**NEPARAMETRICKÉ TESTY**

**Neparametrický jednovýběrový**Jeden výběr jehož medián srovnáváme s nějakou hodnotou – Wilcoxonův jednovýběrový test

**1) Máme data z družice Hipparcos pro deklinaci (obdoba zeměpisné šířky) pro pozici 2717 hvězd. Chceme testovat na hladině statistické významnosti 0,05, střední hodnota této deklinace je -47°. (dataset**  hip DE.sta**)**

**H0**: Medián deklinace je roven -47°. (=-47)
**H1**: Medián deklinace je staticky významně odlišný -47°. (≠ -47)

**Krok A) Testování normality našeho výběru –** p menší jak 0,05 -> Neparametrický test (Jinak by se dal použít parametrický)

**Krok B) STATISTICA nám neumožňuje uskutečnit jednovýběrový Wilcoxonův test přímo. Ale pokud si uvědomíme, že se jedná vlastně o modifikaci párového testu, kde mediánem druhého výběru je právě číslo, s kterým srovnáváme medián našeho výběru, pak jsme schopni tento příklad vypočítat. Z H0 se tak stává (**-=0, kde =-47**). Druhý výběr tak vytvoříme jednoduše tak, že vypočteme novou proměnnou jejíž všechny prvky budou rovny – 47.
**

**Krok C) Výpočet testové statistiky** *Statistiky -> Neparametrická statistika -> Porovnání dvou závislých vzorků -> OK
*

Jako proměnné vybereme naše dvě proměnné (*Deklinace* a *referenční hodnota*) a dáně *Wilcoxonův párový test…

*Z výsledků vidíme, že p-hodnota je menší než 0,05 – nulovou hypotézu o rovnosti mediánu
našeho výběru k hodnotě -47 tedy **zamítáme.**

**2) V různých částech České republiky bylo zjištěno procentuální zastoupení kuřáků nad 60 let. Zjistěte, zda je celkem v těchto regionech České republiky zastoupení kuřáků nad 60 let rovno 12 procentům. (**populace\_nad\_60.sta**)

Pokud se podíváte na normalitu těchto dat – pak zjistíte, že p-hodnota S-W testu je 0,07, což sice nezamítá nulovou hypotézu o normalitě, ale z důvodu malého vzorku a téměř dosažené hranice pro zamítnutí hypotézy by bylo vhodnější použít NEPARAMETRICKÝ test.**

[p-hodnota 0,0499, Zamítáme nulovou hypotézu rovnosti mediánu našeho výběru a předpokládané hodnoty 12%]

**3 ) Byl sledován zisk (v 1000 kč) jedné společnosti v 44 týdnech. Zjistěte zda byla střední hodnota zisku za tuto dobu rovna 175 000. (**zisk.sta**)**

Opět normalita vychází, ale těsně nad hranici zamítnutí. (Opět lepší použít NEPARAMETRICKÝ test)

[p-hodnota menší jak 0,001, Zamítáme nulovou hypotézu rovnosti střední hodnoty zisku našeho výběru a předpokládané hodnoty 175]

**NEPARAMETRICKÝ NEPÁROVÝ TEST**

Srovnáváme hodnoty mediánů dvou na sebe nezávislých výběrů. Alespoň jeden výběr nemá splněný předpoklad normality dat.

**1) Testujte na hladině statistické významnosti 0,05 rovnost šířky okvětních lístků u kosatců Setosy a Versicol. (**kosatce.sta**)**

**H0:** Šířky okvětních lístků u Setosy a Versicolor jsou stejné. (=)
**H1:** Šířky okvětních lístků u Setosy a Versicolor jsou odlišné. (≠)

**Krok A )** Data máme ve tvaru, kdy naměřené hodnoty šířky okvětních lístků máme rozděleny dle skupinové proměnné *IRISTYPE.* To znamená, že při testování normality jednotlivých výběrů musíme zapnout Analyzování dle skupin. (*viz parametrické testy-dvouvýběrový nepárový t-test*) –

Oba výběry nemají normální rozdělení – (normalita počítána *Statistiky->základní statistiky->tabulky četností->*záložka *Normalita*(Lilliefors a Shapiro-Wilk))



**Krok B)** Výpočet statistiky – *Statistiky->Neparametrické statistiky -> Porovnání dvou nezávislých vzorků (skupiny)->OK
*

V případě, že by vstupní data byla ve formátu, že hodnoty okvětních lístků pro Setosu a Versicolor byly jednotlivé proměnné, pak si je musíme upravit do podoby jakou máme zde… tedy vytvořit skupinovou proměnnou

Vybereme proměnné…


A vybereme Mann-Whitneyův U test


Výsledná tabulka:


Nyní se na základě velikosti vzorku (v tomto případě n=100) rozhodneme jakou p-hodnotu použijeme. Pokud je n>30 pak používáme asymptotickou p-hodnotu, pokud je n menší jak 30 pak používáme přesnou hodnotu.

Z výsledků je patrné, že zamítáme nulovou hypotézu.

**2)** Máme dvě odrůdy brambor a sledujeme jejich výnosnost na různých místech. Testujte na hladině významnosti 0,05, že výnosnost obou odrůd je stejná.(odrudy.sta)

- potřeba si vytvořit jednu skupinovou proměnnou a všechny výnosnosti dát pod sebe – vzniknou tak dvě proměnné o 18 případech…

[p – 0,161, nezamítáme nulovou hypotézu o stejné výnosnosti jednotlivých odrůd brambor.]

**3)**  Máme dva typy kreditních karet Visa a M/E kartu. Na hladině významnosti 0,05 testujte, zda se v různých městech provede stejný počet plateb pomocí těchto dvou typů karet (počty v 1000) (kreditni\_karty.sta).

Pozn.

* Pokud provedeme test normality tak vyjdou oba výběry normální, nicméně u jednoho vyjde p-hodnota blízká 0,05 (0,07) a z důvodu menšího n, bych doporučil dělat neparametrický test.

[p=0,252, Nezamítáme nulovou hypotézu o stejném používání M/E karty a Visa karty]

**Neparametrický párový test**

Testujeme, že se závislé proměnné neliší před a po nějaké události, popřípadě jiným způsobem měření.

Párovost lze zjistit pomocí korelací, viz. Parametrické párové testy.

**1) Máme dvě měření křehkosti ocele (před zpracováním a po něm). Zjistěte, zda se statisticky významně liší. (**ocel.sta**)

Krok A)** vytvoříme proměnnou diference obou měření…Zjistíme normalitu rozdílu před a po zpracování dat – viz. Parametrické párové testy. (p=0,03 – zamítáme normalitu diference dat)
****

**Krok B)** Testování – *Statistiky->Neparametrické statistiky ->OK
*

Vybereme proměnné a zvolíme *Wilcoxonův párový test* (větší síla – spíše na symetrická data..dá se posoudit dle tvaru histogramu) nebo *Znaménkový test* (*menší síla testu*)…v podstatě je na Vás jaký použijete – zde použijeme *Wilcoxonův test*

Výsledná tabulka:

Dle p-hodnoty vidíme, že zamítáme nulovou hypotézu o stejné křehkosti ocele před zpracováním a po něm.

**2)** Máme k dispozici nějakou hladinu jistého parametru v krvi (např. kreatininu)před operací a po operaci. Zjistěte zda se na hladině statistické významnosti 0,05 tyto dvě hladiny liší. (parametr\_krve.sta)

[p- 0,017, Zamítáme nulovou hypotézu o rovnosti obsahu tohoto parametru v krvi před a po operaci]

**3)** Naměřili jsme hodnoty diet u jednotlivých druhů krys. Testujte, zda se jednotlivé diety liší (efektivita diety pro krysy.sta)

-Normalita – p=0,09 (Zde se asi dají použít oba druhy testů, jak parametrický tak neparametrický – ze cvičných důvodů použijeme neparametrický)

[p=0,056, Nezamítáme nulovou hypotézu o rovnosti jednotlivých typů diety.]