1. Analýza rozptylu jednoduchého třídění (ANOVA)
1.1. Označení
1.2. Testování hypotézy o shodě středních hodnot
1.3. Testování hypotézy o shodě rozptylů
1.4. Metody mnohonásobného porovnávání
1.5. Doporučený postup při ANOVĚ
1.6. Příklad
1.7. Význam předpokladů v ANOVĚ

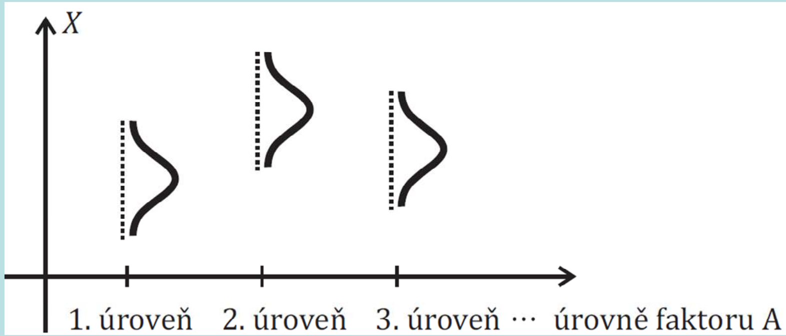
2. Neparametrické obdoby t-testů a ANOVY
2.1. Přehled parametrických a neparametrických testů
2.2. Pojem pořadí a průměrného pořadí
2.3. Jednovýběrový a párový Wilcoxonův test a jeho asymptotická varianta
2.4. Dvouvýběrový Wilcoxonův test a jeho asymptotická varianta
2.5. Kruskalův – Wallisův test
2.6. Mediánový test
2.7. Metody mnohonásobného porovnávání
2.8. Příklad na K-W test a mediánový test

1. Analýza rozptylu jednoduchého třídění (ANOVA)

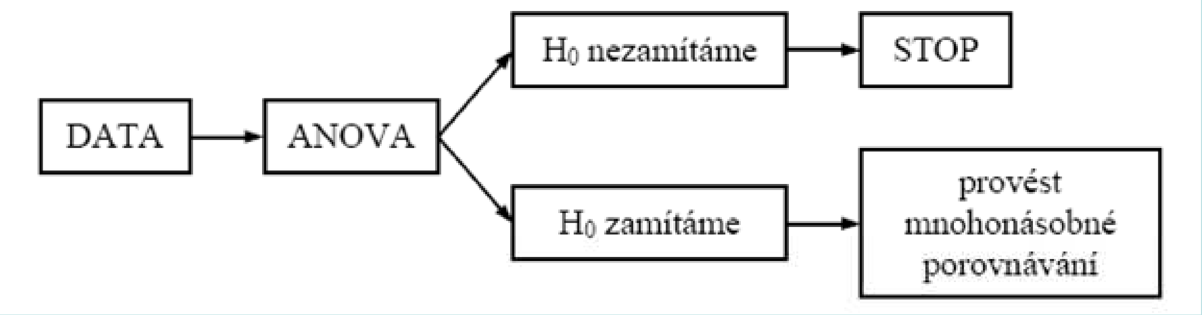
Motivace: Zajímáme se o problém, zda lze určitým faktorem (tj. nominální náhodnou veličinou A) vysvětlit variabilitu pozorovaných hodnot náhodné veličiny X, která je intervalového či poměrového typu. Např. zkoumáme, zda metoda výuky určitého předmětu (faktor A) ovlivňuje počet bodů dosažených studenty v závěrečném testu (náhodná veličina X).
Předpokládáme, že faktor A má $r \geq 3$ úrovní a přitom i-té úrovni odpovídá n_i pozorování $X_{i1},\dots,X_{in_i}$, které tvoří náhodný výběr z rozložení $N(\mu_i, \sigma^2)$, $i = 1, ..., r$ a jednotlivé náhodné výběry jsou stochasticky nezávislé, tedy $X_{ij} = \mu_i + \varepsilon_{ij}$, kde ε_{ij} jsou stochasticky nezávislé náhodné veličiny s rozložením $N(0, \sigma^2)$, $i = 1, \dots, r$, $j = 1, \dots, n_i$.
Výsledky lze zapsat do tabulky

faktor A	výsledky
úroveň 1	$X_{11},\dots,X_{1n_1}$
úroveň 2	$X_{21},\dots,X_{2n_2}$
…	…
úroveň r	$X_{r1},\dots,X_{rn_r}$

Ilustrace:



Na hladině významnosti α testujeme nulovou hypotézu, která tvrdí, že všechny střední hodnoty jsou stejné, tj.
H_0: $\mu_1 = \ldots = \mu_r$ proti alternativní hypotéze H_1, která tvrdí, že aspoň jedna dvojice středních hodnot se liší.
Jedná se tedy o zobecnění dvouvýběrového t-testu a na první pohled se zdá, že stačí utvořit $\binom{r}{2}$ dvojic náhodných výběrů a na každou dvojici aplikovat dvouvýběrový t-test. Hypotézu o shodě všech středních hodnot bychom pak zamítli, pokud aspoň v jednom případě z $\binom{r}{2}$ porovnávání se prokáže odlišnost středních hodnot. Odtud je vidět, že k neoprávněnému zamítnutí nulové hypotézy (tj. k chybě 1. druhu) může dojít s pravděpodobností větší než α. Proto ve 30. letech 20. století vytvořil R. A. Fisher metodu ANOVA (analýza rozptylu, v popsané situaci konkrétně analýza rozptylu jednoduchého třídění), která uvedenou podmínku splňuje.
Pokud na hladině významnosti α zamítneme nulovou hypotézu, zajímá nás, které dvojice středních hodnot se od sebe liší.
K řešení tohoto problému slouží metody mnohonásobného porovnávání, např. Scheffého nebo Tukeyova metoda.



1.1. Označení:

V analýze rozptylu jednoduchého třídění se používá tzv. tečková notace.

$$n = \sum_{i=1}^{r} n_i$$ … celkový rozsah všech r výběrů

$$X_{i.} = \sum_{j=1}^{n_i} X_{ij}$$ … součet hodnot v i-tém výběru

$$M_{i.} = \frac{1}{n_i} X_{i.}$$ … výběrový průměr v i-tém výběru

$$X_{..} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} X_{ij}$$ … součet hodnot všech výběrů

$$M_{..} = \frac{1}{n} X_{..}$$ … celkový průměr všech r výběrů

Zavedeme součty čtverců

$$S_T = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ij} - M_{..}\right)^2$$ … celkový součet čtverců

(charakterizuje variabilitu jednotlivých pozorování kolem celkového průměru),
počet stupňů volnosti $f_T = n - 1$,

$$S_A = \sum_{i=1}^{r} n_i \left(M_{i.} - M_{..}\right)^2$$ … skupinový součet čtverců

(charakterizuje variabilitu mezi jednotlivými náhodnými výběry),
počet stupňů volnosti $f_A = r - 1$.

$$S_E = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ij} - M_{i.}\right)^2$$ … reziduální součet čtverců

(charakterizuje variabilitu uvnitř jednotlivých výběrů),
počet stupňů volnosti $f_E = n - r$.

Lze dokázat, že $S_T = S_A + S_E$.

1.2.Testování hypotézy o shodě středních hodnot

Náhodné veličiny X_{ij} se řídí modelem

M0: $X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

pro $i = 1, \dots, r$, $j = 1, \dots, n_i$, přičemž

ε_{ij} jsou stochasticky nezávislé náhodné veličiny s rozložením $N(0, \sigma^2)$,

μ je společná část střední hodnoty závisle proměnné veličiny,

α_i je efekt faktoru A na úrovni i.

Parametry μ, α_i neznáme.

Požadujeme, aby platila tzv. reparametrizační rovnice: $\sum_{i=1}^{r} n_i \alpha_i = 0$.

(Pokud je třídění vyvážené, tj. pokud mají všechny výběry stejný rozsah: $n_1 = n_2 = \dots = n_r$, pak

lze použít zjednodušenou podmínku $\sum_{i=1}^{r} \alpha_i = 0$.)

Kdyby nezáleželo na faktoru A, platila by hypotéza $\alpha_1 = \ldots = \alpha_r = 0$ a dostali bychom model
M1: $X_{ij} = \mu + \varepsilon_{ij}$.
Během analýzy rozptylu tedy zkoumáme, zda výběrové průměry $M_1, \ldots, M_r$ se od sebe liší pouze v mezích náhodného kolísání kolem celkového průměru M nebo zda se projevuje vliv faktoru A.

Rozdíl mezi modely M0 a M1 ověřujeme pomocí testové statistiky
$F_A = \frac{S_A / f_A}{S_E / f_E}$, která se řídí rozložením F(r-1,n-r), je-li model M1 správný. Hypotézu o nevýznamnosti faktoru A tedy zamítneme na hladině významnosti α, když platí: $F_A \geq F_{1-\alpha}(r-1,n-r)$.

Výsledky výpočtů zapisujeme do tabulky analýzy rozptylu jednoduchého třídění.

Zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	S_A	$f_A = r - 1$	S_A/f_A	$\frac{S_A/f_A}{S_E/f_E}$
reziduální	S_E	$f_E = n - r$	S_E/f_E	-
celkový	S_T	$f_T = n - 1$	-	-

Sílu závislosti náhodné veličiny X na faktoru A můžeme měřit pomocí poměru determinace: $P^2 = \frac{S_A}{S_T}$. Nabývá hodnot z intervalu $\langle 0,1 \rangle$.

1.3. Testování hypotézy o shodě rozptylů

Před provedením analýzy rozptylu je zapotřebí ověřit předpoklad o shodě rozptylů v daných r výběrech.

a) Levenův test: Položme $Z_{ij} = \left| X_{ij} - M_{i.} \right|$. Označíme

$$M_{Zi} = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij},$$

$$M_Z = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Z_{ij},$$

$$S_{ZE} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Z_{ij} - M_{Zi}\right)^2,$$

$$S_{ZA} = \sum_{i=1}^{r} n_i\left(M_{Zi} - M_Z\right)^2$$

Platí-li hypotéza o shodě rozptylů, pak statistika

$$F_{ZA} = \frac{S_{ZA}/(r-1)}{S_{ZE}/(n-r)} \approx F(r\text{-}1,\ n\text{-}r).$$

Hypotézu o shodě rozptylů tedy zamítáme na asymptotické hladině významnosti α, když $F_{ZA} \geq F_{1-\alpha}(r\text{-}1,\ n\text{-}r)$.
(Levenův test je vlastně založen na analýze rozptylu absolutních hodnot centrovaných pozorování. Vzhledem k tomu, že náhodné veličiny $X_{ij} - M_i$ nejsou stochasticky nezávislé a absolutní hodnoty těchto veličin nemají normální rozložení, je Levenův test pouze aproximativní.)

b) Brownův – Forsytheův test je modifikací Levenova testu. Modifikace spočívá v tom, že místo výběrového průměru i-tého výběru se při výpočtu veličiny Z_{ij} používá medián i-tého výběru.

c) Bartlettův test: Platí-li hypotéza o shodě rozptylů a rozsahy všech výběrů jsou větší než 6, pak statistika $B = \frac{1}{C}\left[(n-r)\ln S_*^2 - \sum_{i=1}^{r}(n_i - 1)\ln S_i^2\right]$ se asymptoticky řídí rozložením $\chi^2(r-1)$. Přitom konstanta $C = 1 + \frac{1}{3(r-1)}\left(\sum_{i=1}^{r}\frac{1}{n_i - 1} - \frac{1}{n-r}\right)$ a S_*^2 je vážený průměr výběrových rozptylů.

H_0 zamítáme na asymptotické hladině významnosti α, když B se realizuje v kritickém oboru $W = \langle \chi^2_{1-\alpha}(r-1), \infty)$.

1.4. Post – hoc metody mnohonásobného porovnávání

Zamítneme-li na hladině významnosti α hypotézu o shodě středních hodnot, chceme zjistit, které dvojice středních hodnot se liší na dané hladině významnosti α, tj. na hladině významnosti α testujeme $H_0: \mu_l = \mu_k$ proti $H_1: \mu_l \neq \mu_k$ pro všechna l, k = 1, .., r, $l \neq k$.

a) Mají-li všechny výběry týž rozsah p (říkáme, že třídění je vyvážené), použijeme Tukeyovu metodu.

Testová statistika má tvar $\dfrac{|M_{k.} - M_{l.}|}{\frac{S_*}{\sqrt{p}}}$. Rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$\dfrac{|M_{k.} - M_{l.}|}{\frac{S_*}{\sqrt{p}}} \geq q_{1-\alpha}(r, n-r)$, kde hodnoty $q_{1-\alpha}(r, n\text{-}r)$ jsou kvantily studentizovaného rozpětí a najdeme je ve statistických tabulkách. (Studentizované rozpětí je náhodná veličina $Q = \dfrac{X_{(n)} - X_{(1)}}{s}$.)

Existuje modifikace Tukeyovy metody pro nestejné rozsahy výběrů, nazývá se Tukeyova HSD metoda. V tomto případě má testová statistika tvar $\dfrac{|M_{k.} - M_{l.}|}{S_* \sqrt{\frac{1}{2}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}}$. Rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$$\frac{|M_{k.} - M_{l.}|}{S_* \sqrt{\frac{1}{2}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}} \geq q_{1-\alpha}(r, n-r).$$

b) Nemají-li všechny výběry stejný rozsah, použijeme Scheffého metodu: rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$$\left|M_{k.} - M_{l.}\right| \geq S_* \sqrt{(r-1)\left(\frac{1}{n_k} + \frac{1}{n_l}\right) F_{1-\alpha}(r-1, n-r)}\,.$$

Výhodou Scheffého testu je, že k jeho provedení nepotřebujeme speciální statistické tabulky s hodnotami kvantilů studentizovaného rozpětí, ale stačí běžné statistické tabulky s kvantily Fisherova – Snedecorova rozložení.

V případě vyváženého třídění, kdy lze aplikovat Tukeyovu i Scheffého metodu, použijeme tu, která je citlivější. Tukeyova metoda tedy bude výhodnější, když
$q_{1-\alpha}^2(r, n-r) < 2(r-1)F_{1-\alpha}(r-1, n-r)$.

Metody mnohonásobného porovnávání mají obecně menší sílu než ANOVA.
Může nastat situace, kdy při zamítnutí H_0 nenajdeme metodami mnohonásobného porovnávání významný rozdíl u žádné dvojice středních hodnot. K tomu dochází zvláště tehdy, když p-hodnota pro ANOVU je jen o málo nižší než zvolená hladina významnosti. Pak slabší test patřící do skupiny metod mnohonásobného porovnávání nemusí odhalit žádný rozdíl.

1.5. Doporučený postup při provádění analýzy rozptylu:

a) Ověření normality daných r náhodných výběrů (grafické metody - NP plot, Q-Q plot, histogram, testy hypotéz o normálním rozložení - Lilieforsova varianta Kolmogorovova – Smirnovova testu nebo Shapirův – Wilkův test).
Doporučuje se kombinace obou způsobů. Závěry učiníme až na základě posouzení obou výsledků.
Obecně lze říci, že analýza rozptylu není příliš citlivá na porušení předpokladu normality, zvláště při větších rozsazích výběrů (nad 20), což je důsledek působení centrální limitní věty. Mírné porušení normality tedy není na závadu, při větším porušení použijeme např. Kruskalův – Wallisův test jako neparametrickou obdobu analýzy rozptylu jednoduchého třídění.

b) Po ověření normality se testuje homogenitu rozptylů, tj. předpoklad, že všechny náhodné výběry pocházejí z normálních rozložení s týmž rozpylem. Graficky ověřujeme shodu rozptylů pomocí krabicových diagramů, kdy sledujeme, zda je šířka krabic stejná. Numericky testujeme homogenitu rozptylů pomocí Levenova testu, Brownova – Forsytheova testu (oba jsou implementovány ve STATISTICE, Brownův – Forsytheův test v MINITABu) či Bartlettova testu (je k dispozici v MINITABu).
Slabé porušení homogenity rozptylů nevadí, při větším se doporučuje mediánový test.

c) Pokud jsou splněny předpoklady normality a homogenity rozptylů, můžeme přistoupit k testování shody středních hodnot. Předtím je samozřejmě vhodné vypočítat průměry a směrodatné odchylky či rozptyly v jednotlivých skupinách.

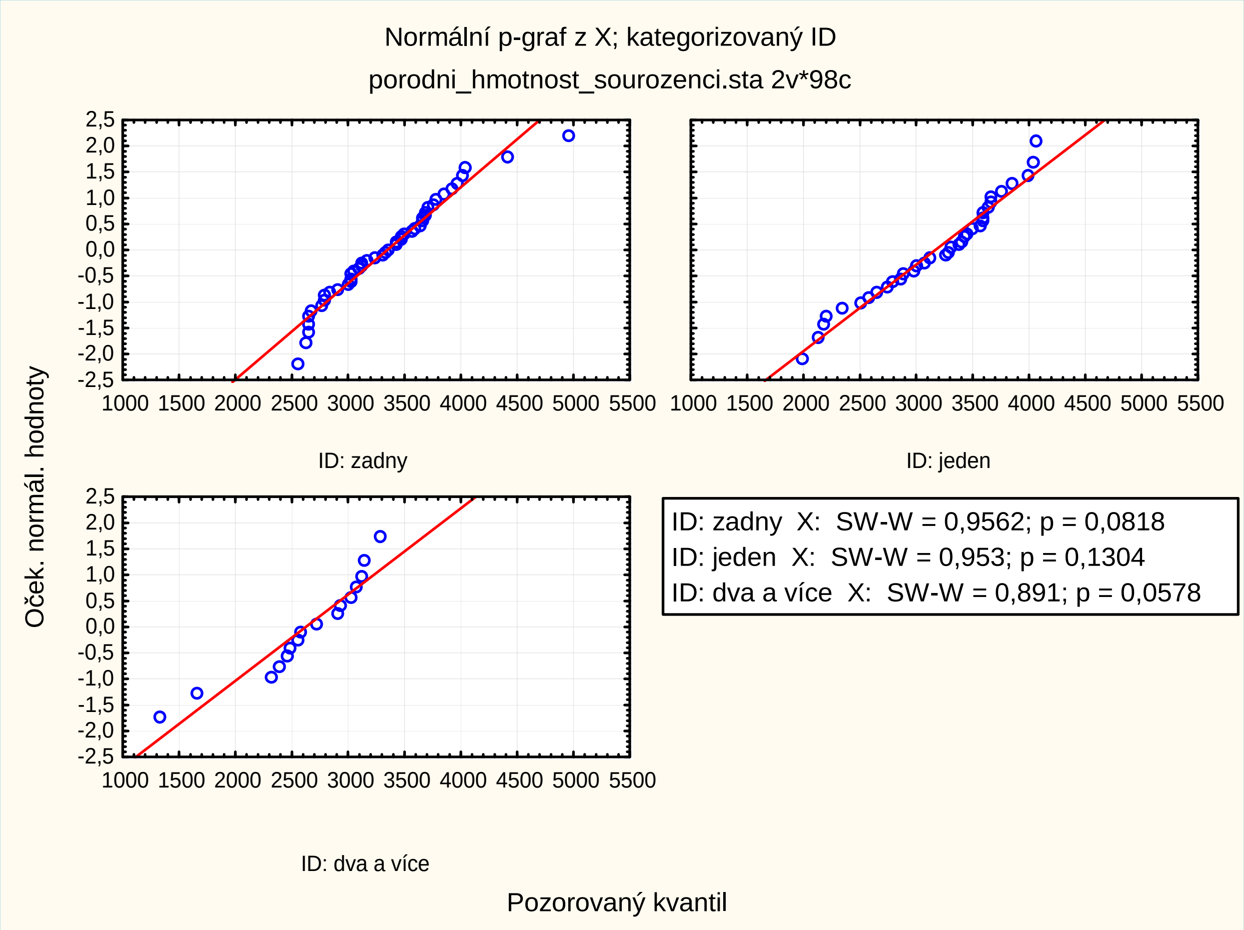
d) Dojde-li na zvolené hladině významnosti k zamítnutí hypotézy o shodě středních hodnot, zajímá nás, které dvojice středních hodnot se od sebe liší. K řešení tohoto problému slouží post-hoc metody mnohonásobného porovnávání, např. Scheffého nebo Tukeyova metoda.

1.6. Příklad: Máme k dispozici údaje o porodní hmotnosti 98 novorozenců. Kromě porodní hmotnosti je také uveden počet starších sourozenců. Má varianty 0, 1, 2 a víc.
Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota porodní hmotnosti nezávisí na počtu starších sourozenců. Zamítnete-li nulovou hypotézu, zjistěte, které dvojice variant počtu starších sourozenců se liší na hladině významnosti 0,05.

Řešení pomocí systému STATISTICA

Otevřeme datový soubor porodni_hmotnost_sourozenci.sta o dvou proměnných X a ID a 98 případech. V proměnné X jsou uloženy zjištěné hmotnosti, v proměnné ID kódy pro počty starších sourozenců (0 pro prvorozené dítě, 1 pro druhorozené dítě, 2 pro dítě se dvěma a více staršími sourozenci).

Ověříme normalitu daných tří náhodných výběrů pomocí N-P plotu a S-W testu:



Hypotézu o normalitě nezamítáme na hladině významnosti pro žádnou skupinu novorozenců.

Vypočteme výběrové průměry a výběrové směrodatné odchylky:
Statistiky – Základní statistiky a tabulky – Rozklad & jednofakt. ANOVA – OK – Proměnné – Závislé – X, Grupovací - ID – OK – Skupiny tabulek - Výpočet.

Rozkladová tabulka popisných statistik (porodni_hmotnost_sourozenci.sta)
N=98 (V seznamu záv. prom. nejsou ChD)

ID	X průměr	X N	X Sm.odch.
zadny	3346,087	46	515,9650
jeden	3170,000	36	573,4258
dva a více	2624,375	16	537,5744
Vš.skup.	3163,571	98	592,1205

Nyní ověříme předpoklad shody rozptylů.
Aktivujeme Statistiky dle skupin – vybereme záložku Jednotlivé tabulky – OK – vybereme záložku ANOVA & testy - Levenův test – Výpočet.

Leveneův test homogenity rozpylů (porodni_hmotnost_sourozenci.sta)
Označ. efekty jsou význ. na hlad. p < ,05000

Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
X	107852,4	2	53926,22	8996979	95	94705,04	0,569412	0,567777

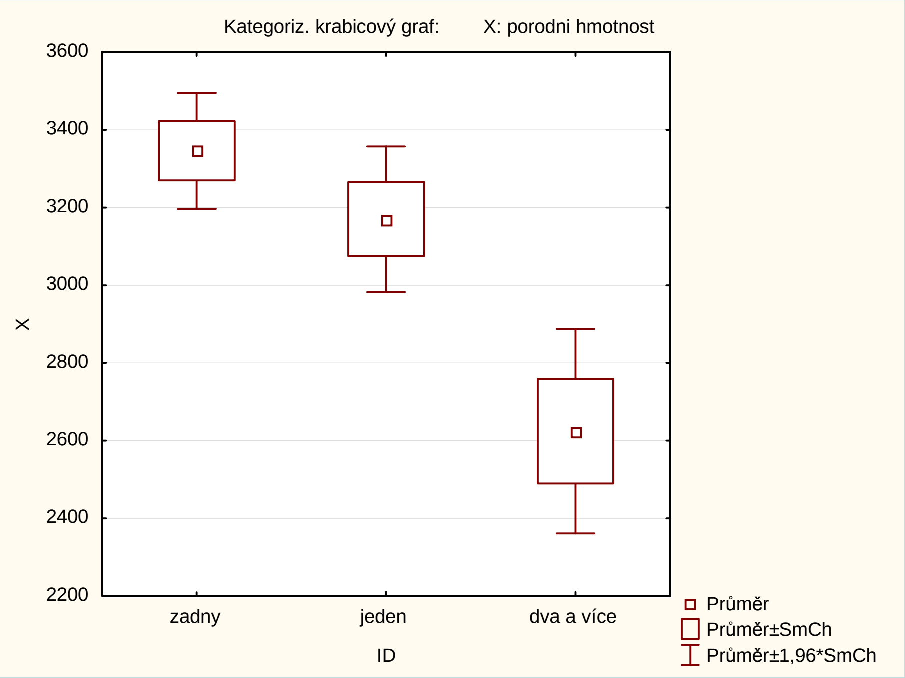
Vidíme, že p-hodnota Levenova testu je 0,5678, tedy větší než hladina významnosti 0,05. Hypotézu o shodě rozptylů nezamítáme na hladině významnosti 0,05.

Přistoupíme k testu hypotézy o shodě středních hodnot.
Aktivujeme Statistiky dle skupin - Analýza rozptylu – Výpočet.

Proměnná	Analýza rozptylu (porodni_hmotnost_sourozenci.sta) Označ. efekty jsou význ. na hlad. p < ,05000							
	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
X	6185561	2	3092780	27823289	95	292876,7	10,56001	0,000072

Jelikož p-hodnota = 0,000072 je menší než hladina významnosti 0,05, hypotézu o shodě středních hodnot zamítáme na hladině významnosti 0,05.

Výpočet doplníme krabicovými diagramy: Na záložce Základní výsledky vybereme Kategorizovaný krabicový graf



Nyní aplikujeme Schefféño metodu mnohonásobného porovnávání, abychom zjistili, které dvojice variant počtu starších sourozenců se liší na hladině významnosti 0,05. Na záložce Post – hoc zvolíme Schefféův test.

ID	Scheffeho test; proměn.:X (porodni_hmotnost_sourozenci.sta) Označ. rozdíly jsou významné na hlad. p < ,05000		
	{1} M=3346,1	{2} M=3170,0	{3} M=2624,4
zadny {1}		0,347438	0,000072
jeden {2}	0,347438		0,004891
dva a více {3}	0,000072	0,004891	

Tabulka obsahuje p-hodnoty pro vzájemné porovnání středních hodnot porodních hmotnosti daných tří sku-pin novorozenců. Vidíme, že na hladině významnosti 0,05 se neliší pouze prvorození od druhorozených.

1.7. Význam předpokladů v analýze rozptylu

a) Nezávislost jednotlivých náhodných výběrů – velmi důležitý předpoklad, musí být splněn, jinak dostaneme nesmyslné výsledky.

b) Normalita – ANOVA není příliš citlivá na porušení normality, zvlášť pokud mají všechny výběry rozsah nad 20 (důsledek centrální limitní věty). Při výraznějším porušení normality se doporučuje Kruskalův – Wallisův test.

c) Shoda rozptylů – mírné porušení nevadí, při větším se doporučuje mediánový test. Test shody rozptylů má smysl provádět až po ověření předpokladu normality.

2. Neparametrické obdoby t-testů a ANOVY

Motivace: Při aplikaci t-testů a ANOVY (tj. parametrických testů) by měly být splněny určité předpoklady:

- normalita dat (pro výběry větších rozsahů ($n \geq 30$) nemá mírné porušení normality závažný dopad na výsledky)
- homogenita rozptylů
- intervalový či poměrový charakter dat

Pokud nejsou tyto předpoklady splněny, použijeme tzv. neparametrické testy, které nevyžadují předpoklad o normalitě, stačí např. předpokládat, že distribuční funkce rozložení, z něhož náhodný výběr pochází, je spojitá.

Nevýhoda - ve srovnání s klasickými parametrickými testy jsou neparametrické testy slabší, tzn., že nepravdivou hypotézu zamítají s menší pravděpodobností než testy parametrické.

Uvedeme několik neparametrických testů, které jsou založeny na pořadí a týkají se mediánů. Nazývají se pořadové testy.

2.1. Přehled parametrických a neparametrických testů

Situace	Parametrický test	Neparametrický test
Jeden jednorozměrný výběr	Jednovýběrový t-test	Jednovýběrový Wilcoxonův test
Jeden dvourozměrný výběr	Párový t-test	Párový Wilcoxonův test
Dva nezávislé výběry	Dvouvýběrový t-test	Dvouvýběrový Wilcoxonův test (Mannův – Whitneyův test)
Aspoň tři nezávislé výběry	Jednofaktorová ANOVA	Kruskalův – Wallisův test Mediánový test

2.2. Pojem pořadí a průměrného pořadí

Nechť $X_1, ..., X_n$ je náhodný výběr.

Vektor $(X_{(1)}, ..., X_{(n)})$, kde $X_{(1)} \leq ... \leq X_{(n)}$ se nazývá uspořádaný náhodný výběr a statistika $X_{(i)}$ se nazývá i-tá pořádková statistika, $i = 1, ..., n$.

Pořadím R_i statistiky X_i rozumíme počet těch náhodných veličin $X_1, ..., X_n$, které nabývají hodnoty menší nebo rovné X_i.

V praxi se může stát, že některá pozorování jsou si rovna a vytvářejí skupiny shodných čísel. Pak těmto shodným číslům přiřadíme průměrné pořadí odpovídající takové skupině.

Příklad: Máme čísla 2 1,8 2,1 2,4 1,9 2,1 2 1,8 2,3 2,2. Stanovte jejich pořadí.

Řešení:

usp.hodnoty	1,8	1,8	1,9	2	2	2,1	2,1	2,2	2,3	2,4
pořadí	1	2	3	4	5	6	7	8	9	10
průměrné pořadí	1,5	1,5	3	4,5	4,5	6,5	6,5	8	9	10

2.3. Jednovýběrový a párový Wilcoxonův test a jeho asymptotická varianta

Frank Wilcoxon (1892 – 1965): Americký statistik a chemik

Nechť $X_1, \ldots, X_n$ je náhodný výběr ze spojitého rozložení s hustotou $\varphi(x)$, která je symetrická kolem mediánu $x_{0,50}$, tj.
$\varphi(x_{0,50} + x) = \varphi(x_{0,50} - x)$. Nechť c je reálná konstanta.
Testujeme hypotézu H_0: $x_{0,50} = c$
proti oboustranné alternativě H_1: $x_{0,50} \neq c$ nebo
proti levostranné alternativě H_1: $x_{0,50} < c$ nebo
proti pravostranné alternativě H_1: $x_{0,50} > c$.

Postup provedení testu:

a) Utvoříme rozdíly $Y_i = X_i - c$, i = 1, ..., n. (Jsou-li některé rozdíly nulové, pak za n bereme jen počet nenulových hodnot.)

b) Absolutní hodnoty $|Y_i|$ uspořádáme vzestupně podle velikosti a spočteme pořadí R_i.

c) Zavedeme statistiky

$S_W^+ = \sum_{Y_i>0} R_i^+$, což je součet pořadí přes kladné hodnoty Y_i,

$S_W^- = \sum_{Y_i<0} R_i^-$, což je součet pořadí přes záporné hodnoty Y_i.

Přitom platí, že součet $S_W^+ + S_W^- = n(n+1)/2$.

Je-li H_0 pravdivá, pak $E(S_W^+) = n(n+1)/4$ a $D(S_W^+) = n(n+1)(2n+1)/24$.

d) Testová statistika = $\min(S_W^+, S_W^-)$ pro oboustrannou alternativu,

= S_W^+ pro levostrannou alternativu,

= S_W^- pro pravostrannou alternativu.

e) H_0 zamítáme na hladině významnosti α, když testová statistika je menší nebo rovna tabelované kritické hodnotě.

Asymptotická varianta jednovýběrového Wilcoxonova testu:

Pro $n \geq 30$ lze využít asymptotické normality statistiky S_W^+.

Platí-li H_0, pak $U_0 = \dfrac{S_W^+ - E(S_W^+)}{\sqrt{D(S_W^+)}} = \dfrac{S_W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \approx N(0,1)$.

Kritický obor:

pro oboustrannou alternativu $W = \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$,

pro levostrannou alternativu $W = \left(-\infty, -u_{1-\alpha}\right\rangle$,

pro pravostrannou alternativu $W = \left\langle u_{1-\alpha}, \infty\right)$

H_0 zamítáme na asymptotické hladině významnosti α, když $U_0 \in W$.

Předpoklady použití jednovýběrového Wilcoxonova testu:

- rozložení, z něhož daný náhodný výběr pochází, je spojité
- hustota tohoto rozložení je symetrická kolem mediánu
- sledovaná veličina X má aspoň ordinální charakter

(Není-li splněn předpoklad o symetrii hustoty kolem mediánu, lze použít např. znaménkový test.)

Příklad na jednovýběrový Wilcoxonův test

U 12 náhodně vybraných zemí bylo zjištěno procento populace starší 60 let:
4,9 6,0 6,1 17,6 4,5 12,3 5,7 5,3 9,6 13,5 15,7 7,2.
Na hladině významnosti 0,05 testujte hypotézu, že aspoň polovina zemí má 12 % obyvatel nad 60 let.

Řešení: Jde o úlohu na jednovýběrový test. Jednovýběrový t-test nelze použít, protože daný náhodný výběr se neřídí normálním rozložením. Použijeme jednovýběrový Wilcoxonův test. Testujeme H_0: $x_{0,50} = 12$ proti H_1: $x_{0,50} \neq 12$.

Výpočet pomocí systému STATISTICA:

Otevřeme datový soubor populace_nad_60.sta se dvěma proměnnými a 12 případy. Proměnná X obsahuje procento populace starší 60 let a v proměnné konst je číslo 12.
Statistiky – Neparametrická statistika – Porovnání dvou závislých vzorků – OK – 1. seznam proměnných X, Druhý seznam proměnných konst – OK – Wilcoxonův párový test.

	Wilcoxonův párový test (populace_nad_60.sta) Označené testy jsou významné na hladině p <,05000			
Dvojice proměnných	Počet platných	T	Z	p-hodn.
X & konst	12	13,00000	2,039608	0,041390

Výstupní tabulka poskytne hodnotu testové statistiky SW^+ (zde označena T), hodnotu asymptotické testové statistiky U_0 a p-hodnotu pro U_0. V tomto případě je p-hodnota 0,04139, tedy nulová hypotéza se zamítá na asymptotické hladině významnosti 0,05. Tento postup však není korektní, protože rozsah výběru je menší než 30. Proto porovnáme testovou statistiku s tabelovanou kritickou hodnotou. Pro n = 12 a hladinu významnosti 0,05 je kritická hodnota 13. Protože $13 \leq 13$, H_0 zamítáme na hladině významnosti 0,05.

Párový Wilcoxonův test

Nechť $(X_1, Y_1), \ldots, (X_n, Y_n)$ je náhodný výběr ze spojitého dvourozměrného rozložení.

Testujeme H_0: $x_{0,50} - y_{0,50} = c$ proti H_1: $x_{0,50} - y_{0,50} \neq c$ (resp. proti jednostranným alternativám).

Utvoříme rozdíly $Z_i = X_i – Y_i$, $i = 1, \ldots, n$ a testujeme hypotézu o mediánu $z_{0,50}$, tj. H_0: $z_{0,50} = c$ proti H_1: $z_{0,50} \neq c$.

Příklad na párový Wilcoxonův test

V rámci psychologického výzkumu se zjišťoval vliv vážné hudby na koncentraci studentů. Každý z devíti náhodně vybraných studentů vyřešil úkol v tichém prostředí a poté jiný, stejně obtížný úkol v prostředí, v němž hrála vážná hudba. Zaznamenával se čas (v s) potřebný k řešení úkolů.

Číslo studenta	1	2	3	4	5	6	7	8	9
Ticho	63	52	55	50	70	72	51	74	70
Vážná hudba	104	92	71	88	72	49	86	42	101

Na hladině významnosti 0,05 testujte hypotézu, že poslech vážné hudby neovlivňuje čas potřebný k řešení úkolu.

Řešení: Jde o úlohu na párový test. Nelze použít párový t-test, protože je porušena normalita rozdílového náhodného výběru. Přejdeme tedy k párovému Wilcoxonovu test.
Testujeme hypotézu H_0: $z_{0,50} = 0$ proti oboustranné alternativě H_1: $z_{0,50} \neq 0$.

Výpočet pomocí systému STATISTICA:

Otevřeme datový soubor koncentrace_vazna_hudba.sta se dvěma proměnnými X, Y a 9 případy. Statistiky – Neparametrická statistika – Porovnání dvou závislých vzorků – OK – 1. seznam proměnných X, 2. seznam proměnných Y – OK – Wilcoxonův párový test.

	Wilcoxonův párový test (koncentrace_vazna_hudba.sta) Označené testy jsou významné na hladině p <,05000			
Dvojice proměnných	Počet platných	T	Z	p-hodn.
X & Y	9	8,000000	1,717812	0,085832

Testová statistika (zde označená jako T) nabývá hodnoty 8, asymptotická testová statistika (označená jako Z) nabývá hodnoty 1,7178, odpovídající asymptotická p-hodnota je 0,0858, tedy na asymptotické hladině významnosti 0,05 nulovou hypotézu nezamítáme.

Nejsou však splněny podmínky pro použití asymptotické varianty testu. Proto porovnáme testovou statistiku s tabelovanou kritickou hodnotou. Pro n = 9 a hladinu významnosti 0,05 je kritická hodnota 5. Protože 8 > 5, nelze H_0 zamítnout na hladině významnosti 0,05. Neprokázali jsme tedy, že poslech vážné hudby ovlivňuje koncentraci studentů.

Příklad na asymptotickou variantu Wilcoxonova testu:
30 náhodně vybraných osob mělo nezávisle na sobě bez předchozího nácviku odhadnout, kdy od daného signálu uplyne právě 1 minuta. Byly získány následující výsledky (v sekundách):
53 48 45 55 63 51 66 56 50 58 61 51 64 63 59 47 46 58 52 56 61 57 48 62 54 49 51 46 53 58.
Na asymptotické hladině významnosti 0,05 testujte hypotézu, že polovina osob délku jedné minuty podhodnotí a druhá nadhodnotí.

Řešení: Testujeme H_0: $x_{0,50} = 60$ proti oboustranné alternativě H_1: $x_{0,50} \neq 60$.
Výpočet pomocí systému STATISTICA:
Otevřeme datový soubor odhad_minuty.sta se dvěma proměnnými a 30 případy. V proměnné odhad jsou zjištěné hodnoty a v proměnné konst je číslo 60.
Statistiky – Neparametrická statistika – Porovnání dvou závislých vzorků – OK – 1. seznam proměnných odhad, 2. seznam proměnných konst – OK – Wilcoxonův párový test.

	Wilcoxonův párový test (odhad minuty) Označené testy jsou významné na hladině p <,05000			
Dvojice proměnných	Počet platných	T	Z	Úroveň p
odhad & konst	30	55,00000	3,650880	0,000261

Testová statistika (zde označená jako T) nabývá hodnoty 55, asymptotická testová statistika (označená jako Z) nabývá hodnoty 3,65088, odpovídající asymptotická p-hodnota je 0,000261, tedy na asymptotické hladině významnosti 0,05 nulovou hypotézu zamítáme. S rizikem omylu nejvýše 5% jsme tedy prokázali, že pravděpodobnost nadhodnocení jedné minuty není stejná jako pravděpodobnost podhodnocení.

2.4. Dvouvýběrový Wilcoxonův test a jeho asymptotická varianta

Nechť $X_1, ..., X_n$ a $Y_1, ..., Y_m$ jsou dva nezávislé náhodné výběry ze dvou spojitých rozložení, jejichž distribuční funkce se mohou lišit pouze posunutím. Označme $x_{0,50}$ medián prvního rozložení a $y_{0,50}$ medián druhého rozložení. Na hladině významnosti 0,05 testujeme hypotézu, že distribuční funkce těchto rozložení jsou shodné neboli mediány jsou shodné proti alternativě, že jsou rozdílné, tj.

H_0: $x_{0,50} - y_{0,50} = 0$ proti H_1: $x_{0,50} - y_{0,50} \neq 0$.

Postup provedení testu:

a) Všech n + m hodnot $X_1, ..., X_n$ a $Y_1, ..., Y_m$ uspořádáme vzestupně podle velikosti.

b) Zjistíme součet pořadí hodnot $X_1, ..., X_n$ a označíme ho T_1.
Součet pořadí hodnot $Y_1, ..., Y_m$ označíme T_2.

c) Vypočteme statistiky $U_1 = mn + n(n+1)/2 – T_1$, $U_2 = mn + m(m+1)/2 - T_2$.
Přitom platí $U_1 + U_2 = mn$.

d) Pokud $\min(U_1,U_2) \leq$ tabelovaná kritická hodnota (pro dané rozsahy výběrů m, n a dané α), pak nulovou hypotézu o totožnosti obou distribučních funkcí zamítáme na hladině významnosti α. V tabulkách: $n = \min\{m,n\}$ a $m = \max\{m,n\}$.

Asymptotická varianta dvouvýběrového Wilcoxonova testu:

Pro velká n, m (n, m > 30) lze využít asymptotické normality statistiky U_1'.

Platí-li H_0, pak $U_0 = \dfrac{U_1' - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \approx N(0,1)$, kde $U_1' = \min(U_1, U_2)$.

Kritický obor:

pro oboustrannou alternativu $W = \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$,

pro levostrannou alternativu $W = \left(-\infty, -u_{1-\alpha}\right\rangle$,

pro pravostrannou alternativu $W = \left\langle u_{1-\alpha}, \infty\right)$

H_0 zamítáme na asymptotické hladině významnosti α, když $U_0 \in W$.

Předpoklady použití dvouvýběrového Wilcoxonova testu:

- dané dva náhodné výběry jsou nezávislé
- rozložení, z nichž dané dva náhodné výběry pocházejí, jsou spojitá
- distribuční funkce těchto rozložení se mohou lišit pouze posunutím
- sledovaná veličina má aspoň ordinální charakter

(Není-li splněn předpoklad, že distribuční funkce se mohou lišit pouze posunutím, lze použít např. dvouvýběrový Kolmogorovův – Smirnovův test.)

Příklad na dvouvýběrový Wilcoxonův test:

Bylo vybráno 10 polí stejné kvality. Na čtyřech z nich se zkoušel nový způsob hnojení, zbylých šest bylo ošetřeno starým způsobem. Pole byla oseta pšenicí a sledoval se její hektarový výnos. Je třeba zjistit, zda nový způsob hnojení má týž vliv na průměrné hektarové výnosy pšenice jako starý způsob hnojení.

hektarové výnosy při novém způsobu: 51 52 49 55

hektarové výnosy při starém způsobu: 45 54 48 44 53 50

Test proveďte na hladině významnosti 0,05.

Řešení:

Na hladině významnosti 0,05 testujeme H_0: $x_{0,50} - y_{0,50} = 0$ proti oboustranné alternativě H_1: $x_{0,50} - y_{0,50} \neq 0$.

Kritická hodnota pro $\alpha = 0,05$, min(4,6) = 4, max(4,6) = 6 je 2.

Otevřeme datový soubor hojeni.sta se dvěma proměnnými a 10 případy. V proměnné vynos jsou zjištěné hodnoty a v proměnné hnojeni je 4x číslo 1 pro nový způsob hnojení a 6x číslo 2 pro starý způsob hnojení.

Statistiky – Neparametrická statistika – Porovnání dvou nezávislých vzorků – OK – Proměnné – Seznam závislých proměnných vynos, Nezáv. (grupov.) proměnná hnojeni – OK – M-W U test.

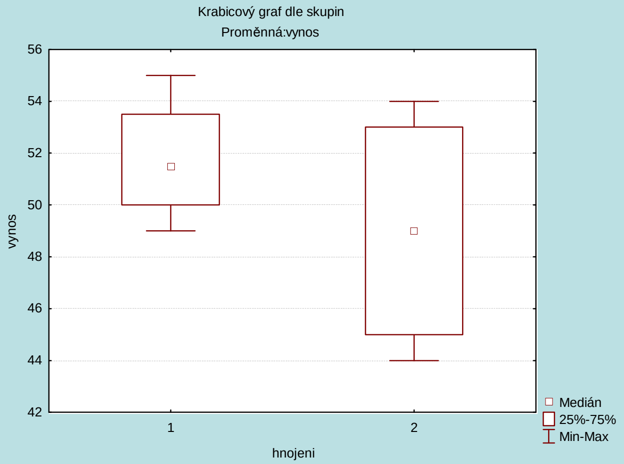
Upozornění: Ve STATISTICE je dvouvýběrový Wilcoxonův test uveden pod názvem Mannův – Whitneyův test.

Mann-Whitneyův U test (vynos)
Dle proměn. hnojeni
Označené testy jsou významné na hladině p <,05000

Proměnná	Sčt poř. skup. 1	Sčt poř. skup. 2	U	Z	Úroveň p	Z upravené	Úroveň p	N platn. skup. 1	N platn. skup. 2	2*1str. přesné p
vynos	27,00000	28,00000	7,000000	1,066004	0,286423	1,066004	0,286423	4	6	0,352381

Ve výstupní tabulce jsou součty pořadí T_1, T_2, hodnota testové statistiky min(U_1, U_2) označená U, hodnota asymptotické testové statistiky U_0 (označená Z), asymptotická p-hodnota pro U_0 a přesná p-hodnota (ozn. 2*1str. přesné p – ta se používá pro rozsahy výběrů pod 30). V našem případě přesná p-hodnota = 0,352381, tedy H_0 nezamítáme na hladině významnosti 0,05.

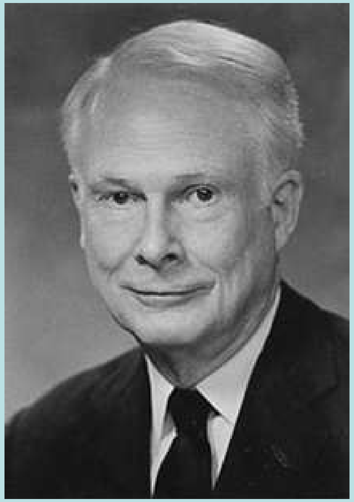
Výpočet je vhodné doplnit krabicovým diagramem.



Je zřejmé, že výnosy při starém způsobu hnojení jsou vesměs nižší než při novém způsobu a také vykazují mnohem větší variabilitu.

2.5. Kruskalův - Wallisův test

William Kruskal (1919 – 2005):
Americký matematik



Wilson Allen Wallis (1912 – 1988):
Americký matematik

Nechť je dáno $r \geq 3$ nezávislých náhodných výběrů o rozsazích $n_1, \dots , n_r$. Předpokládáme, že tyto výběry pocházejí ze spojitých rozložení. Označme $n = n_1 + \dots + n_r$. Na asymptotické hladině významnosti α chceme testovat hypotézu, že všechny tyto výběry pocházejí z téhož rozložení.

Postup testu:

a) Všech n hodnot seřadíme do rostoucí posloupnosti.
b) Určíme pořadí každé hodnoty v tomto sdruženém výběru.
c) Označme T_j součet pořadí těch hodnot, které patří do j-tého výběru, j = 1, ..., r (kontrola: musí platit $T_1 + ... + T_r = n(n+1)/2$).
d) Testová statistika má tvar: $Q = \frac{12}{n(n+1)} \sum_{j=1}^{r} \frac{T_j^2}{n_j} - 3(n+1)$. Platí-li H_0, má statistika Q asymptoticky rozložení $\chi^2(r-1)$.
e) Kritický obor: $W = \langle \chi^2_{1-\alpha}(r-1), \infty)$.
f) H_0 zamítneme na asymptotické hladině významnosti α, když $Q \geq \chi_{1-\alpha}^2(r-1)$.

2.6. Mediánový test

Výchozí situace je stejná jako u K-W testu

Postup testu:

a) Všech n hodnot uspořádáme do rostoucí posloupnosti.

b) Najdeme medián $x_{0,50}$ těchto n hodnot.

c) Označme P_j počet hodnot v j-tém výběru, které jsou větší nebo rovny mediánu $x_{0,50}$.

d) Testová statistika má tvar $Q_M = 4\sum_{j=1}^{r}\frac{P_j^{\,2}}{n_j} - n$. Platí-li H_0, má statistika Q_M asymptoticky rozložení $\chi^2(r-1)$.

d) Kritický obor: $W = \left\langle \chi^2_{1-\alpha}(r-1), \infty \right)$.

e) H_0 zamítneme na asymptotické hladině významnosti α, když $Q_M \geq \chi_{1-\alpha}{}^2(r-1)$.

2.7. Metody mnohonásobného porovnávání

Zamítneme-li hypotézu, že všechny náhodné výběry pocházejí z téhož rozložení, zajímá nás, které dvojice náhodných výběrů se liší na zvolené hladině významnosti. Testujeme H_0: k-tý a l-tý náhodný výběr pocházejí z téhož rozložení, $k, l = 1, .., r$, $k \neq l$ proti H_1: aspoň jedna dvojice výběrů pochází z různých rozložení.

a) Neményiho metoda (Peter Neményi 1927 – 2002: Americký matematik maďarského původu)

- Všechny výběry mají týž rozsah p (třídění je vyvážené).
- Vypočteme $|T_l - T_k|$.
- V tabulkách najdeme kritickou hodnotu (pro dané p, r, α).
- Pokud $|T_l - T_k| \geq$ tabelovaná kritická hodnota, pak na hladině významnosti α zamítáme hypotézu, že l-tý a k-tý výběr pocházejí z téhož rozložení.

b) Obecná metoda mnohonásobného porovnávání

- Vypočteme $\left|\frac{T_l}{n_l} - \frac{T_k}{n_k}\right|$.
- Ve speciálních statistických tabulkách najdeme kritickou hodnotu $h_{KW}(\alpha)$. Při větších rozsazích výběrů je možno ji nahradit kvantilem $\chi_{1-\alpha}^{2}(r-1)$.
- Jestliže $\left|\frac{T_l}{n_l} - \frac{T_k}{n_k}\right| \geq \sqrt{\frac{1}{12}\left(\frac{1}{n_l} + \frac{1}{n_k}\right) n(n+1) h_{KW}(\alpha)}$, pak na hladině významnosti α zamítáme hypotézu, že l-tý a k-tý výběr pocházejí z téhož rozložení.

2.8. Příklad na Kruskalův – Walisův a mediánový test:

Čtyři laboranti provedli analytické stanovení procenta niklu v oceli. Každý hodnotil pět vzorků.

Laborant A: 4,15 4,26 4,10 4,30 4,25
Laborant B: 4,38 4,40 4,29 4,39 4,45
Laborant C: 4,23 4,16 4,20 4,24 4,27
Laborant D: 4,41 4,31 4,42 4,37 4,43

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že všechny čtyři náhodné výběry pocházejí ze stejného rozložení. Pokud nulovou hypotézu zamítnete, zjistěte, které dvojice výběrů se liší.

Výpočet pomocí systému STATISTICA:

Otevřeme datový soubor nikl v oceli.sta o dvou proměnných a 20 případech. V proměnné nikl jsou změřené hodnoty, v proměnné laborant je 5x1 pro 1. laboranta atd. až 5x4 pro 4. laboranta.
Statistiky – Neparametrická statistika – Porovnání více nezávislých vzorků - OK – Seznam závislých proměnných nikl, Nezáv. (grupovací) proměnná laborant – OK – Summary: Kruskal-Wallis ANOVA & Median test. Ve dvou výstupních tabulkách se objeví výsledky K-W testu a mediánového testu.

Závislá: nikl	Kruskal-Wallisova ANOVA založ. na poř.; nikl (nikl v oceli) Nezávislá (grupovací) proměnná : laborant Kruskal-Wallisův test: H (3, N= 20) =13,77714 p =,0032		
	Kód	Počet platných	Součet pořadí
1	1	5	29,00000
2	2	5	75,00000
3	3	5	27,00000
4	4	5	79,00000

Závislá: nikl	Mediánový test, celk. medián = 4,29500; nikl (nikl v oceli) Nezávislá (grupovací) proměnná : laborant Chi-Kvadr. = 13,60000 sv = 3 p = ,0035				
	1	2	3	4	Celkem
<= Medián: pozorov.	4,00000	1,00000	5,00000	0,00000	10,00000
očekáv.	2,50000	2,50000	2,50000	2,50000	
poz.-oč.	1,50000	-1,50000	2,50000	-2,50000	
> Medián: pozorov.	1,00000	4,00000	0,00000	5,00000	10,00000
očekáv.	2,50000	2,50000	2,50000	2,50000	
poz.-oč.	-1,50000	1,50000	-2,50000	2,50000	
Celkem: oček.	5,00000	5,00000	5,00000	5,00000	20,00000

Oba testy zamítají hypotézu o shodě mediánů v daných čtyřech skupinách na asymptotické hladině významnosti 0,05.

Nyní provedeme mnohonásobné porovnávání, abychom zjistili, které dvojice laborantů se liší. Zvolíme Vícenás. porovnání průměrného pořadí pro vš. skupiny.

	Vícenásobné porovnání p hodnot (oboustr.) nikl (nikl v oceli Nezávislá (grupovací) proměnná: laborant Kruskal-Wallisův test: H (3, N= 20) =13,77714 p =,0032			
Závislá: nikl	1 R:5,8000	2 R:15,000	3 R:5,4000	4 R:15,800
1		0,083641	1,000000	0,045158
2	0,083641		0,061779	1,000000
3	1,000000	0,061779		0,032664
4	0,045158	1,000000	0,032664	

Tabulka obsahuje p-hodnoty pro porovnání dvojic skupin. Vidíme, že na hladině významnosti 0,05 se liší laboranti A, D a laboranti C, D.

Grafické znázornění výsledků