**Řešení domácích úkolů z Biostatistiky**

05\_samostatny ukol

Datovy soubor **05\_studenti.sta**

U proměnné Známka zjistěte absolutní, relativní četnost, dále absolutní a relativní kumulativní četnost.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Frequency table: Známka z biostatistiky (05\_studenti.sta) | | | |
| |  | | --- | | Count | | |  | | --- | | Cumulative Count | | |  | | --- | | Percent | | |  | | --- | | Cumulative Percent | |
| |  | | --- | | 1 | | 5 | 5 | 19,23077 | 19,2308 |
| |  | | --- | | 2 | | 11 | 16 | 42,30769 | 61,5385 |
| |  | | --- | | 3 | | 7 | 23 | 26,92308 | 88,4615 |
| |  | | --- | | 4 | | 3 | 26 | 11,53846 | 100,0000 |
| |  | | --- | | Missing | | 0 | 26 | 0,00000 | 100,0000 |

Zjistěte průměr, medián pro proměnnou Věk.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Descriptive Statistics (05\_studenti.sta) | | | | | |
| |  | | --- | | Valid N | | |  | | --- | | Mean | | |  | | --- | | Median | | |  | | --- | | Minimum | | |  | | --- | | Maximum | | |  | | --- | | Std.Dev. | |
| |  | | --- | | Věk | | 26 | 23,07692 | 22,50000 | 20,00000 | 33,00000 | 2,528606 |

U proměnné pohlaví zjistěte modus.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Descriptive Statistics (05\_studenti.sta) | | |
| |  | | --- | | Valid N | | |  | | --- | | Mode | | |  | | --- | | Frequency of Mode | |
| |  | | --- | | Pohlavi | | 26 | 1,000000 | 16 |

Kategorie 1 byly muži, tj. modus je „muži“, bylo jich 16.

Pro proměnnou známka zjistěte medián, modus

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Descriptive Statistics (05\_studenti.sta) | | | |
| |  | | --- | | Valid N | | |  | | --- | | Median | | |  | | --- | | Mode | | |  | | --- | | Frequency of Mode | |
| |  | | --- | | Známka z biostatistiky | | 26 | 2,000000 | 2,000000 | 11 |

Nejčastější známkou byla 2, bylo jich 11.

08\_domaci ukol

Datovy soubor **08\_01\_priklad.sta**

Na hladině významnosti testujte hypotézu, že střední hodnota výšky lidí je 175 cm proti oboustranné alternativě. Před provedením testu ověřte normalitu dat pomocí S-W testu.

Shapiro-Wilkův test (testování normality rozložení výšky): P = 0.464; tj. rozložení je normální

Jednovýběrový t-test: srovnání s referenční hodnotou 175 cm:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Test of means against reference constant (value) (08\_01\_priklad.sta) | | | | | | | |
| |  | | --- | | Mean | | |  | | --- | | Std.Dv. | | |  | | --- | | N | | |  | | --- | | Std.Err. | | |  | | --- | | Reference Constant | | |  | | --- | | t-value | | |  | | --- | | df | | |  | | --- | | p | |
| |  | | --- | | výška | | 175,6374 | 11,08757 | 21 | 2,419507 | 175,0000 | 0,263461 | 20 | 0,794891 |

Závěr: Výška lidí se neliší od 175 cm (jednovýběrový t-test: P=0.795).

Na hladině významnosti testujte hypotézu, že střední hodnota výšky lidí je 181 cm proti oboustranné alternativě. Jednovýběrový t-test: srovnání s referenční hodnotou

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Test of means against reference constant (value) (08\_01\_priklad.sta) | | | | | | | |
| |  | | --- | | Mean | | |  | | --- | | Std.Dv. | | |  | | --- | | N | | |  | | --- | | Std.Err. | | |  | | --- | | Reference Constant | | |  | | --- | | t-value | | |  | | --- | | df | | |  | | --- | | p | |
| |  | | --- | | výška | | 175,6374 | 11,08757 | 21 | 2,419507 | 181,0000 | -2,21638 | 20 | 0,038422 |

Závěr: Výška lidí se liší od 181 cm (jednovýběrový t-test: P = 0.038).

Datovy soubor **08\_02\_priklad.sta**

Ověřte normalitu výšky v 1. skupině a ve 2. skupině pomocí testů.

Výška v 1. skupině má normální rozložení (Shapiro-Wilkův test: P = 0.077).

Výška v 2. skupině má normální rozložení (Shapiro-Wilkův test: P = 0.343).

Na hladině významnosti 0,05 testujte hypotézu, že rozptyly výšek skupiny 1 a 2 jsou shodné.

Rozptyly výšek jsou shodné (F-test: P = 0.905).

Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty výšek skupiny 1 a 2 jsou shodné.

Výška v 1. a 2 skupině se statisticky významně liší (dvouvýběrový nepárový t-test: P < 0.001)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | T-tests; Grouping: skupina (08\_02\_priklad.sta) Group 1: 1 Group 2: 2 | | | | | | | | | | |
| |  | | --- | | Mean 1 | | |  | | --- | | Mean 2 | | |  | | --- | | t-value | | |  | | --- | | df | | |  | | --- | | p | | |  | | --- | | Valid N 1 | | |  | | --- | | Valid N 2 | | |  | | --- | | Std.Dev. 1 | | |  | | --- | | Std.Dev. 2 | | |  | | --- | | F-ratio Variances | | |  | | --- | | p Variances | |
| |  | | --- | | výška | | 180,2506 | 159,7945 | 44,71177 | 1998 | 0,00 | 1000 | 1000 | 10,21082 | 10,24963 | 1,007616 | 0,904578 |

Datovy soubor **08\_03\_priklad.sta**

Na hladině významnosti 0,05 testujte hypotézu, že dieta neměla významný vliv na změnu hmotnosti, tj. že rozdíl středních hodnot hmotnosti se neliší.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | T-test for Dependent Samples (08\_03\_priklad.sta) Marked differences are significant at p < ,05000 | | | | | | | | | |
| |  | | --- | | Mean | | |  | | --- | | Std.Dv. | | |  | | --- | | N | | |  | | --- | | Diff. | | |  | | --- | | Std.Dv. Diff. | | |  | | --- | | t | | |  | | --- | | df | | |  | | --- | | p | | |  | | --- | | Confidence -95,000% | | |  | | --- | | Confidence +95,000% | |
| |  | | --- | | pred\_dietou | | 86,80000 | 6,610598 |  |  |  |  |  |  |  |  |
| |  | | --- | | po\_diete | | 82,20000 | 5,263079 | 5 | 4,600000 | 2,190890 | 4,694855 | 4 | 0,009344 | 1,879650 | 7,320350 |

Závěr: U žen došlo působením diety k statisticky významné změně hmotnosti (párový t-test pro závislé výběry: P = 0.009).

Datovy soubor **08\_04\_priklad.sta**

Na hladině významnosti 0,05 testujte hypotézu, že lék neměl významný vliv na změnu krevního tlaku.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | T-test for Dependent Samples (08\_04\_priklad.sta) Marked differences are significant at p < ,05000 | | | | | | | | | |
| |  | | --- | | Mean | | |  | | --- | | Std.Dv. | | |  | | --- | | N | | |  | | --- | | Diff. | | |  | | --- | | Std.Dv. Diff. | | |  | | --- | | t | | |  | | --- | | df | | |  | | --- | | p | | |  | | --- | | Confidence -95,000% | | |  | | --- | | Confidence +95,000% | |
| |  | | --- | | placebo | | 190,3636 | 17,41421 |  |  |  |  |  |  |  |  |
| |  | | --- | | hydrochlothiazid | | 166,3636 | 21,42089 | 11 | 24,00000 | 13,09198 | 6,079980 | 10 | 0,000119 | 15,20469 | 32,79531 |

Závěr: Lék měl statistický významný vliv na změnu krevního tlaku (párový t-test pro závislé výběry: P < 0.001).

Datovy soubor **08\_05\_priklad.sta**

Na hladině významnosti 0,05 otestujte, zda se skutečná spotřeba automobilu odlišuje od toho, co udává výrobce. Výrobce udává, že průměrná spotřeba paliva je 12,5 l/100 km.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Test of means against reference constant (value) (08\_05\_priklad.sta) | | | | | | | |
| |  | | --- | | Mean | | |  | | --- | | Std.Dv. | | |  | | --- | | N | | |  | | --- | | Std.Err. | | |  | | --- | | Reference Constant | | |  | | --- | | t-value | | |  | | --- | | df | | |  | | --- | | p | |
| |  | | --- | | spotřeba | | 13,91429 | 0,555888 | 14 | 0,148567 | 12,50000 | 9,519502 | 13 | 0,000000 |

Závěr: Spotřeba automobilu se liší od výrobcem udávané hodnoty spotřeby 12,5 l/100 km (jednovýběrový t-test – srovnání s referenční hodnotou 12,5: P < 0.001).

09\_Domaci ukol

Datovy soubor **09\_01\_priklad.sta**

Pomocí vhodného testu testujte hypotézu, že spotřeba benzínu A i B byla stejná (hladina významnosti = 0,05).

Ke srovnání použijeme párový neparametrický test (Wilcoxonův párový test (pro N < 30) nebo znaménový test (pro menší N)). Vhodnější bude znaménkový test kvůli nízkému počtu záznamů.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pair of Variables | Sign Test (09\_01\_priklad.sta) Marked tests are significant at p <,05000 | | | |
| |  | | --- | | No. of Non-ties | | |  | | --- | | Percent v < V | | |  | | --- | | Z | | |  | | --- | | p-value | |
| |  | | --- | | benzín A & benzín B | | 10 | 50,00000 | -0,316228 | 0,751830 |

Závěr: Spotřeba benzínu A i B je stejná (Znaménkový test: P = 0.752).

Datovy soubor **09\_02\_priklad.sta**

Pomocí vhodného testu testujte hypotézu, že porovnávané způsoby výkrmů (1-klasická směs, 2- směs s vitamínovým doplňkem) se neliší (hladina významnosti = 0,05).

Použijeme nepárový neparametrický test, tj. Mann-Whitneyův U test.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | Mann-Whitney U Test (w/ continuity correction) (09\_02\_priklad.sta) By variable skupina Marked tests are significant at p <,05000 | | | | | | | | | |
| |  | | --- | | Rank Sum standardní směs | | |  | | --- | | Rank Sum směs i vitamín | | |  | | --- | | U | | |  | | --- | | Z | | |  | | --- | | p-value | | |  | | --- | | Z adjusted | | |  | | --- | | p-value | | |  | | --- | | Valid N standardní směs | | |  | | --- | | Valid N směs i vitamín | | |  | | --- | | 2\*1sided exact p | |
| |  | | --- | | vaha | | 282,0000 | 459,0000 | 92,00000 | -2,56914 | 0,010196 | -2,57069 | 0,010150 | 19 | 19 | 0,009047 |

Závěr: Váhové přírůstky u selat krmených klasickou směsí a směsí s vitamínovým doplňkem se statisticky významně liší (Mann-Whitney U test: P = 0.010).

10\_Domaci\_ukol

Datovy soubor **10\_01\_priklad.sta**

Testujte hypotézu, že barva vlasů a barva očí spolu nesouvisí. Vypočítejte Cramérův koeficient a interpretujte jej.

Použijeme Pearsonův chi-kvadrát test; podmínky dobré aproximace byly splněny.

|  |  |  |  |
| --- | --- | --- | --- |
| Statistic | Statistics: Vlasy(4) x Oči(3) (10\_01\_priklad.sta) | | |
| |  | | --- | | Chi-square | | |  | | --- | | df | | |  | | --- | | p | |
| |  | | --- | | Pearson Chi-square | | 1073,508 | df=6 | p=0,0000 |
| |  | | --- | | M-L Chi-square | | 1137,606 | df=6 | p=0,0000 |
| |  | | --- | | Phi | | ,3973270 |  |  |
| |  | | --- | | Contingency coefficient | | ,3692482 |  |  |
| |  | | --- | | Cramér's V | | ,2809526 |  |  |

Závěr: barva vlasů a barva očí spolu souvisí (Pearsonův chi-kvadrát test: P < 0.001). Mezi barvou očí a barvou vlasů je slabá závislost (Cramerův koeficient: V = 0.281).

Datovy soubor **10\_02\_priklad.sta**

Ve Skotsku byla provedena studie, která měla prokázat, zda procentuální zastoupení krevních skupin na celém území je homogenní nebo není.

Použijeme pearsonův chi-kvadrát test; podmínky dobré aproximace byly splněny.

|  |  |  |  |
| --- | --- | --- | --- |
| Statistic | Statistics: Krevni skupina(4) x Oblast(3) (10\_02\_priklad.sta) | | |
| |  | | --- | | Chi-square | | |  | | --- | | df | | |  | | --- | | p | |
| |  | | --- | | Pearson Chi-square | | 10,45372 | df=6 | p=,10680 |
| |  | | --- | | M-L Chi-square | | 10,85063 | df=6 | p=,09310 |
| |  | | --- | | Phi | | ,1478841 |  |  |
| |  | | --- | | Contingency coefficient | | ,1462931 |  |  |
| |  | | --- | | Cramér's V | | ,1045699 |  |  |

Závěr: Procentuální zastoupení krevních skupin je na třech lokalitách homogenní (Pearsonův chi-kradrát test: P = 0.107).