

1. cvičení

Příklad 1

ručně

Jaká je pravděpodobnost, že náhodná veličina $X \sim N(20, 16)$ nabude hodnotu menší, než 12 nebo větší, než 28?

Příklad 2

ručně

Nechť X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(0, 1)$, $i = 1, 2$. Zjistěte, jaké rozložení má transformovaná náhodná veličina $Y = 3 + X_1 - 2X_2$, určete její parametry a najděte dolní kvartil náhodné veličiny Y .

Příklad 3

ručně

Nechť devizový kurs eura je náhodná veličina $X_1 \sim N(19, 0.5^2)$ a devizový kurs dolaru je náhodná veličina $X_2 \sim N(32, 0.6^2)$. Korelace $R(X_1, X_2) = -0.8$. Jaká je pravděpodobnost, že měnový koš $0.65X_1 + 0.35X_2$ bude mít hodnotu větší než 24?

Následující příklady pomocí sw.:

Příklad 4

Do datového okna programu STATISTICA načteme datový soubor *lide.txt*, který obsahuje následující údaje o 32 náhodně vybraných osobách: Jméno (křestní jméno osoby); Pohlaví (1 muž, 2 žena); Věk (věk osoby v dosažených letech); Výška (výška osoby v cm); Hmotnost (hmotnost osoby v kg).

Příklad 5

Proměnné nazveme Jméno, Sex, Věk, Výška, Hmotnost a vytvoříme jim návěští (Jméno - křestní jméno osoby, Sex - pohlaví osoby, Věk - věk v letech, Výška - výška v cm, Hmotnost - hmotnost v kg). Popíšeme, co znamenají jednotlivé varianty u proměnné Sex (1 - muž, 2 - žena).

Příklad 6

Pomocí správce jmen případů převedte proměnou Jméno na jména případů. Proměnnou Jméno poté zrušte.

Příklad 7

Pro každou osobu vypočtete BMI.

(Body Mass Index se počítá podle vzorce $\frac{\text{hmotnost v kg}}{\text{výška}^2 \text{ v m}^2}$. Osoby, které mají BMI pod 18,5, trpí podvýživou; BMI mezi 18,5 a 25 ukazuje na normální stav; hodnoty mezi 25 a 30 svědčí o nadváze a hodnoty nad 30 pak o obezitě.)

Příklad 8

Pro proměnné Věk, Výška, Hmotnost, BMI zjistěte minimum, maximum a průměr.

Příklad 9

Vytvořte sloupcový graf, kde na x -ové ose budou jednotlivé osoby a na y -ové ose hodnoty BMI. Pod každý sloupec umístěte jméno příslušné osoby.

Příklad 10

Nyní vytvořte sloupcové grafy pro BMI zvlášť pro muže a zvlášť pro ženy.

Doma teoretické příklady:

Příklad 11

Nechť X_1, X_2, X_3, X_4 jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(0, 1)$,
 $i = 1, 2, 3, 4$. Jaké rozložení má transformovaná náhodná veličina $X = \frac{X_1\sqrt{3}}{\sqrt{X_2^2+X_3^2+X_4^2}}$?

Příklad 12

Nechť náhodná veličina $X \sim F(n_1, n_2)$. Zjistěte rozložení transformované náhodné veličiny
 $Y = \frac{1}{X}$. [$Y \sim F(n_2, n_1)$]

Příklad 13

Nechť náhodná veličina $X \sim t(n)$. Zjistěte rozložení transformované náhodné veličiny
 $Y = X^2$. [$Y \sim F(1, n)$]

Pamatujte, že:

$$\left. \begin{array}{l} u_\alpha = -u_{1-\alpha} \\ t_{\alpha(n)} = -t_{1-\alpha}(n) \\ F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)} \end{array} \right\}$$

kde u_α je α kvantil normálního rozložení; t_α je α kvantil studentova rozložení; $F_\alpha(n_1, n_2)$ je α kvantil Fisher-Snedocerova rozložení;

Příklad 14

Příklad 1 pomocí sw.

SW. návody k 1. cvičení

Návod k řešení příkladu 4:

Soubor - Otevřít - *Lide.txt*-importovat jako tabulku,definovat - OK - OK

Návod k řešení příkladu 5:

Kurzor nastavíme na Prom1 - 2x klikneme myší - Jméno: Jméno - Dlouhé jméno: křestní jméno osoby - OK.

Kurzor nastavíme na Prom 2 - 2x klikneme myší - Jméno: Sex - Dlouhé jméno: pohlaví osoby, Text. hodnoty - 1 muž, 2 žena (Přepínání mezi číselnými hodnotami a jejich textovým popisem se děje pomocí tlačítka s ikonou štítku.)

Podobně vytvoříme návěští pro další proměnné.

Návod k řešení příkladu 6:

Data - Správce jmen případů - Přenést jména případů z proměnné Jméno - OK - OK. Proměnnou Jméno nyní zrušíme takto: Kurzorem se nastavme na proměnnou Jméno - Data - Proměnné - Odstranit - OK.

Návod k řešení příkladu 7:

Za proměnnou Hmotnost vložíme novou proměnnou: Data - Proměnné - Přidat - Za Hmotnost - Jméno BMI. Do Dlouhého jména napíšeme vzorec $= 10000 * \text{Hmotnost} / \text{Výška}^2$

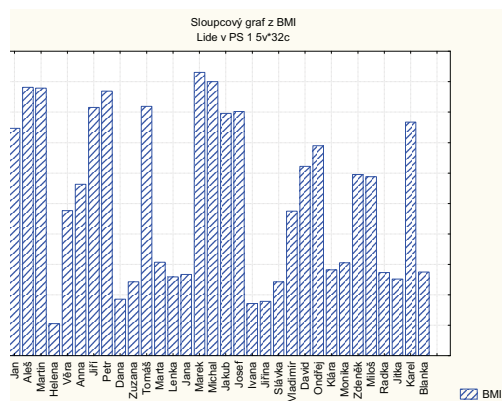
Návod k řešení příkladu 8:

Tento úkol lze splnit více způsoby, zatím si ukážeme postup pomocí funkce Statistika bloku dat, která slouží k umístování základních číselných charakteristik proměnných (např. průměr, směrodatná odchylka, medián, maximum, minimum atd.) přímo do datového okna. Kliknutím myší označíme proměnné, pro které chceme výpočet provést - Statistiky - Statistiky bloku dat - Blok sloupců - zvolíme danou charakteristiku.

Návod k řešení příkladu 9:

Grafy - 2D Grafy - Sloupcové/pruhové grafy - Proměnné BMI - OK - OK Vytvoří se sloupcový graf, který upravíme tak, aby každý sloupec byl popsán jménem příslušné osoby. 2x klikneme na některé jméno v popisu vodorovné osy. Otevře se okno "Možnosti grafu", kde v stromové struktuře je aktivní "Hodnoty měřítka" - v Přeskakovat hodnoty zaškrtneme Vypnuto a v Rozložení vybereme Kolmo na osu - OK.

Dále je vhodné na svislé ose znázorňovat hodnoty pouze od 16 do 26. 2x klikneme na některé číslo v popisu svislé osy. V okně "Možnosti grafu" přejdeme na Měřítka - Mód Ručně, Minimum 16, Maximum 26 - Upravit krok - Mód Ručně - Velikost kroku 1 - OK. Výsledný graf:



Návod k řešení příkladu 10:

Je zapotřebí nejprve vybrat pouze muže. Grafy - 2D Grafy - Sloupcové/pruhové grafy - Filtr případů - zaškrtneme Zapnout filtr, některé, vybrané pomocí výrazu $\text{Sex} = 1$, OK. Dále postupujeme analogicky jako v předchozím úkolu. Stejným způsobem vytvoříme graf pro ženy, kde ve filtru zadáme podmínku $\text{Sex} = 2$.

2. cvičení

Příklad 1

ručně

Jsou dány hodnoty 10, 12, 8, 16, 9. Považujme je za realizace náhodného výběru, které má střední hodnotu μ , rozptyl σ^2 a distribuční funkci $F(x)$. Vypočtěte realizace výběrového průměru, výběrového rozptylu, výběrové směrodatné odchylky a sestrojte graf výběrové distribuční funkce $F_n(x)$.

Příklad 2

a) pomocí sw.

V souboru `2_test_body.sta` máme k dispozici výsledky testů ze dvou předmětů zjištěné u osmi náhodně vybraných studentů určitého oboru:

Číslo studenta	1	2	3	4	5	6	7	8
Počet bodů v 1. testu	80	50	36	58	42	60	56	68
Počet bodů v 2. testu	65	60	35	39	48	44	48	61

Vypočtěte realizace výběrového průměru, výběrové směrodatné odchylky a výběrový koeficient korelace.

b) ručně

Vypočtěte a interpretejte realizaci výběrového koeficientu korelace. Pro usnadnění výpočtů máte k dispozici tyto součty:

$$\sum_{i=1}^8 x_i = 450, \quad \sum_{i=1}^8 y_i = 400, \quad \sum_{i=1}^8 x_i^2 = 26\,684, \quad \sum_{i=1}^8 y_i^2 = 20\,836, \quad \sum_{i=1}^8 x_i y_i = 23\,214,$$

Výpočetní (ne definiční) vztah pro výběrový rozptyl je:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} M^2$$

Výpočetní (ne definiční) vztah pro výběrovou kovarianci je:

$$S_{12} = \frac{1}{n-1} \sum_{i=1}^n X_i Y_i - \frac{n}{n-1} M_1 M_2$$

Příklad 3

ručně

Nechť X_1, \dots, X_n je náhodný výběr z rozložení, které má střední hodnotu μ a rozptyl σ^2 . Odvoďte střední hodnotu výběrového průměru $E(M)$ a rozptyl výběrového průměru $D(M)$. Dá se dále ukázat, že $E(S^2) = \sigma^2$ a $E(S_{12}) = \sigma_{12}$. Výsledky je potřeba si pamatovat.

Příklad 4

pomocí sw. **Průzkum chování výběrového průměru a výběrového rozptylu.**

- 1.) Nejdříve vytvořte nový datový soubor o 103 proměnných a 100 případech. Potom pomocí programu `2_gener.svb`, který si stáhnete z učebních materiálů, se naplní prvních 100 proměnných 100 realizacemi náh. veličin $X_i \sim Rs(0, 1)$, $i = 1, \dots, 100$, do proměnné `v101` se uloží pořadová čísla 1 až 100, do proměnné `v102` se uloží průměry a do prom. `v103` rozptyly proměnných `v1` až `v100`. (Po otevření makra se objeví panel nástrojů, kde modrá šipka spouští makro.) Proměnnou `v101` přejmenujte na `PORADI`, `v102` na `PRUMER` a `v103` na `ROZPTYL`. Vzniklý datový soubor uložte pod názvem `uniform.sta`.
- 2.) Graficky znázorněte do jednoho grafu hodnoty některé z proměnných `v1, \dots, v100` (např. `v1`) a hodnoty proměnné `PRUMER`.

- 3.) Vypočtete průměr a rozptyl např. proměnné $v1$ a proměnné PRUMER. Jakým hodnotám by měl být blízký průměr proměnné $v1$, rozptyl proměnné $v1$, průměr proměnné PRUMER a rozptyl proměnné PRUMER? Spočítejte ručně a porovnejte s výsledky Vašeho datového souboru.
- 4.) Nakreslete histogram pro proměnnou $v1$ a pro proměnnou PRUMER. (První histogram se blíží úsečce, druhý Gaussově křivce.)

Komentář:

V sloupcích (s výjimkou PORADI) máme realizace náhodných veličin: X_1, X_2, \dots, X_{100} , dále výběrového průměru $M = \frac{1}{100} \sum_{i=1}^{100} X_i$ a výběrového rozptylu $S^2 = \frac{1}{99} \sum_{i=1}^{100} (X_i - M)^2$. Z příkladu 3 víme, že $E(X_i) = \mu$, $D(X_i) = \sigma^2$, $E(M) = \mu$, $D(M) = \frac{\sigma^2}{n}$. Dále $X_i \sim Rs(0, 1)$, tedy hodnoty μ a σ umíme určit.

Průměr proměnné $v1$ slouží k ilustraci $E(X_1)$; rozptyl proměnné $v1$ slouží k ilustraci $D(X_1)$; průměr proměnné PRUMER slouží k ilustraci $E(M)$ a rozptyl proměnné PRUMER slouží k ilustraci $D(M)$.

Náhodná veličina M má pro $n \rightarrow \infty$ dle centrální limitní věty normální rozložení. Proto histogram proměnné PRUMER ($n = 100$) připomíná Gaussovu křivku.

Doma: Vypočtete průměr proměnné ROZPTYL. Jaké hodnotě by měl být "blízký"? (Průměr proměnné ROZPTYL slouží k ilustraci $E(S^2) = \sigma^2$.)

Příklad 5

ručně

Předpokládejme, že velký ročník na vysoké škole má výsledky ze statistiky normálně rozloženy kolem střední hodnoty 72 bodů se směrodatnou odchylkou 9 bodů. Najděte pravděpodobnost, že

- náhodně vybraný student bude mít výsledek nad 80 bodů.
- průměr výsledků náhodného výběru 10 studentů bude větší, než 80 bodů.

Příklad 6

ručně

Odvoďte pravděpodobnostní funkci náhodného výběru pocházejícího z alternativního rozložení $A(\vartheta)$.

samostatná práce doma:

V samostatném souboru *2_speciál_zadání.pdf* jsou úkoly k datovému souboru, s kterým jsme pracovali v podzimním semestru. Návodů pro práci se softwarem jsou v *2_sw_návody.pdf*.

SW. návody ke 2. cvičení

Návod k řešení příkladu 2a:

Otevřeme tabulku *2_test_body.sta*.

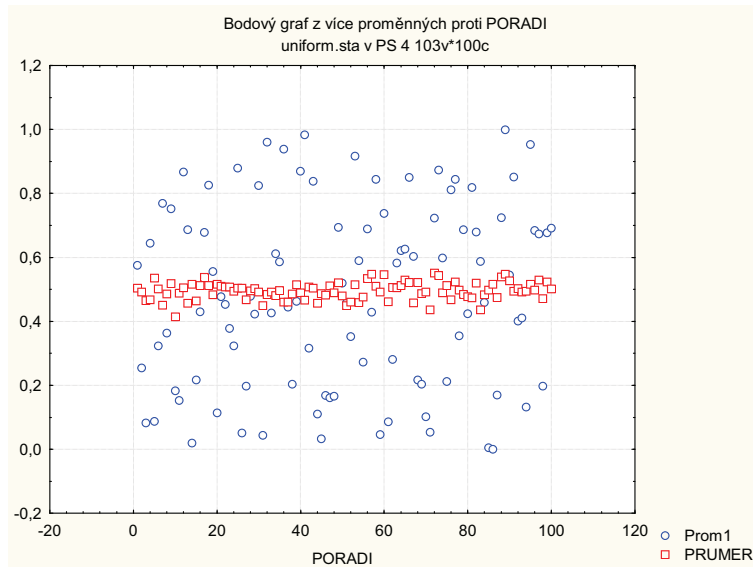
Roletka statistiky-základní statistiky a tabulky-Popisné statistiky-proměnné-vybrat vše-OK-výpočet. Takto jsme získali výběrové průměry a rozptyly pro obě proměnné.

Výběrovou korelaci získáme postupem:

Roletka statistiky-základní statistiky a tabulky-korelační matice-1seznam proměnných-vše-Souhrn:Korelace. V posledních dvou sloupcích na vedlejší diagonále je výběrový korelační koeficient mezi počtem bodů v prvníma druhém testu. V prvních dvou sloupcích jsou opět výběrové průměry a rozptyly pro obě proměnné.

Návod k řešení příkladu 4:

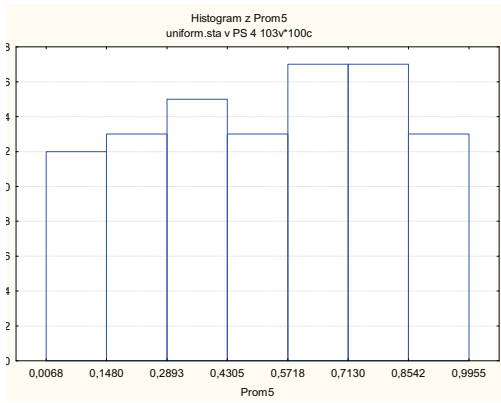
- ad 2.) Grafy - Bodové grafy - Typ grafu Vícenásobný - vypneme Lineární proložení - Proměnné X-PORADI; Y-v1, PRUMER, OK, OK. Vidíme, že hodnoty proměnné v1 se nacházejí mezi 0 a 1, zatímco hodnoty proměnné PRUMER se koncentrují v úzkém pásmu kolem 0,5.



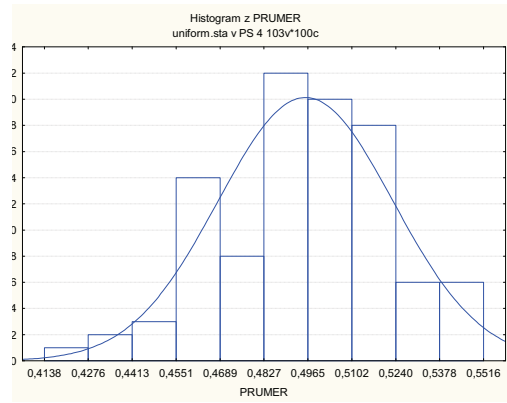
- ad 3.) Roletka statistiky-základní statistiky a tabulky-Popisné statistiky-proměnné-v1,PRUMER-OK-záložka Detailní výsledky, zaškrtnout rozptyl, odškrtnout smer. odch.- výpočet. Ve výstupní tabulce si všimněte, že průměr v1 a průměr PRUMERU se moc neliší, ale rozptyl průměru je přibližně 100-násobně menší, než rozptyl v1.

Proměnná	Popisné statistiky (uniform.sta v PS 4)		
	N platných	Průměr	Rozptyl
Prom1	100	0,483526	0,081679
PRUMER	100	0,495279	0,000746

- ad 4.) grafy-histogramy-proměnné v1(resp. prumer) OK-typ grafu-běžný OK.



kjg



kjg

2. cvičení speciál

Vedení pojišťovny (zaměřené na pojištění automobilů) požádalo manažera oddělení marketingového výzkumu o provedení průzkumu, který by ukázal názory zákazníků na uvažovaný nový systém pojištění aut. Náhodně bylo vybráno 110 současných zákazníků pojišťovny a ti byli telefonicky seznámeni s následujícím textem: "Naše pojišťovna nabízí nový systém pojištění aut výhradně pro cesty nad 300 km. Za roční poplatek 12 tisíc Kč budete pojištěni pro případ libovolných potíží s autem při všech cestách nad 300 km. V případě nehody pojišťovna uhradí opravu, cestovní náklady a popř. i některé další výlohy, jako je ubytování a stravování v hotelu, telefon atd. Stupnicí od 1 (jednoznačný nezájem) do 5 (jednoznačný zájem) laskavě vyjádřete svůj postoj k nabízenému novému typu pojištění. Dále uveďte svůj věk, počet cest nad 300 km v loňském roce, stáří vašeho auta a váš rodinný stav. Děkujeme."

Získané odpovědi byly zaznamenány do datového souboru *pojist.sta* a zakódovány takto: POSTOJ ... postoj k novému typu pojištění (jednoznačný nezájem = 1, lehký nezájem = 2, neutrální postoj = 3, lehký zájem = 4, jednoznačný zájem = 5). RODSTAV ... rodinný stav (svobodný = 1, rozvedený, ovdovělý = 2, ženatý = 3). VEK ... věk v dokončených letech. STARIAUT ... stáří auta v letech. CESTY ... počet cest nad 300 km v předešlém roce.

Příklad 1

Datový soubor *pojist.sta* načtete do systému STATISTICA. Všem proměnným vytvořte návěští a popište význam jednotlivých variant proměnných POSTOJ a RODSTAV.

Příklad 2

Zjistěte absolutní a relativní četnosti a absolutní a relativní kumulativní četnosti proměnných POSTOJ a RODSTAV.

Příklad 3

Absolutní četnosti proměnných POSTOJ a RODSTAV znázorněte graficky pomocí výsečového diagramu.

Příklad 4

Proměnnou VEK zakódujte do 6 třídicích intervalů $\langle 23, 29 \rangle$, $\langle 29, 35 \rangle$, $\langle 35, 41 \rangle$, $\langle 41, 47 \rangle$, $\langle 47, 53 \rangle$, $\langle 53, 59 \rangle$ a zjistěte jejich četnosti.

Příklad 5

Vytvořte histogram proměnné VEK se šesti třídicími intervaly $\langle 23, 29 \rangle$, $\langle 29, 35 \rangle$, $\langle 35, 41 \rangle$, $\langle 41, 47 \rangle$, $\langle 47, 53 \rangle$, $\langle 53, 59 \rangle$.

Příklad 6

Vytvořte kategorizovaný histogram proměnné VEK podle proměnné RODSTAV.

Příklad 7

Vypočítejte následující číselné charakteristiky: modus, medián, dolní a horní kvartil, kvartilová odchylka (t.j. kvartilové rozpětí) pro POSTOJ (ordinální proměnná); modus pro RODSTAV (nominální proměnná); průměr, směrodatná odchylka, variační koeficient, šikmost, špičatost pro VEK, STARIAUT, CESTY (poměrové proměnné).

(Jsou-li hodnoty x_1, \dots, x_n rozloženy symetricky kolem aritmetického průměru, je šikmost rovna nule. Při záporné šikmosti mluvíme o záporně zešikmených souborech, kde nadprůměrné hodnoty jsou četnější, než podprůměrné; při kladné šikmosti mluvíme o kladně zešikmených souborech, kde jsou četnější podprůměrné hodnoty.)

Špičatost popisuje, jakým způsobem jsou hodnoty x_1, \dots, x_n koncentrovány kolem aritmetického průměru. Je-li špičatost nulová, mluvíme o normální špičatosti. Při záporné špičatosti je rozložení souboru plošší, než normální rozložení; při kladné špičatosti je rozložení souboru více špičaté, než normální rozložení.)

Příklad 8

Pro proměnnou STARIAUT nakreslete histogram s proloženou hustotou normálního rozložení. Ponechte implicitní počet třídících intervalů.

Další příklady k samostatnému řešení:

Načtěte datový soubor lide.sta, s nímž jste pracovali v 1. cvičení.

1. Vytvořte tabulku absolutních a relativních četností proměnné SEX. Četnosti znázorněte pomocí výsečového diagramu.

2. Vytvořte histogram proměnné VEK se šesti třídícími intervaly $(16, 23)$, $(23, 30)$, $(30, 37)$, $(37, 43)$, $(43, 50)$, a zakreslenou Gaussovou křivkou.

3. Vytvořte kategorizované histogramy proměnné BMI pro muže a pro ženy.

4. Vypočtěte průměr, směrodatnou odchylku, variační koeficient, šikmost a špičatost proměnné BMI pro muže a pro ženy.

5. Vytvořte kategorizované krabicové diagramy pro proměnnou Vyska pro muže a pro ženy.

SW. návody ke 2. cvičení speciál

Návod k řešení příkladu 1:

Soubor - Otevřít - *pojist.sta* - Otevřít.

Názvy a vlastnosti proměnných se upravují v okně, do něhož vstoupíme, když 2x klikneme myší na název proměnné. Návěští se píše do Dlouhého jména, význam variant do Text. hodnot.

Návod k řešení příkladu 2:

Statistiky - Základní statistiky/Tabulky - Tabulky četností - OK - Proměnné POSTOJ, RODSTAV - OK - Výpočet. Tabulky se uloží do pracovního sešitu, listovat v nich můžeme pomocí stromové struktury v levé části okna.

Návod k řešení příkladu 3:

V menu zvolíme Grafy - 2D Grafy - Výšečové grafy. Vybereme proměnné POSTOJ, RODSTAV

Editace obrázku poklepáním na pozadí grafu, otevře se okno "možnosti grafu", v záložce výšečové grafy lze upravovat např. typ (2D nabo 3D), tvar (elipsa, kružnice) a pod.

Návod k řešení příkladu 4:

Za proměnnou VEK vložíme novou proměnnou RVEK (Data - Proměnné - Přidat - Za VEK, Jméno RVEK, Dlouhé jméno zakódovaný věk, OK). Nastavíme se kurzorem na RVEK. Data - Překódovat - Kategorie 1 Zahrnout pokud $VEK \geq 23$ and $VEK \leq 29$, do okénka Nová hodnota1 zapíšeme 1 atd. až Kategorie 6 Zahrnout pokud $VEK > 53$ and $VEK \leq 59$, do okénka Nová hodnota 6 zapíšeme 6, OK. Četnosti zjistíme analogicky jako v úkolu 4.

Návod k řešení příkladu 5:

V menu vybereme Grafy - Histogramy - Proměnné VEK, OK, záložka Detaily - zaškrtneme Hranice - Určit hranice - zaškrtneme Zadejte hraniční rozmezí, Minimum 23, Krok 6, Maximum 59 - OK - Vypneme normální proložení - OK.

(Soubor vykazuje kladné zešikmení, protože mladší věkové kategorie jsou zastoupeny s vyšší četností než starší věkové kategorie.)

Návod k řešení příkladu 6:

Postupujeme stejně jako v předešlém případě a dále v záložce Kategorizovaný zvolíme - Kategorie X - Zapnuto - Změnit proměnnou RODSTAV - OK - OK.

Návod k řešení příkladu 7:

Statistiky - Základní statistiky/tabulky - Popisné statistiky - OK, Proměnné - zadáme název příslušné proměnné, Detailní výsledky - vybereme příslušné charakteristiky.

Návod k řešení příkladu 8:

Grafy - Histogramy - Proměnné STARIAUT - OK.

3. cvičení

Příklad 1

ručně

11.A;

c) Dále dokažte, že posloupnosti: $M_1, M_2, \dots, M_n, \dots$ a $L_1, L_2, \dots, L_n, \dots$ tvoří posloupnosti asymptoticky nestranných odhadů konstanty μ .

d) Rozhodněte, zda posloupnosti: $M_1, M_2, \dots, M_n, \dots$ a $L_1, L_2, \dots, L_n, \dots$ tvoří posloupnosti konzistentních odhadů konstanty μ .

Příklad 2

pomocí sw.:

V přednášce bylo uvedeno, že výběrová distribuční funkce je pro pevné x nestranným odhadem (teoretické) distribuční funkce, tedy $E(F_n(x)) = F(x)$ pro lib. x . Ilustrujte pomocí náhodně vygenerovaných čísel z norm. standardizovaného rozložení.

1. Vytvořte nový datový soubor o třech proměnných a sto případech.
2. Do první proměnné vložte náhodně vygenerovaná čísla z normálního standardizovaného rozložení a uspořádejte je podle velikosti
3. Do druhé proměnné vložte hodnoty výběrové distribuční funkce $F_n(x)$ odpovídající jednotlivým hodnotám vygenerovaných čísel.
4. Do třetí proměnné vložte hodnoty (teoretické) distribuční funkce $F(x)$ odpovídající jednotlivým hodnotám vygenerovaných čísel.
5. Sestrojte bodový graf, kde na x -ové ose budou vygenerované hodnoty a na y -ové ose budou znázorněny hodnoty jak výběrové, tak teoretické distribuční funkce. Editujte obrázek tak, že hodnoty distribuční funkce zůstanou jako izolované body, zatímco hodnoty teoretické distribuční funkce spojte spojnicemi. Interpretujte obrázek.

Příklad 3

ručně

Konstantní rychlost letadla μ byla odhadována v pěti měřeních (které jsou v datovém souboru *letadlo_sta*) a z jejich výsledků byl vypočten průměr $m = 870,3 m/s$. Najděte 95% interval spolehlivosti pro neznámou rychlost μ , je-li známo, že směrodatná odchylka měření rychlosti je $\sigma = 2,1 m/s$ a výsledky měření jsou v souladu s normálním rozložením. (Dolní a horní odhad odvoďte. V dalších cvičeních již odhady nebudeme odvozovat, budeme dosazovat do již odvozených vzorců.)

Příklad 4

pomocí sw.:

Budeme sledovat vliv rozsahu výběru n na šířku intervalu spolehlivosti pro μ při pevném $\alpha = 0,05$. Uvažujme 41 hypotetických náhodných výběrů různých rozsahů $n = 5, 7, 9, \dots, 85$ z rozložení $N(0,1)$. Tyto výběry jsou takové, že jejich výběrové průměry se vždy realizovaly hodnotou 0. Pro každý výběr vypočítejte dolní a horní mez 95%-ního intervalu spolehlivosti pro μ . Graficky znázorněte závislost těchto mezí na rozsahu n a sledujte šířku intervalu spolehlivosti v závislosti na rozsahu n výběrového souboru.

1. Nejdříve vytvořte nový datový soubor o 3 proměnných a 41 případech. Potom, po spuštění programu *intsp1.svb*, který si stáhnete ze studijních materiálů, se do proměnné *v1* uloží rozsahy výběrů 5, 7, ..., 85; do *v2* se uloží dolní meze 95%-ního intervalu spolehlivosti pro μ a do *v3* horní meze int. sp. pro μ . Tyto meze jsou pro $\alpha = 0,05$, realizaci $m = 0$ a známé $\sigma = 1$ spočteny ze vzorců: $d = m - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$, $h = m + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$
2. Sestrojte graf, kde na *x*-ové ose budou rozsahy výběrových souborů a na *y*-ové ose budou jak horní, tak dolní odhady int. sp. pro μ v závislosti na *n*. Jak spolu souvisí rozsah *n* a šířka intervalu spolehlivosti?

Příklad 5

pomocí sw.:

Budeme sledovat vliv velikosti rizika α na šířku intervalu spolehlivosti pro μ při pevném $n = 25$.

1. Nejdříve vytvořte nový datový soubor o 3 proměnných a 20 případech. Potom, po spuštění programu *intsp2.svb*, který si stáhnete ze studijních materiálů, se do proměnné *v1* uloží hodnoty rizika $\alpha = 0,20; 0,19; \dots; 0,01$; do *v2* se uloží dolní meze $100(1 - \alpha)\%$ -ního intervalu spolehlivosti pro μ a do *v3* horní meze int. sp. pro μ . Tyto meze jsou pro $n = 25$, realizaci $m = 0$ a známé $\sigma = 1$ spočteny ze vzorců: $d = m - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$, $h = m + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}$
2. Sestrojte graf, kde na *x*-ové ose budou hodnoty rizika α a na *y*-ové ose budou jak horní, tak dolní odhady int. sp. pro μ v závislosti na α .

Příklady k samostatnému řešení:

Příklad 6

Nechť X_1, \dots, X_n je náhodný výběr z rozložení $Rs(0, b)$, kde $b > 0$ je neznámý parametr. Jsou definovány statistiky $T_1 = X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3 + \frac{1}{6}X_4$ a $T_2 = \frac{1}{2}(X_1 + X_2 + X_3 + X_4)$.

- a) Ukažte, že T_1 a T_2 jsou nestranné odhady parametru b .
- b) Rozhodněte, který odhad je lepší.
- c) Jaký jiný odhad byste sami navrhli?
Označme $X_{(n)} = \max\{X_1, \dots, X_n\}$. Ukažte, že statistika $T_3 = X_{(n)}$ je konzistentním odhadem parametru b .

Příklad 7

Řeste 2. příklad pro $n=1000$.

Příklad 8

11.C,11.D

SW. návody ke 3. cvičení

Návod k řešení příkladu 2:

ad 2) Do Dlouhého jména proměnné napíšeme =RndNormal(1). (Funkce *RndNormal(x)* generuje náhodná čísla z normálního rozložení se střední hodnotou 0 a rozptylem x .) Po naplnění proměnné $v1$ ji setřídíme podle velikosti: roletka Data-Setřídít...

ad 3) Do Dlouhého jména proměnné napíšeme = $v0/100$. $v0$ je označení pro šedý sloupec pořadových čísel případů, 100 je počet všech vygenerovaných dat.

ad 4) Do Dlouhého jména napíšeme příkaz =INormal($v1;0;1$). (Funkce *INormal(v1;0;1)* počítá hodnotu distribuční funkce standardizovaného normálního rozložení v bodě $v1$.)

ad 5) Grafy - Bodové grafy - Typ grafu Vícenásobný - vypneme Lineární proložení - Proměnné X: $v1$; Y: $v2,v3$; OK, OK

Editace obrázku: poklepeme na značku odpovídající proměnné $v3$ v pravém dolním rohu obrázku; odškrtneme značky, zaškrtneme spojnice a vybereme jako spojnicí spojitou čáru-zavřít-OK

Návod k řešení příkladu 4:

ad 2) Grafy - Bodové grafy - Typ grafu Vícenásobný - vypneme Lineární proložení - Proměnné X: $v1$; Y: $v2, v3$; OK, OK. (Chceme-li přidat přímku vyjadřující hodnotu výběrového průměru $m = 0$, do tabulky vložíme novou proměnnou, kterou naplníme nulami. Do výše uvedeného grafu přidáme na y -ovou osu i tuto novou proměnnou a editací obrázku izolované body nové proměnné nahradíme spojnicemi.)

Návod k řešení příkladu 5:

Analogicky jako v předchozím příkl.

4. cvičení

Příklad 1

a) *ručně*

Hloubka moře se měří přístrojem, jehož systematická chyba je nulová a náhodné chyby měření mají normální rozložení se směrodatnou odchylkou $\sigma = 1 \text{ m}$. Kolik měření je nutno provést, aby se hloubka stanovila s chybou nejvýše $\pm 0,25 \text{ m}$ při riziku 0,05? (Tedy aby šířka intervalu spolehlivosti pro skutečnou hloubku moře byla nejvýše $2 \cdot 0,25 \text{ m}$.)

b) *pomocí sw*

Řekněme, že skutečná hloubka moře na daném místě je $\mu = 1000 \text{ m}$, tedy jednotlivá měření pochází z rozložení $N(1000, 1)$. Dle řešení a) je potřeba provést alespoň 62 měření, aby šířka intervalu spolehlivosti pro skutečnou hloubku μ byla nejvýše půl metru. S pravděpodobností 95% by hloubka 1000 m měla být uvnitř odhadnutého intervalu spolehlivosti. Vygenerujte náhodný výběr rozsahu 62 z rozložení $N(1000, 1)$, spočítejte odpovídající dolní a horní odhad intervalu spolehlivosti, zkontrolujte jeho šířku a skutečnost, zda 1000 m leží uvnitř IS. Zjistěte, kolik studentů v seminární skupině (jestli vůbec nějaký) má skutečnou hloubku 1000 m mimo interval spolehlivosti. Jaký počet "špatných" intervalů spolehlivosti se dá přibližně očekávat?

Příklad 2

Předpokládáme, že u jisté společnosti, která doručuje zásilky ve velkém městě, se čas doručení řídí normálním rozložením. Náhodně vybereme 100 zásilek u kterých zaznamenáme čas potřebný k doručení. Výběrový průměr vypočítaný z tohoto výběrového souboru je $m = 31,5$ minut. Z minulosti víme, že směrodatná odchylka času doručení je 5 minut. Na základě těchto předpokladů určete na hladině významnosti 0,05 horní odhad pro stření hodnotu času doručení. *Po odvození horního odhadu proveďte samotný výpočet pomocí sw.*

Příklad 3

Před zahájením výroby nového pracího prášku výrobce provedl mezi sty náhodně vybranými zákazníky anketu. Její součástí bylo ohodnocení kvality nového prášku na stobodové škále. Sestrojte 95%-ní interval spolehlivosti pro neznámou střední hodnotu μ bodového hodnocení pracího prášku. (Na základě dřívějších zkušeností předpokládáme, že směrodatná odchylka bodového hodnocení je $\sigma = 8$.) Řešení proveďte *pomocí sw*; jednotlivé odpovědi ankety jsou uloženy v souboru *prací_prášek.sta*. Zadání příkladu nepředpokládá normalitu náhodného výběru, proto ji alespoň orientačně posuďte pomocí histogramu. (Testům normality se budeme věnovat později.)

Příklad 4

Výrobce nového lacinějšího typu žárovek tvrdí, že jeho výrobky jsou přesně tak spolehlivé, jako cenově vyšší žárovky konkurenčního podniku s deklarovanou střední životností 5000 hodin. Pro ověření výrobcova tvrzení jsme vybrali 50 žárovek a sledovali jejich životnost. Jejich průměrná životnost byla 5100 hodin. Na hladině významnosti 0,05 ověřme, že je výrobcovo tvrzení správné. Předpokládáme, že náhodný výběr pochází z normálního rozložení a $\sigma = 500$ hodin. Řešení proveďte pomocí

- a) intervalu spolehlivosti
- b) kritického oboru

c) p-hodnoty *ručně i softwarem*

Poznámka: Všechny dosud uvedené příklady tohoto cvičení předpokládaly, že náhodný výběr pochází z normálního rozložení s neznámou střední hodnotou μ a známou směrodatnou odchylkou σ . Jelikož tato situace v praxi nastává málokdy (obvykle σ není známé), sw. Statistica nemá implementovaný interval spolehlivosti a test hypotézy o parametru μ přímo a výpočty bylo potřeba provádět pomocí "dlouhých jmen". Situace, které jsou v praxi běžné, umí Statistica řešit elegantněji.

Příklad 5

Eskalátor je navržen s celkovou nosností 8000 kg. Jeho plánovaná kapacita je 100 lidí. Předpokládejme, že hmotnost člověka má normální rozložení se střední hodnotou 75 kg a směrodatnou odchylkou 10 kg. Jaká je pravděpodobnost, že náhodný výběr 100 lidí bude mít celkovou hmotnost vyšší, než je povolené zatížení eskalátoru? *Řešte ručně i pomocí pravděpodobnostního kalkulátoru.*

(Řešení porovnejte s 5. příkladem z 2. cvičení; rozlišujte mezi standardizací výběrového průměru a výběrového úhrnu.)

Příklady k samostatnému řešení:

Příklady z učebnice 11.B-11.E

SW. návody k 4. cvičení

Návod k řešení příkladu 1:

1. Vytvoříme nový datový soubor o 1 proměnné nazvané *měření* a 62 případech. Do dlouhého jména proměnné *měření* zapíšeme výraz `=rndnormal(1)+1000` - OK.
2. Určíme realizaci výběrového průměru *m*: Statistiky - Základní statistiky a tabulky - Popisné statistiky; na záložce "Detailní výsledky" necháme zaškrtnutý pouze průměr - OK
3. Z pracovního sešitu vybereme výstupní tabulku s průměrem jako samostatné okno: Nastavíme se kurzorem na tabulku s průměrem - klik pravým tlačítkem myši - Extrahovat jako samostatné okno - Kopie. (Výstupní tabulku teď použijeme jako datovou tabulku s jedním řádkem.) Přidáme 3 proměnné pro dolní odhad *d*, horní odhad *h* a pro šířku intervalu spolehlivosti *delta*. (Data - Proměnné - Přidat ...)
Do dlouhého jména proměnné *d* zapíšeme výraz `=Průměr-1/sqrt(62)*VNormal(0,975;0;1)` - OK.
Do dlouhého jména proměnné *h* zapíšeme výraz `=Průměr+1/sqrt(62)*VNormal(0,975;0;1)` - OK.
Do dlouhého jména proměnné *delta* zapíšeme výraz `=h-d` - OK.
Funkce Vnormal(x;0;1) počítá x-quantil rozložení $N \sim (0, 1)$.

Návod k řešení příkladu 2:

Vytvoříme datový soubor o 1 proměnné (nazvané *h*) a jednom případu. Do dlouhého jména proměnné *h* zapíšeme výraz `= 31,5 + Vnormal(0,95;0;1)*5/sqrt(100)`.

Návod k řešení příkladu 3:

1. Orientačně posoudíme normalitu:
Grafy - Histogramy - proměnné: *body* OK - Typ proložení: normální OK.
2. Určíme průměr náhodného výběru:
"Vybarvíme" sloupec *body* - Statistiky - Statistika bloku dat - Blok sloupců - Průměry OK
3. Přidáme dvě proměnné *d* a *h* pro dolní a horní odhad.
Data - Proměnné - Přidat ...
4. Do dlouhého jména proměnných *d* a *h* zapíšeme výrazy pro výpočet odhadů.

Návod k řešení příkladu 4:

ad c)

Statistiky - Pravděpodobnostní kalkulátor - Rozdělení - Z (normální); zaškrtneme: 1-kumul p; oboustranné; X:1,41; průměr:0; SmOdch:1 - Výpočet. P-hodnota je v okýnku p:

5. cvičení

Příklad 1

Z populace stejně starých selat téhož plemene bylo vylosováno 6 selat a po dobu půl roku jim byla podávána stejná výkrmná dieta. Byly zaznamenány průměrné denní přírůstky v dg, které jsou v datovém souboru *Prasata_I.sta*. Z dřívějších pokusů je známo, že v populaci mívají takové přírůstky normální rozložení. Při riziku $\alpha = 0,05$ odvoďte:

a) dolní odhad střední hodnoty μ ; b) intervalový odhad směrodatné odchylky σ .

- I. ručně (Realizace výběrových charakteristik a kvantily možno i softwarem, ; kvantily nejrychleji pomocí pravděpodobnostního kalkulátoru.)
- II. pomocí sw.

Příklad 2

Mlékárna má zájem na udržení horní hranice variability obsahu tuku (udávaného v procentech) ve svém mléku. Jestliže předpokládaný obsah tuku je μ %, potom aktuální obsah tuku v kartónu mléka by se neměl příliš odchýlovat od této hodnoty. Pro mlékárnu je přijatelná směrodatná odchylka nejvýše $\sigma = 0,1$ % tuku v mléce. Vybrali jsme náhodně 20 kartónů mléka a naměřili jsme procentuální hodnoty obsahu tuku v mléku, které jsou v datovém souboru *Mléko.sta*.

- a) Pomocí kritického oboru testujte (ručně) na hladině $\alpha = 0.05$ hypotézu $H_0 : \sigma \leq 0.1$ proti pravostranné alternativě $H_1 : \sigma > 0.1$. (Test není implementován, sw lze případně využít jako inteligentní kalkulačku - do dlouhého jména zadat tvar vzorce testové statistiky...)
- b) Stejnou hypotézu testujte softwarem pomocí intervalu spolehlivosti.
- c) Označuje-li X obsah tuku v mléce, určete pravděpodobnost

$$P(|X - \mu| > \sigma)$$

$$P(|X - \mu| > 3\sigma)$$

JE POTŘEBA OVĚŘIT NORMALITU. (Zatím alespoň orientačně histogramem.)

Příklad 3

Pro data z příkladu 1. testujte hypotézu, že střední hodnota váhových přírůstků populace, z níž pochází náhodný výběr je 61 Dg. Test proveďte na hladině významnosti 5%. (Ručně i sw.) Interpretuje softwarem získanou p-hodnotu.

Příklad 4

Bylo vylosováno 6 vrhů selat a z nich vždy dva sourozenci. Jeden z nich dostal dietu č. 1 a druhý dietu č. 2 (volba diet v páru byla náhodná). Váhové přírůstky v dg. jsou v datovém souboru *Prasata_II.sta*. Sestrojte 95% interval spolehlivosti pro $\mu = \mu_1 - \mu_2$. (Motivace k dalšímu příkladu: Která dieta se jeví lepší? Proč je vhodné hledat odpověď na tuto otázku přes náhodný výběr "dvojčat"?)

Příklad 5

Pro data z příkladu 4. testujte na hladině významnosti 0,05 hypotézu, že obě výkrmné diety mají stejný vliv. (Ručně i sw.) Interpretuje softwarem získanou p-hodnotu.

Příklad 6

Je známo, že výška hochů ve věku 9,5 až 10 let má normální rozložení. Dětský lékař náhodně vybral 15 hochů uvedeného věku, změřil je a vypočítal realizaci výběrového průměru $m = 139,13$ cm a realizaci výběrového rozptylu $s^2 = 39,2$ cm². Podle jeho názoru by výška hochů v tomto věku neměla přesáhnout 142 cm s pravděpodobností aspoň 0,95. Lze tvrzení lékaře akceptovat? (Testování proveďte pomocí intervalu spolehlivosti i pomocí kritického oboru.)

doma k procvičení: příklady 13.A, 13.B

SW. návody k 5. cvičení

Návod k řešení příkladu 1:

ad a) STATISTICA má implementovaný výpočet oboustranného intervalu spolehlivosti pro parametr μ , když σ není známé. V naší úloze potřebujeme ale levostranný interval spolehlivosti na hladině 5%. Proto softwarem určíme oboustranný interval spolehlivosti na hladině 10% a dolní mez 10-ti %ního oboustranného intervalu spolehlivosti bude stejná, jako dolní mez 5-ti %ního levostranného intervalu spolehlivosti.

Roletka: Statistika - Základní statistiky a tabulky - Popisné statistiky - záložka: Detailní výsledky; zaškrtneme: Meze spolehl. prům.; nastavíme 90%. Vybereme proměnnou PRIRUSTKY - Výpočet.

ad b) ...záložka: Detailní výsledky; zaškrtneme: Meze sp. směr. odch.; nastavíme 95%....

Návod k řešení příkladu 2:

Otevřeme datový soubor *Mléko.sta* a orientačně posoudíme normalitu. Grafy - Histogramy...

ad b) Roletka: Statistika - Základní statistiky a tabulky - Popisné statistiky - záložka: Detailní výsledky; zaškrtneme: Meze sp. směr. odch.; nastavíme 90%. Vybereme proměnnou TUK - Výpočet. Kterou z uvedených mezí použijeme?

ad c) V pravděpodobnostním kalkulátoru pro normální rozložení zaškrtneme: oboustranné a (1-kumul. p); okýnko X : vyplníme 1, případně 3 - (již po standardizaci!), proto i parametry Průměr:0; Sm.Odch.:1 - Výpočet.

Návod k řešení příkladu 3:

Statistika - Základní statistiky a tabulky - t-test, samost. vzorek - Proměnné:PRIRUSTKY - Test všech průměrů vůči: 61 - Výpočet t-testy. (V záložce detailní výsledky lze nastavit "p-hodnota pro zvýraznění", která je implicitně nastavená pro hladinu $\alpha = 0,05$. Při zamítnutí hypotézy na nastavené hladině α se výsledky zobrazí červeně.)

Návod k řešení příkladu 4:

Nejdříve vytvoříme rozdílový náhodný výběr a jeho hodnoty vložíme do nové proměnné. Data - Proměnné - Přidat. Nazveme ji ROZDIL, do dlouhého jména: =v1-v2 OK

Roletka: Statistika - Základní statistiky a tabulky - Popisné statistiky - záložka: Detailní výsledky; zaškrtneme: Meze spolehl. prům.; nastavíme 95%. Vybereme proměnnou ROZDIL - Výpočet.

Návod k řešení příkladu 5:

Máme dvě možnosti:a) buď pomocí rozdílového náhodného výběru, nebo b) přímo z dvojrozměrného výběru.

ad a) Statistika - Základní statistiky a tabulky - t-test, samost. vzorek - Proměnné:ROZDIL - Test všech průměrů vůči: 0 - Výpočet t-testy.

ad b) Statistika - Základní statistiky a tabulky - t-test, závislé vzorky - OK - Proměnné: 1.seznam proměnných: DIET1; 2.seznam proměnných: DIET2 - OK - Výpočet t-testy.

Interpretujte p-hodnotu a rozhodněte o nulové hypotéze.

6. cvičení

Načtete datový soubor žáci.sta, kde interpretace proměnných je následující:

SEX	Pohlaví (1 – chlapec, 2 – dívka)
IQ_VERB	Verbální IQ
IQ_PERF	Performační IQ
IQ_CELK	Celkové IQ
TRIDA	Třída (1. až 9.)
VZDEL_M	Vzdělání matky (1 – základní, 2 – SŠ, 3 – VŠ)
VZDEL_O	Vzdělání otce (1 – základní, 2 – SŠ, 3 – VŠ)
SIDLO	Sídlo (1 – město, 2 – venkov)

Příklad 1

Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty performační a verbální inteligence se neliší. Dvojrozměrnou normalitu výběru budeme předpokládat, jelikož ji zatím neumíme ověřit.

- Ručně, ovšem všechny potřebné realizace číselných charakteristik a kvantily naleznete pomocí sw.
- Pomocí sw. Ve výstupu naleznete realizaci testové statistiky a interpretujte p-hodnotu.

Příklad 2

Pomocí sw. sestrojte 95% interval spolehlivosti pro střední hodnotu celkového IQ a to zvlášť pro chlapce a zvlášť pro dívky. Nezapomeňte ověřit (zatím jen histogramem) normalitu obou výběrů. Interval spolehlivosti doplňte krabicovými diagramy.

Příklad 3

Testujte (nejdříve ručně) na hladině 1% hypotézu, že rozptyl celkového IQ u mužů a u žen je stejný. Test pomocí sw. bude součástí řešení 4b).

Příklad 4

Testujte na hladině 1% hypotézu, že celkové IQ se u mužů a žen neliší.

- Ručně, ovšem všechny potřebné realizace číselných charakteristik a kvantily naleznete pomocí sw.
- Pomocí sw. Ve výstupu naleznete realizace testových statistik a interpretujte p-hodnoty pro t-test a F-test.

Příklad 5

Pomocí sw. sestrojte 99% interval spolehlivosti pro rozdíl středních hodnot celkového IQ mužů a celkového IQ žen.

Příklad 6

Jsou dány dva nezávislé náhodné výběry, první pochází z rozložení $N(2; 1, 5)$ a má rozsah 10, druhý pochází z rozložení $N(3; 4)$ a má rozsah 5. Jaká je pravděpodobnost, že výběrový průměr 1. výběru bude menší než výběrový průměr 2. výběru?

doma k procvičení**Příklad 7**

Bylo vylosováno 11 stejně starých selat téhož plemene. Šesti z nich byla předepsána výkrmná dieta č.1 a zbylým pěti výkrmná dieta č.2. Průměrné denní přírůstky v D_g za dobu půl roku jsou následující:

dieta č.1: 62, 54, 55, 60, 53, 58

dieta č.2: 52, 56, 49, 50, 51.

Zjištěné hodnoty považujeme za realizace dvou nezávislých náhodných výběrů pocházejících z rozložení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$.

- a) Sestrojte 95% empirický interval spolehlivosti pro podíl rozptylů.
- b) Za předpokladu, že data pocházejí z rozložení $N(\mu_1, \sigma^2)$ a $N(\mu_2, \sigma^2)$, sestrojte 95% empirický interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$.

Příklad 8

Pro údaje ze 7. příkladu testujte na hladině významnosti 0,05 hypotézu, že

- a) rozptyly hmotnostních přírůstků selat při obou výkrmných dietách jsou shodné
- b) obě výkrmné diety mají stejný vliv na hmotnostní přírůstky selat.

Test proveďte pomocí intervalu spolehlivosti i pomocí testového kritéria a kritického oboru.

Příklad 9

Učebnice 13.C-13.E

SW. návody k 6. cvičení

Návod k řešení příkladu 1:

ad a) Vytvoříme novou proměnnou ROZDIL=IQ_VERB-IQ_PERF. Potřebné realizace číselných charakteristik najdeme buď 1) Statistika - Základní statistiky - Popisné statistiky; nebo nejrychleji 2) Poklepeme na hlavičku nové proměnné a v otevřeném okně klikneme na tlačítko Hodn./Statist. Ve výstupu jsou potřebné realizace číselných charakteristik. Kvantily nalezneme pomocí pravděpodobnostního kalkulátoru rozdělení - t (Studentovo) - stupně volnosti doplníme do okýnka sv; hladinu kvantilu do okýnka p.

ad b) Statistika - Základní statistiky a tabulky - t-test, závislé vzorky - OK - Proměnné: 1. seznam prom: IQ_VERB; 2. seznam prom: IQ_PERF - OK - Výpočet t-testy.

Návod k řešení příkladu 2:

Histogram pro chlapce: Grafy - Histogramy - Proměnné: IQ_CELK - tlačítko Filtr případů - Zapnout filtr - některé, vybrané výrazem: SEX=1 - OK - OK. Stejným způsobem i histogram pro dívky.

Interval spolehlivosti pro střední hodnotu celkového IQ chlapců: Statistika - Základní statistiky - Popisné statistiky - Proměnné: IQ_CELK - tlačítko Filtr případů - Zapnout filtr - některé, vybrané výrazem: SEX=1 - OK - záložka Detailní výsledky: zaškrtnout meze spolehlivosti prům. - Výpočet. Analogicky pro dívky.

(V záložce detailní výsledky se hodí zaškrtnout i rozptyl, který bude potřeba ve 3. příkladě.)

Krabicový graf: Grafy - 2D grafy - Krabicové grafy - Proměnné: Závislé prom: IQ_CELK; Grupovací prom: SEX - OK - Střední bod - Hodn: průměr OK. (V záložce detaily lze nastavit požadovaný vzhled krabicového diagramu, bližší vysvětlení v Rejstříku pod heslem "Outliers and Extremes".)

Návod k řešení příkladu 3:

Kvantily Fisher-Snedocerova rozložení opět v pravděpodobnostním kalkulátoru, F(Fisherovo)...

Návod k řešení příkladu 4:

ad a) Při výpočtu realizací číselných statistik použijte filtr případů.

ad b) Statistika - Základní statistiky a tabulky - t-test, nezávislé, dle skupin - Proměnné: Závislé proměnné: IQ_CELK; Grupovací proměnná: SEX - OK - Kód pro skup.1: 1; Kód pro skup.2: 2; - (Zkontrolovat, že není zapnutý filtr případů) - Výpočet t-testy.

Návod k řešení příkladu 5:

Statistika - Základní statistiky a tabulky - t-test, nezávislé, dle skupin - záložka Možnosti - zaškrtnout meze spol. pro odhady, ostatní viz. předchozí návod.

7. cvičení

Příklad 1

V každém odvětví jsou určité obecně přijatelné obchodní praktiky. V náhodném souboru 73 bankéřů byla položena tato otázka: "Jsou ve Vašem odvětví některé praktiky, které lze považovat za neetické?" V tomto souboru bylo zjištěno 39 odpovědí "ne". Najděte 95% - ní interval spolehlivosti pro podíl bankéřů v populaci, kteří nepovažují žádnou z obecně přijatelných obchodních praktik v jejich odvětví za neetickou. IS odvoďte ručně i pomocí sw.

Příklad 2

Pro vstup do parlamentu potřebují politické strany získat více, než 5% hlasů voličů. Mezi 1000 dotazovanými osobami v předvolebním průzkumu se pro jistou stranu vyslovilo 60 osob. Na asymptotické hladině významnosti $\alpha = 0,05$ testujte hypotézu, že se uvažovaná strana do parlamentu nedostane. Test pomocí testovacího kritéria proveďte ručně; pomocí IS lze softwarem i bez datového souboru.

Příklad 3

Načtete datový soubor *žáci.sta*. Pro celý soubor 856 dětí vypočtete průměr proměnné IQ_CELK. Na asymptotické hladině významnosti 0,05 testujte hypotézu, že podíl dětí s nadprůměrným IQ_CELK je mezi městskými a venkovskými dětmi stejný. Test proveďte ručně i pomocí sw.

Příklad 4

Při průzkumu trhu, když jednotlivcům nebo domácnostem rozesíláme dotazníky, je důležité získat co největší počet odpovědí. Jedna z možností, jak zvýšit zájem domácností o vyplnění dotazníků je v úvodu uvést pohnutky vedoucí k anketě.

Dotazníky obsahující uvedení pohnutek byly rozeslány náhodnému výběru 250 domácností a bylo získáno 101 zodpovězených dotazníků. Obdobné dotazníky bez uvedení pohnutek byly rozeslány náhodnému výběru 250 domácností a bylo získáno 75 zodpovězených dotazníků. Na hladině 5% testujte hypotézu, že podíl zodpovězených dotazníků s úvodní motivací je stejný, nebo větší, než podíl zodpovězených dotazníků bez úvodní otázky. Test proveďte pouze pomocí sw.

Pozor, při výpočtu p -hodnoty testu s levostrannou alternativou je kritický obor soustředěný v levé části rozložení a proto $p = P(T_0 \leq t_0)$. Sw vždy počítá $p = P(T_0 \geq |t_0|)$. Proto je potřeba si rozmyslet, zda rozdíl $m_1 - m_2$ je kladné, nebo záporné číslo a jestli je p -hodnota soustředěná v pravé, nebo levé části rozložení. V našem příkladě získáme p -hodnotu jako (1 - sw. p jednostr.)

doma k procvičení:

Příklad 5

V rámci celosvětového průzkumu, týkajícího se srovnávání kvality života v různých zemích, byla pro zjištění názoru o zločinnosti pokládána tato otázka: "Bojíte se v noci jít po málo frekventovaných ulicích?" Odpovědi jsou uvedeny v následující tabulce. Za předpokladu, že průzkum v každé zemi má stejnou přesnost a stejný rozsah náhodného výběru $n = 300$, sestrojte 95% interval spolehlivosti pro rozdíl podílu odpovědí "Ano" mezi

- a) USA a Japonskem
- b) USA a Latinskou Amerikou

Odpověď	USA	Japonsko	Latinská Amerika
Ano	40%	33%	57%
Ne	56%	63%	42%
Žádný názor	4%	4%	1%

Příklad 6

V prodejně mají zjištěno, že 25% nákupů přesahuje 300 Kč. Při nákupu nad 300 Kč obdrží zákazník slosovateľný kupón. Jaká je pravděpodobnost, že mezi 100 zákazníky bude rozdáno 20 až 25 kupónů? Výpočet proveďte a) přesně, b) pomocí aproximace normálním rozložením.

Příklad 7

Učebnice 14.A až 14.F

SW. návody k 7. cvičení

Návod k řešení příkladu 1:

$$n = 73, m = 39/73 = 0,53425, \alpha = 0,05$$

Statistiky - Analýza síly testu - levé okno: odhad intervalu; pravé okno: Jeden podíl, Z, Chí-kvadrát test - OK - Pozorovaný podíl p: 0,53425; Velik. Vzorku (N): 73; Spolehlivost: 0,95 - Vypočítat. Ve výstupu jsou 3 intervaly spolehlivosti, z nichž třetí (Pí (původ)) odpovídá tvaru z přednášky.

Návod k řešení příkladu 3:

1. Nejdříve určíme průměr proměnné IQ_CELK
2. Určíme počet dětí, které bydlí ve městě a na venkově
Statistiky - Zákl. statistiky a tabulky - Tabulky četností - OK - Proměnné: SIDLO - OK (V záložce možnosti můžeme zrušit všechny další přednastavené četnosti - nejsou pro řešení potřebné) - Výpočet.
3. Určíme počet městských dětí, které mají nadprůměrné IQ a obdobně venkovských dětí, které mají nadprůměrné IQ.
Statistiky - Zákl. statistiky a tabulky - Tabulky četností - OK - Proměnné: IQ_CELK - Select cases - Zapnout filtr - některé vybrané výrazem IQ_CELK > vypočtený průměr proměnné IQ_CELK - záložka: detailní výsledky - Uživatel. definov. kateg. - klik na tlačítko vedle - Kategorie 1: Zahrnout, když: SIDLO=1; Kategorie 2: Zahrnout, když: SIDLO=2 - OK - Výpočet. (V záložce možnosti můžeme zrušit všechny další přednastavené četnosti.)
4. Nalezneme p -hodnotu testu o shodě "teoretických podílů" $\vartheta_1 = \vartheta_2$
Statistiky - Zákl. statistiky a tabulky - Testy rozdílů: r, %, průměry - OK - Rozdíly mezi dvěma poměry: N_1, N_2 jsou rozsahy výběrů měšťáků a venkovanů, P_1, P_2 jsou dle značení přednášky M_1, M_2 (tedy relativní četnosti nadprůměrných měšťáků a nadprůměrných venkovanů) - oboustr. - výpočet. Interpretujte p -hodnotu. (p -hodnota je softwarem počítaná z **absolutní hodnoty** testovacího kritéria v poznámce 8.11.)
5. Dodatečně určíme realizaci testové statistiky pomocí pravděpodobnostního kalkulátoru. V pravděpodobnostním kalkulátoru normálního rozložení zaškrtneme Inverze, Oboustr., (1-kumul p), a do okna p: doplníme již vypočtenou p -hodnotu testu - Výpočet. V okně X je realizace testové statistiky. (V důsledku zaokrouhlování je mírně zkreslená.)

8. cvičení

Příklad 1

Jsou známy měsíční tržby (v tisících Kč) tří prodavačů za dobu půl roku.

1. prodavač:	12	10	9	10	11	9
2. prodavač:	10	12	11	12	14	13
3. prodavač:	19	18	16	16	17	15

Data jsou uložena v souboru *Měsíční_tržby.sta*

Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty tržeb všech tří prodavačů jsou stejné. (Předpokládáme, že výběry jsou navzájem nezávislé.) Pokud zamítnete nulovou hypotézu, zjistěte, tržby kterých dvou prodavačů se liší na hladině významnosti 0,05. Podrobné instrukce:

- Sestrojte model, rozmyslete si předpoklady modelu, formulujte hypotézu.
- Sestrojte krabicový diagram, interpretujte jej. Vytvořte tabulku popisných statistik, která obsahuje průměry, směr. odchylky a četnosti pro tržby jednotlivých prodavačů.
- Pomocí sw. ověřte předpoklady normality 1.) testy, 2.) vizuálně pomocí grafů.
- Pomocí sw. ověřte předpoklad shody rozptylů.
- Nejdříve ručně (s podklady k 8. přednášce), potom pomocí sw. testujte hypotézu o shodě středních hodnot.
- V případě zamítnutí hypotézy o shodě středních hodnot zjistěte, které dvojice středních hodnot zamítnutí způsobily.

Příklad 2

Načtěte soubor *žáci.sta*. Pomocí sw. testujte hypotézu, že vzdělání matky nemá vliv na celkové IQ. Pokud zamítnete nulovou hypotézu, zjistěte, IQ kterých dětí (vzhledem k vzdělání matky) se liší na hladině významnosti 0,05.

Příklad 3

Je dáno pět nezávislých náhodných výběrů o rozsazích 5, 7, 6, 8, 5, přičemž i -tý výběr pochází z rozložení $N \sim (\mu_i, \sigma^2)$, $i = 1, \dots, 5$. Byl vypočten celkový součet čtverců $S_T = 15$ a reziduální součet čtverců $S_E = 3$. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

Příklad 4

Je dána neúplná tabulka ANOVA. Místo otazníků doplňte chybějící čísla.

zdroj variability	součet čtverců	stupně volnosti	podíl	FA
skupiny	?	2	?	?
reziduální	16,033	?	?	-
celkový	17,301	35	-	-

doma k procvičení:

Příklad 5

Pomocí sw. testujte hypotézu, že vzdělání otce nemá vliv na celkové IQ.

Příklad 6

Učebnice 15.A-15.E

SW. návody k 8. cvičení

Návod k řešení příkladu 1:

- b) **Krabicové grafy:** Statistika - Základní statistiky/tabulky - Rozklad & jednofakt. ANOVA - Proměnné - Závislé: TRZBY; Grupovací: PRODAVAC - OK - Kódy pro grupovací proměnné: Vše - OK - OK - otevře se okno: Statistika dle skupin. Záložka: Základní výsledky - Kategoriz. krabicový graf
Tabulka popisných statistik: Návrat do okna Statistika dle skupin - záložka: Základní výsledky - Výpočet: Tabulka statistik (zobrazí se průměry, směrodatné odchylky a rozsahy všech tří výběrů).
- c) **Testy normality:** Grafy - Histogramy - Proměnné: TRZBY - OK - záložka: Kategorizovaný - Kategorie X - zapnuto; Změnit proměnnou - PRODAVAC - OK - záložka: Detaily - zaškrtnout Shapiro-Wilksův test a Kolmogorov-Smirnovův test - OK. Ve výstupu jsou kromě histogramů i p -hodnoty obou testů pro jednotlivé výběry. Další cesty k oběma testům jsou:
1) Statistika - Základní statistiky/tabulky - Tabulky četností - OK - Proměnné ... - OK - Normalita - zaškrtneme Lilieforsův test a S-W test - Testy normality.
2) Statistika - Základní statistiky/tabulky - Popisné statistiky - OK - Proměnné ... - OK - Normalita - zaškrtneme K-S test & Lilieforsův test a S-W test - Tabulky četností (nebo Histogram).
Tímto postupem ovšem nelze kategorizovat proměnnou TRZBY podle PRODAVAC, takže testy je potřeba provádět natřikrát pomocí "select cases".
Vizuální posouzení normality pomocí grafů: Návrat do okna Statistika dle skupin (př.1b) - záložka: ANOVA & testy - Kategoriz. normál. pravd. grafy.
Jiná cesta k normálnímu pravděpodobnostnímu grafu je: Grafy - 2D Grafy - Normální pravděpodobnostní grafy - Proměnná ... - odškrtneme Neurčit průměrnou pozici svázaných pozorování - OK.
Pokud se body grafu vyskytují v těsné blízkosti přímky, lze usuzovat, že výběr pochází z normálního rozložení.
(K vizuálnímu posouzení normality lze také použít P-P plot a Q-Q plot, Grafy - 2D Grafy - Grafy typu Q-Q, nebo P-P.)
- d) **Předpoklad shody rozptylů:** návrat do okna Statistika dle skupin - záložka: ANOVA & testy - Leveneovy testy.
- e) **Test shody středních hodnot:** návrat do okna Statistika dle skupin - záložka: ANOVA & testy - Analýza rozptylu.
- f) **Mnohonásobné porovnávání:** návrat do okna Statistika dle skupin - záložka: Post-hoc - Tukeyův HSD.

9. cvičení

Příklad 1

U sedmi náhodně vybraných strojů v určitém podniku se zjišťovalo stáří stroje v letech (proměnná X) a týdenní náklady na údržbu stroje v Kč (proměnná Y). Data načtete z datového souboru *stroje.sta* a znázorníte je graficky. Na základě grafu posuďte, zda závislost Y na X může být modelována modely I.-IV., případně navrhněte jiný model.

$$\text{I.: } Y = \beta_0 + \beta_1 x + \varepsilon$$

$$\text{II.: } Y = \beta_0 + \beta_1 \sqrt{x} + \varepsilon$$

$$\text{III.: } Y = \beta_0 + \beta_1 \log_{10} x + \varepsilon$$

$$\text{IV.: } Y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$$

Dále předpokládáme, že pro vektor náhodných chyb platí $\varepsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$

Příklad 2

Úkoly a)- h) proveďte nejdříve pro model III., kde závislost počtu oprav na stáří stroje vystihuje funkce $y = \beta_0 + \beta_1 \log_{10} x$.

- Pro uvažovaný model sestavte regresní matici. Nalezněte vztah (v přednášce), z něhož lze určit odhady b_0, b_1 regresních parametrů β_0, β_1 a tyto odhady určete již pomocí sw. Napište rovnici odhadnuté regresní funkce.
- Najděte odhad s^2 rozptylu σ^2 , dále vypočtete index determinace ID^2 (někdy zvaný koeficient determinace a značený R^2) a interpretujte ho. (Nejdříve nalezněte odpovídající vzorce v přednášce, interpretujte jejich význam a vyčíslení proveďte pomocí sw.)
- Určete 95 % intervaly spolehlivosti pro regresní parametry β_0 a β_1 . (Vyčíslení pomocí sw, vzorce najít v přednášce.)
- Na hladině významnosti 0,05 proveďte celkový F-test. (Vyčíslení pomocí sw, vzorce najít v přednášce.)
- Na hladině významnosti 0,05 proveďte dílčí t-testy. (Vyčíslení pomocí sw, vzorce najít v přednášce.)
- Určete regresní odhad týdenních nákladů na údržbu stroje starého 4 roky.
- Znázorněte data s proloženou regresní funkcí.
- Posuďte normalitu reziduí. (Na reziduích by pochopitelně bylo potřeba posoudit i další předpoklady, kladené na náhodné chyby: nezávislost, homoskedasticitu a nulovou střední hodnotu náhodných chyb.)

Příklad 3

Pro modely I.-IV. určete odpovídající indexy determinace. Který z modelů je s ohledem na ID^2 nejvhodnější? (Pozor, posuzovat adekvátnost modelu pouze s ohledem na ID^2 je nedostatečné!)

doma k procvičení:

Příklad 4

Pro modely I., II., IV. proveďte úkoly a) - h) z 2. příkladu.

SW. návody k 9. cvičení

Návod k řešení příkladu 1:

Grafy - Bodové grafy - Proměnné: X na X-ovou osu; Y na Y-ovou osu -OK - Typ proložení vypnout - OK

Návod k řešení příkladu 2:

- a) Přidejte proměnnou s názvem LOGX, kde v dlouhém jménu bude: $=\log_{10}(v1)$; Statistiky - Vícerozměrná regrese - Závisle proměnná Y, nezávisle proměnné LOGX - OK - OK. Otevře se okno "Výsledky-vícenásobná regrese" - záložka Základní výsledky - "Výpočet: Výsledky regrese". Regresní parametry jsou v 3. sloupci "b".
- b) Vrátime se do "Výsledky - vícenásobná regrese" - záložka Detailní výsledky - ANOVA. Odhad rozptylu najdeme na řádce Rezid., ve sloupci Průměr čtverců. Index determinace je uveden v záhlaví původní výstupní tabulky pod označením R2.
- c) Nejdříve z pracovního sešitu vybereme tabulku "Výsledky regrese se závislou proměnnou: Y (stroje.sta)" Pravoklik na tabulku - Extrahovat jako samostatné okno - Kopie. Se samostatnou tabulkou dále pracujeme jako s datovým souborem. Přidáme v ní 2 proměnné DM (dolní mez) a HM (horní mez). Do jejich dlouhého jména napíšeme $=v3-v4*VStudent(0,975;5)$ pro dolní mez, resp. $=v3+v4*VStudent(0,975;5)$ pro horní mez. (Ve čtvrtém sloupci "sm. chyba z b" jsou hodnoty směrodatných chyb s_{b_j} .)
- d) Testovou statistiku F-testu a odpovídající p-hodnotu najdeme v záhlaví původní výstupní tabulky "Výsledky regrese", nebo v tabulce ANOVA.
- e) Obě testové statistiky t-testů najdeme v sloupci t(5) v tabulce "Výsledky regrese".
- f) Nejdříve určíme hodnotu $\log(4) = 0,60206$. Vrátime se do "Výsledky - vícenásobná regrese" - záložka Rezidua/předpoklady/předpovědi - Předpovědi závisle proměnné LOGX: 0,60206 - OK. Ve výstupní tabulce je hledaná hodnota označena jako Předpověď.
- g) Návrat do Bodové grafy - záložka Details: Typ proložení - Logaritmické; záložka Možnosti 2: Základ logaritmu - 10 -OK.
- h) Vrátime se do "Výsledky - vícenásobná regrese" - záložka Rezidua/předpoklady/předpovědi - Reziduální analýza. Otevře se okno "Reziduální analýza" - záložka Základ - použít postupně obě tlačítka: Normální p-graf reziduí; Výpočet Rezidua & předpovědi. Užitečný je i histogram reziduí: záložka Rezidua - histogram reziduí.

Návod k řešení příkladu 3:

Datový soubor s proměnnými X, Y a LOGX doplníme o proměnné SQRTX a INVX. Hodnoty proměnné SQRTX resp. INVX získáme tak, že do Dlouhého jména napíšeme $=\text{sqrt}(v1)$ resp. $=1/v1$.

ad 2g)

Pokud proložená funkce není v základní nabídce, můžeme do grafu vložit vlastní funkci následujícím postupem: Sestrojíme bodový graf bez proložení - pravoklik na pozadí grafu - Typ: Vlastní funkce - do okna zapsat rovnici vlastní funkce - OK.

10. cvičení

Příklad 1

U 65 náhodně vybraných zaměstnanců jisté firmy byla zjišťována délka praxe v letech (veličina X) a výška prémie v Kč (veličina Y). Dvourozměrné rozložení četností je dáno kontingenční tabulkou:

x / y	1250	1750	2250	2750	3250	3750	4250
12,5	5	3	0	0	0	0	0
17,5	2	4	4	0	0	0	0
22,5	0	1	6	7	4	0	0
27,5	0	0	1	3	7	1	0
32,5	0	0	0	1	10	5	1

Vypočtete realizaci r_{12} výběrového koeficientu korelace R_{12} a interpretujte jeho hodnotu. Pro úsporu času máte uvedeny následující součty:

$$\sum_{j=1}^5 n_{.j} x_{[j]} = 1\,562,5; \quad \sum_{k=1}^7 n_{.k} y_{[k]} = 172\,750; \quad \sum_{j=1}^5 n_{.j} x_{[j]}^2 = 40\,456; \quad \sum_{k=1}^7 n_{.k} y_{[k]}^2 = 498\,562\,500;$$

$$\sum_{j=1}^5 \sum_{k=1}^7 n_{jk} x_{[j]} y_{[k]} = 4\,446\,875;$$

Příklad 2

V souboru *sňatky_rozvody.sta* jsou údaje o počtu sňatků a rozvodů (na tisíc obyvatel) v Praze a v Brně v letech 1999 - 2008. Vypočtete výběrový koeficient korelace mezi sňatky v Brně a Praze. Na hladině 5 % testujte hypotézu, že "sňatkovost" v Brně a v Praze jsou nezávislé náhodné veličiny. (Rozmyslete si předpoklady testu.) Řešte ručně i pomocí sw.

Příklad 3

Sestrojte 95-% interval spolehlivosti pro korelační koeficient mezi sňatky v Praze a v Brně. Řešte ručně i pomocí sw.

Příklad 4

Načtete soubor *žáci.sta*. Na pětiprocentní hladině významnosti testujte hypotézu, že korelační koeficient mezi verbální a performační inteligencí je pro chlapce i dívky stejný. Řešte ručně i pomocí sw. (Rozmyslete si interpretaci nulové hypotézy.)

Příklad 5

Pro náhodný výběr rozsahu 50 z dvourozměrného normálního rozložení s koeficientem korelace ρ byl vypočten výběrový koeficient korelace $r_{12} = 0,5$. Na asymptotické hladině významnosti 0,05 testujte hypotézu $H_0 : \rho = 0,6$ proti $H_1 : \rho \neq 0,6$. **doma k procvičení:**

Příklad 6

Zadání příkladu 2 modifikujte pro rozvody.

Příklad 7

Učebnice 19.A-19.E

SW. návody k 10. cvičení

Návod k řešení příkladu 2:

- a) **Oprávněnost předpokladu dvojrozměrné normality:** Grafy - Bodové grafy - Proměnné X: SBRNO; Y: SPRAHA - OK - záložka: Details - Proložení vypnuto; Elipsa Normální - OK. Po vytvoření grafu upravíme měřítka na obou osách. Poklepeme na pozadí grafu - v stromové struktuře vybereme Měřítka - Mód: ručně - minimum pro osu X je 4,5; maximum pro osu X je 7; minimum pro osu Y je 4,8; maximum pro osu Y je 6,5 - OK.
- b) **Test nezávislosti:** Statistiky - Základní statistiky/tabulky - Korelační matice - OK - 2 seznamy - 1. seznam proměnných: SBRNO; 2. seznam proměnných: SPRAHA - OK - záložka: Možnosti - Zobrazit detailní tabulku výsledků - Výpočet. (V sloupci $r(X, Y)$ je realizace výběrového koeficientu korelace, v sloupci t je realizace testové statistiky a v sloupci p je odpovídající p-hodnota.)

Pokud známe výběrový koeficient korelace a rozsah výběru, můžeme test nezávislosti veličin X, Y provést pomocí Pravděpodobnostního kalkulátoru.

Statistiky - Pravděpodobnostní kalkulátor - Korelace - zadáme n a r , zaškrtneme Výpočet p z r - Výpočet. Kromě p-hodnoty bude ve výstupu i Fisherovo Z .

Návod k řešení příkladu 3:

Statistiky - Analýza síly testu - Odhad intervalu - Jedna korelace, t-test - v okýnku "pozorované R" doplníme pozorovaný výběrový koef. korelace;...; Výpočetní algoritmus: Fisherovo Z (původní) - Vypočítat.

Návod k řešení příkladu 4:

Statistiky - Základní statistiky a tabulky - Testy rozdílů: r , %, průměry - OK - Rozdíl mezi dvěma korelačními koeficienty: do políčka $r1$ napíšeme hodnotu výběrového korelačního koeficientu mezi performační a verbální inteligencí chlapců, do políčka $N1$ napíšeme počet chlapců ve výběru, analogicky $r2$ a $N2$ pro dívky - Výpočet.

11. cvičení

Příklad 1

U 100 náhodně vybraných vysokoškolských učitelů bylo zjišťováno jejich pohlaví (veličina X) a jejich pedagogická hodnota (veličina Y). Na asymptotické hladině významnosti 0,05 testujte hypotézu o nezávislosti pedagogické hodnosti a pohlaví a vypočtete Cramérův koeficient, jsou-li k dispozici následující údaje:

X/Y	odp. asistent	docent	profesor
muž	32	15	8
žena	34	8	3

Simultánní četnosti znázorněte graficky. Příklad řešte ručně i pomocí sw. Všimněte si, jak jsou v souboru *ped_hodnost.sta* zadána data z kontingenční tabulky. Proč je kritický obor soustředěn na pravém konci rozložení?

Jak by dopadl test, kdybychom sloučili sloupce docent a profesor?

Příklad 2

200 respondentů, z nichž bylo 73 žen, hodnotilo úroveň jistého časopisu. 34 žen ji hodnotilo kladně, stejně jako 47 mužů. Ostatní respondenti se o úrovni časopisu vyjádřili záporně. Vytvořte kontingenční tabulku simultánních absolutních četností a znázorněte je graficky.

- Na asymptotické hladině významnosti 0,05 testujte pomocí χ^2 testu hypotézu, že hodnocení úrovně časopisu nezávisí na pohlaví respondenta. Vypočtete také Cramérův koeficient. Řešte ručně i pomocí sw.
- Pomocí Fisherova přesného testu testujte hypotézu z úkolu a). Řešte pomocí sw.
- Na asymptotické hladině významnosti 0,05 testujte hypotézu o nezávislosti proti alternativě, že časopis se více líbí ženám. (Tedy šance kladného hodnocení ku zápornému, je vyšší u žen, než u mužů.) Řešte ručně.

Příklad 3

Načtete soubor *žáci.sta*. Pro proměnné VZDEL_M a SIDLO sestavte kontingenční tabulku a simultánní četnosti znázorněte graficky. Na asymptotické hladině významnosti 0,05 testujte hypotézu, že vzdělání matky a velikost sídla jsou na sobě nezávislé. Řešte jen pomocí sw.

Příklad 4

12 různých softwarových firem nabízí programy pro vedení účetnictví. Programy byly posouzeny odbornou komisí a komisí složenou z profesionálních účetních. Výsledky v 1. a 2. komisi jsou: (6,4), (7,5), (1,2), (8,10), (4,6), (2,1), (9,7), (12,11), (10,8), (3,3), (5,12), (11,9). Vypočtete Spearmanův koeficient pořadové korelace, interpretujte jej a na hladině významnosti 0,05 testujte hypotézu o nezávislosti pořadí v obou komisích. Test proveďte na základě kritických hodnot, které jsou v tabulkách v Učebních materiálech v ISu. Hodnotu Spearmanova korelačního koeficientu určete ručně i pomocí sw.

Příklad 5

Načtete soubor *TV_IQ.sta*. Proměnná TV udává čas týdně (v hodinách) strávený před televizí, proměnná IQ udává (v bodech) inteligenci. Určete hodnotu Spearmanova korelačního koeficientu mezi proměnnými IQ a TV. Na základě Spearmanova korelačního koeficientu testujte hypotézu, že IQ a čas strávený před televizí spolu nesouvisí. Test proveďte pomocí sw na hladině významnosti 0,05. Testové kritérium, které používá sw, má tvar $T_0 = r_S \sqrt{\frac{n-2}{1-r_S^2}}$ a při platné nulové hypotéze má asymptoticky studentovo rozložení

$t(n - 2)$.

Nejdříve nakreslete bodový graf a všimněte si "odlehých hodnot". Jelikož data jsou poměrového typu, lze uvažovat i o "obyčejném" korelačním koeficientu, ale právě tyto odlehle hodnoty vedou k volbě testů odvozených od Spearmanova korelačního koeficientu.

doma k procvičení

Příklad 6

Načtěte soubor *žáci.sta*. Testujte závislost proměnných vzdělání otce a vzdělání matky. Proměnné považujte za nominálního typu. (Mohou být nahlíženy jako ordinální, ale v souboru velkého rozsahu se opakují pouze 3 hodnoty, proto test pomocí Spearmanova koeficientu není vhodný.)

Příklad 7

učebnice 18.A-18.F

SW. návody k 11. cvičení

Návod k řešení příkladu 1:

- a) **Kontingenční tabulka s marginálními četnostmi.** Statistiky - Základní statistiky/tabulky - OK - Kontingenční tabulky - Specif. Tabulky - List 1: POHLAVI, List 2: PED_HODNOST - OK; zapneme tlačítko vah (vedle Select cases) - zaškrtnout "Použít váhy z tabulky" - Proměnná vah: CETNOST - OK - OK. Otevře se okno: "Výsledky; kontingenční tabulky". Na záložce Základní výsledky - Výpočet: Souhrn tabulek.
- b) **Grafické znázornění simultánních četností.** V okně Výsledky: kontingenční tabulky - záložka Detailní výsledky - 3D histogramy. Graf lze natáčet. 2× poklepeme na pozadí - v stromové struktuře vybereme: Zorný bod - Automatická rotace. (Případně ručně nastavit horizontální a vertikální úhel.)
- c) **Ověření předpokladů χ^2 rozložení testovací statistiky K .** V okně Výsledky: kontingenční tabulky - záložka Možnosti - zaškrtneme Očekávané četnosti; zaškrtneme Zvýraznit četnosti > 5 - Výpočet.
- c) **Samotný test a výpočet Cramerova V .** V okně Výsledky: kontingenční tabulky - záložka Možnosti - zaškrtneme Pearsonův & M-V chí kvadrát; zaškrtneme F_i (tabulky 2×2) & Cramérovo V & C - záložka Detailní výsledky - Detailní 2 rozm. tabulky.
- c) **Sloučení druhé a třetí varianty proměnné PED_HODNOST** Vytvoříme sloupec pro novou proměnnou PED_H_2 a vyznačíme ji. Roletka Data - Překódovat - Zahrnout pokud: PED_HODNOST=1, nová hodnota: 1; Zahrnout pokud: PED_HODNOST=2, nová hodnota: 2; Zahrnout pokud: PED_HODNOST=3, nová hodnota: 2 - OK. Dále dle předchozích bodů.

Návod k řešení příkladu 2:

ad b) **Fisherův přesný test.**V okně Výsledky: kontingenční tabulky - záložka Možnosti - zaškrtneme Fisher exakt., Yates, McNemar- záložka Detailní výsledky - Detailní 2 rozm. tabulky. Ve výstupu je pro oboustrannou alternativu p -hodnota v řádku (Fisherův přesný) 2-stranný.

Návod k řešení příkladu 4:

Statistiky - Neparametrické statistiky - Korelace (Spearman...) - OK - v okýnku Vytvořit: vybereme detailní report - Proměnné - 1.seznam proměnných: KOMISE1; 2.seznam proměnných: KOMISE2 - OK - Spearmanův koef. R.

Návod k řešení příkladu 5:

Postup viz. 4. příklad. (Není potřeba přiřazovat proměnným IQ a TV jejich pořadí, sw. to udělá sám.)

12. cvičení

Příklad 1

Z populace jisté základní školy byla vybraná 25-členná skupina žáků a u každého z nich bylo zjištěno, kolik hodin týdně strávili s internetem. Zjištěná data jsou v souboru *Internet.sta*.

- a) Zajímáme se o "polohu" rozložení (tedy střední hodnotu, či medián). Nakreslete histogram doplněný o test normality proměnné HODINY. Bude vhodný test o parametru μ , nebo o mediánu?
- b) Pomocí znaménkového testu testujte na hladině $\alpha = 0,05$ hypotézu, že medián proměnné HODINY = 10, proti alternativě, že medián proměnné HODINY > 10.
 - Pomocí se zjistěte hodnotu výběrového mediánu.
 - Zjistěte, kolik dětí strávilo více, než 10 hodin s internetem (případy vyznačte symbolem +), kolik méně, než 10 hodin (případy vyznačte symbolem -) a kolik přesně 10 hodin (případy vyznačte symbolem 0). Jak se změní rozsah souboru n ?
 - Pomocí distribuční funkce binomického rozložení určete p -hodnotu znaménkového testu a interpretujte závěr testu.
- c) Pomocí znaménkového testu testujte hypotézu, že medián proměnné HODINY = 10, proti alternativě, že medián proměnné HODINY \neq 10.
 - Pomocí distribuční funkce binomického rozložení určete p -hodnotu znaménkového testu a interpretujte závěr testu.
 - p -hodnotu znaménkového testu nalezněte pomocí implementovaného sotwarového testu. (Implementovaný test udává pouze p -hodnotu pro oboustrannou alternativu, navíc je pouze asymptotická tedy méně přesná, než řešení přes binomické rozložení.)

Příklad 2

U náhodného výběru 15-ti křižovatek byl zaznamenán počet nehod za rok před instalací semaforů a po jejich instalaci. Data jsou v souboru *Krizovatky.sta*.

- a) Chceme zjistit, zda se počet nehod snížil. Který test je lepší: párový t-test, nebo párový znaménkový test?
- b) Pomocí znaménkového testu testujte na hladině $\alpha = 0,05$ hypotézu, že počet nehod se nezměnil, proti alternativě, že počet nehod klesl. Test rozhodněte na základě p -hodnoty určené pomocí distribuční funkce binomického rozložení.

Příklad 3

Jistá studie porovnávala skupinu vegetariánů (X) se skupinou konzumentů masa (Y). Jednou ze sledovaných veličin byla délka života. Pomocí Wilcoxonova testu na hladině $\alpha = 0,05$ zjistěte, jestli jsou délky života v obou skupinách stejné.

- a) Nejdříve nakreslete histogramy pro obě skupiny a doplňte je o testy normality. Dále posuďte, zda se tvary histogramů podstatně liší i v tvaru, nebo jen v posunutí.
- b) Určete průměrnou dobu života v obou skupinách, průměrné pořadí v obou skupinách a medián délky života v obou skupinách. Která skupina s ohledem na výběrové charakteristiky má delší život?
- c) Test o shodě mediánů proveďte proti oboustranné alternativě pouze pomocí sw.
- d) Jaká by byla p -hodnota pro alternativu $H_1 : x_{0,5} > y_{0,5}$

doma k procvičení

Příklad 4

učebnice 16.A, 16.C

SW. návody k 12. cvičení

Návod k řešení příkladu 1:

- b2) Nejdříve přidáme sloupec s novou proměnnou ZNAMENKO. Roletka Data - Překódovat - v otevřeném okně vyplníme: Zahrnout pokud HODINY > 10, nová hodnota "+"; HODINY < 10, nová hodnota "-"; HODINY = 10, nová hodnota "0"; OK
Zjistíme četnosti znaků +,-,0. Roletka Statistika - Základní statistika a tabulky - Tabulky četností - OK - Proměnné: ZNAMENKO - Výpočet.
- b3) Otevřeme nový datový soubor o jedné proměnné a jednom případě. Do dlouhého jména napíšeme $= 1 - IBinom(12; 0, 5; 23)$ (Proč 12 a ne 13? Rozmyslete si, co by bylo v dlouhém jménu pro ostatní alternativy.)
- c1) Otevřeme nový datový soubor o jedné proměnné a jednom případě. Do dlouhého jména napíšeme $= 2 * (1 - IBinom(12; 0, 5; 23))$
(Kdy by bylo správné napsat $= 2 * (IBinom(?, 0, 5; 23))$?)
- c2) Nejdříve přidáme novou proměnnou c do které vložíme samé desítky. Dále Statistika - Neparametrická statistika - Porovnání dvou závislých vzorků - OK -1. seznam proměnných: HODINY, 2. seznam proměnných: 10 - OK - Znaménkový test.

Návod k řešení příkladu 2:

Vytvoříme novou proměnnou do které uložíme rozdíly proměnných PRED a PO. Do dlouhého jména napíšeme $= PRED - PO$. Dále postup dle příkladu 1.

Návod k řešení příkladu 3:

- b) Pořadové statistiky T_1 a T_2 budou ve výstupu popsáném v c). Jsou uvedeny pod názvy *Sčt poř skup. 1* a *Sčt poř skup. 2*
- c) Statistika - Neparametrická statistika - Porovnání dvou nezávislých vzorků - OK. Proměnné - Seznam závislých proměnných: DELKA_ZIV; Nezáv (grupovací) proměnná: KATEGORIE - OK. Tlačítko Mann-Whitneyův U test.