pro n = 6 a x = 4, pro 10 000 výběrů nutno výsledek (P(x=4)) vynásobit * 10000, tj. pro x=4 to je 2456,5, pro x=5 to je 1031,7; pro x = 6 to je 180,6 výběrů.*



37) *v NCSS volte Probability Calculator, normální rozdělení, pak pro a) 0.0521, b) 0.9620, c) 0.3373. Níže příklad pro a)*



38) *dvouvýběrové testy, dle testů normality lze užít parametrický dvouvýběrový t-test, ve všech případech (ať parametrický nebo neparametrické testy) nezamítáme Ho o rovnosti středních hodnot.*

	n	Mean	Std. Deviation	Error of Mean
sever	7	117.8571	4.059087	1.534191
jih	8	118.125	2.997022	1.059607

Tests of Assumptions Section

Assumption	Value	Probability	Decision(5%)
Skewness Normality (sever)	0.0000		
Kurtosis Normality (sever)		1.000000	Cannot reject normality
Omnibus Normality (sever)			
Skewness Normality (jih)	0.5290	0.596808	Cannot reject normality
Kurtosis Normality (jih)	-0.4080	0.683253	Cannot reject normality
Omnibus Normality (jih)	0.4463	0.799985	Cannot reject normality
Variance-Ratio Equal-Variance Test	1.8343	0.445585	Cannot reject equal variances
Modified-Levene Equal-Variance Test	0.3333	0.573566	Cannot reject equal variances

Equal-Variance T-Test Section

Alternative Hypothesis	T-Value	Prob Level	Decision (5%)
Difference <> 0	-0.1467	0.885595	Accept Ho
Difference < 0	-0.1467	0.442797	Accept Ho
Difference > 0	-0.1467	0.557203	Accept Ho

Difference: (sever)-(jih)

Median Statistics

Variable	Count	Median	95% LCL of Median	95% UCL of Median
sever	7	118	113	125
jih	8	117.5	114	121

Mann-Whitney U or Wilcoxon Rank-Sum Test for Difference in Medians

Variable	Mann Whitney U	W Sum Ranks	Mean of W	Std Dev of W

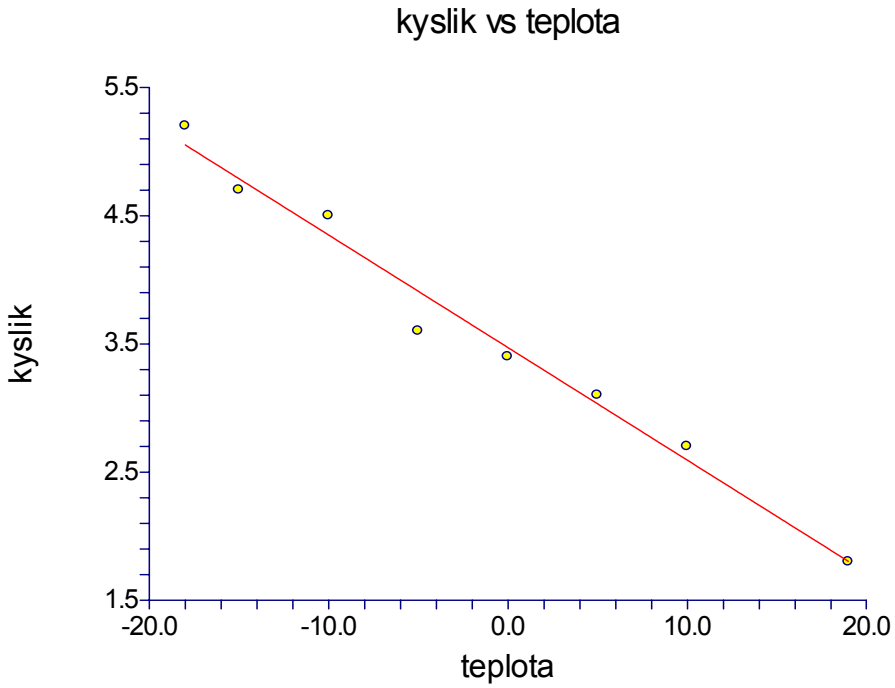
sever	25.5	53.5	56	8.586812
jih	30.5	66.5	64	8.586812

Number Sets of Ties = 4, Multiplicity Factor = 42

Approximation Without Correction

Alternative Hypothesis	Z-Value	Prob Level	Decision (5%)
Diff<>0	-0.2911	0.770941	Accept Ho
Diff<0	-0.2911	0.385471	Accept Ho
Diff>0	-0.2911	0.614529	Accept Ho

39) *jednoduchá lineární regrese, kyslík je závislou a teplota nezávislou proměnnou. Regrese je signifikantní, tj. zamítáme H0: b=0. Model vysvětluje cca 98% chování spotřeby kyslíku na teplotě.*



Regression Estimation Section

Parameter	Intercept (A)	Slope (B)
Regression Coefficients	3.4714	-0.0878
Lower 95% Confidence Limit	3.3243	-0.1000
Upper 95% Confidence Limit	3.6185	-0.0755
Standard Error	0.0601	0.0050
T Value	57.7385	-17.5765
Prob Level (T Test)	0.0000	0.0000
Reject H0 (Alpha = 0.0500)	Yes	Yes

Estimated Model

(3.47142228093351) + (-8.77586966094232E-02) * (teplota)

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1	105.125	105.125			
Slope	1	8.745154	8.745154	308.9326	0.0000	1.0000
Error	6	0.1698459	2.830765E-02			
Adj. Total	7	8.915	1.273571			
Total	8	114.04				

Correlation and R-Squared Section

Parameter	Pearson Correlation Coefficient	R-Squared	Spearman Rank Correlation Coefficient
Estimated Value	-0.9904	**0.9809**	-1.0000