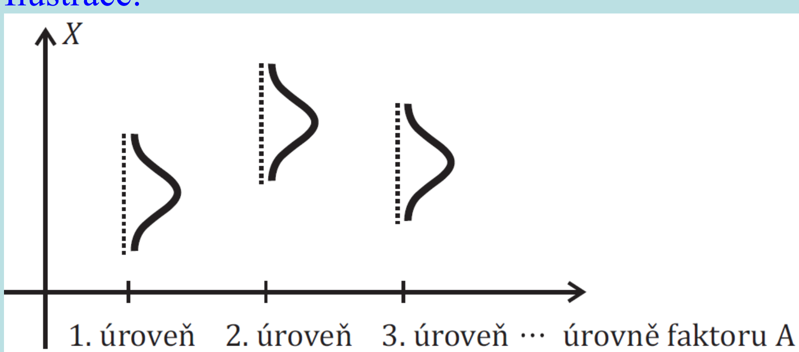
Parametrické úlohy o více nezávislých náhodných výběrech

I. Případ r ≥ 3 nezávislých náhodných výběrů z normálních rozložení (Analýza rozptylu jednoduchého třídění)

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 ≥ 3 úrovní a přitom i-té úrovni odpovídá n_i pozorování $X_{i1}, \ldots, X_{in_i}$, 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, …, r, j = 1, …, n_i.
Výsledky lze zapsat do tabulky

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

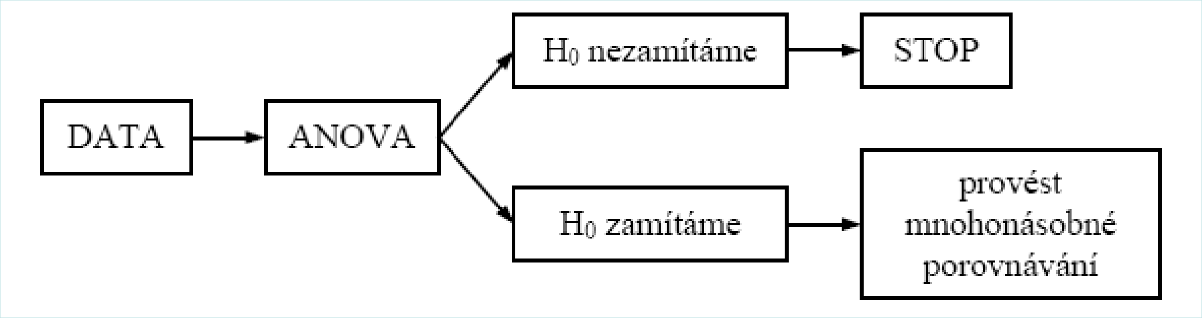
Ilustrace:



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. Schefféo nebo Tukeyova metoda.



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

$S_T = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij} - M_{..})^2$ … celkový součet čtverců (charakterizuje variabilitu jednotlivých pozorování kolem celkového průměru),

počet stupňů volnosti $f_T = n - 1$,

$S_A = \sum_{i=1}^{r} n_i (M_{i.} - M_{..})^2$ … skupinový součet čtverců (charakterizuje variabilitu mezi jednotlivými náhodnými výběry),

počet stupňů volnosti $f_A = r - 1$.

Sčítanec $(M_{i.} - M_{..})$ představuje bodový odhad efektu α_i.

$S_E = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij} - M_{i.})^2$ … reziduální součet čtverců (charakterizuje variabilitu uvnitř jednotlivých výběrů),

počet stupňů volnosti $f_E = n - r$.

Lze dokázat, že $S_T = S_A + S_E$.

(Důkaz je proveden např. ve skriptech Budíková, Mikoláš, Osecký: Popisná statistika v poznámce 5.20.)

Testování hypotézy o shodě středních hodnot

Náhodné veličiny X_{ij} se řídí modelem

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

pro $i = 1, \ldots, r$, $j = 1, \ldots, n_i$, 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: $\sum_{i=1}^{r} \ldots 0$.

(Pokud je třídění vyvážené, tj. pokud mají všechny výběry stejný rozsah: $n_1 = n_2 = \ldots = n_r$, pak lze použít zjednodušenou podmínku $\sum_{i=1}^{r} \ldots 0$.)

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 $\langle 0,1 \rangle$.

Testování hypotézy o shodě rozptylů

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

a) Levenův test: Položme $Z_{ij} = |X_{ij} - M_i|$. Označíme

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

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

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

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

Platí-li hypotéza o shodě rozptylů, pak statistika

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

Hypotézu o shodě rozptylů tedy zamítáme na asymptotické hladině významnosti α, když $F_{ZA} \geq F_{1-\alpha}(r-1, n-r)$.

(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 $W = \langle \chi^2_{1-\alpha}(r-1), \infty)$.

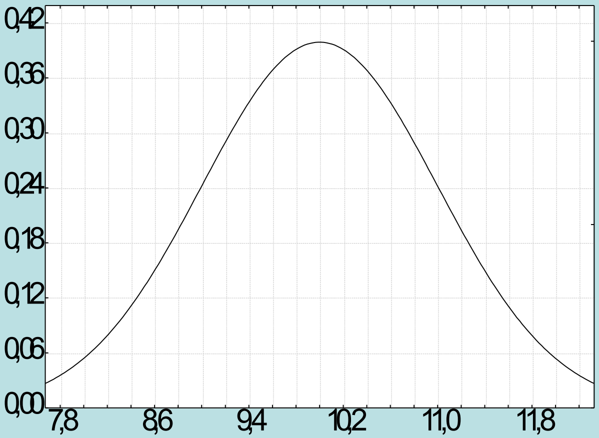
Zkoumání vlastností uvedených tří testů

Pro odhad pravděpodobnosti chyby 1. druhu bylo vždy vygenerováno 100 000 náhodných výběrů, a to postupně z těchto rozložení: N(10; 1), t(10), LN(1; 0,4), Ex(0,85). Všechny výběry měly stejný rozsah od 3 do 11 s krokem 2, počet výběrů byl od 2 do 10 s krokem 2.

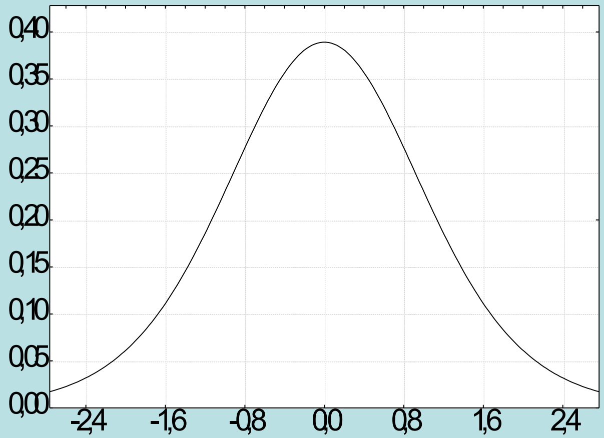
Jako odhad pravděpodobnosti chyby 1. druhu sloužila relativní četnost těch případů, kdy se na hladině významnosti 0,05 zamítla nulová hypotéza o shodě rozptylů. Simulace byly provedeny v programu MathCad.

Grafy hustot zkoumaných rozložení

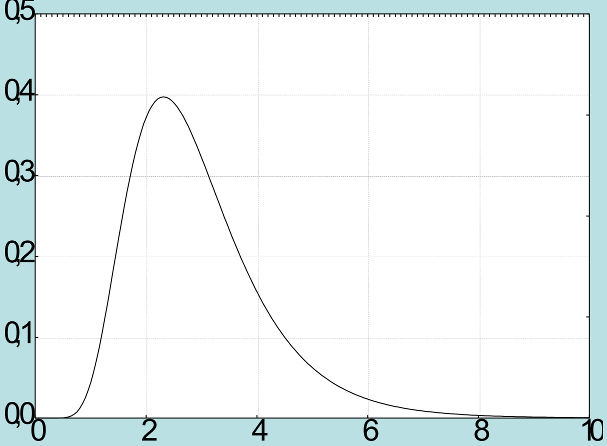
Normální rozložení N(10; 1)



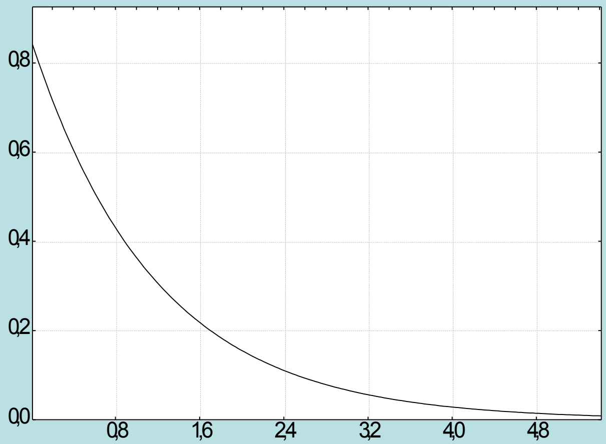
Studentovo rozložení t(10)



Log – normální rozložení LN(1; 0,4)



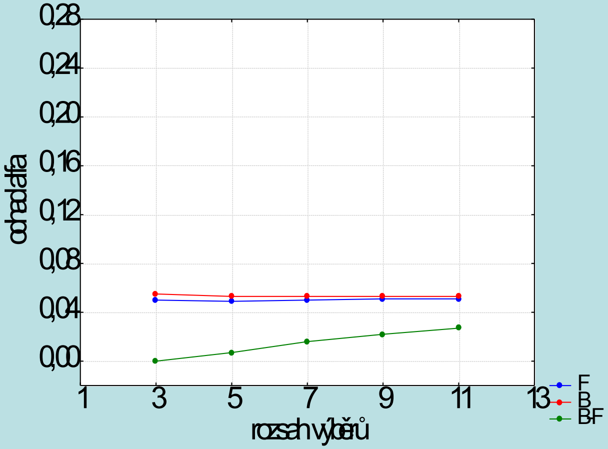
Exponenciální rozložení Ex(0,85)



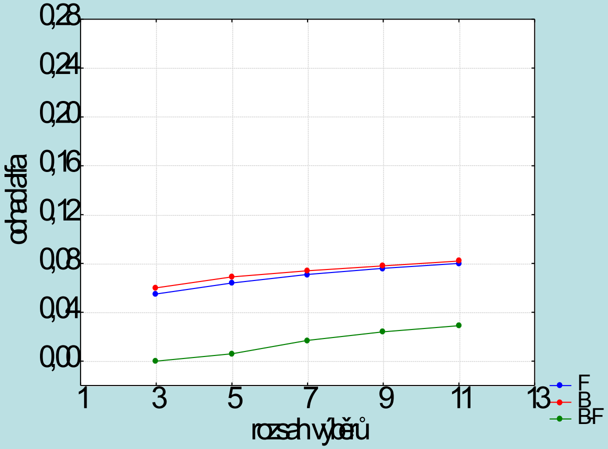
Případ dvou nezávislých náhodných výběrů

Nejprve bylo provedeno srovnání F-testu s Bartlettovým testem a Brownovým – Forsytheovým testem pro dva nezávislé náhodné výběry. V grafech se modrá barva vztahuje k F-testu, červená k Bartlettovu testu a zelená k Brownovu – Forsytheovu testu.

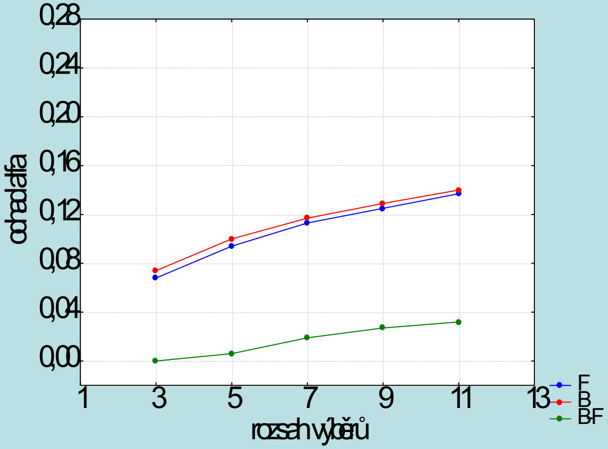
Normální rozložení N(10; 1)



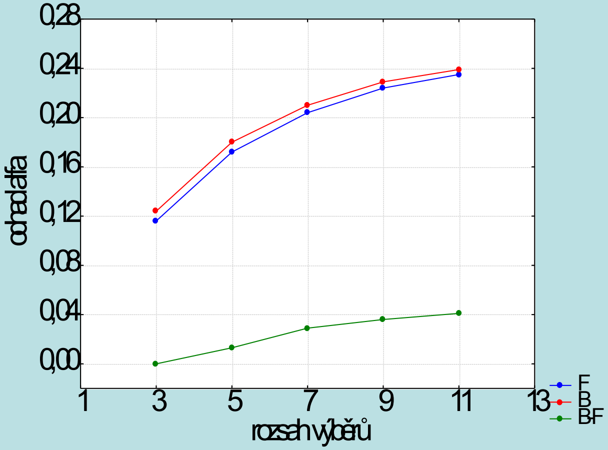
Studentovo rozložení t(10)



Log - normální rozložení LN(1; 0,4)



Exponenciální rozložení Ex(0,85)



Komentář: Podle očekávání je nejnižších odhadů pravděpodobnosti chyby 1. druhu dosahováno pro výběry z normálního rozložení, kdy všechny testy udrží odhad pod hladinou významnosti 0,05. S postupným „vzdalováním se“ od normality relativní četnost neoprávněného zamítnutí nulové hypotézy roste, nejvyšší je pro výběry z exponenciálního rozložení, kde se pro F-test a Bartlettův test blíží k 0,24.

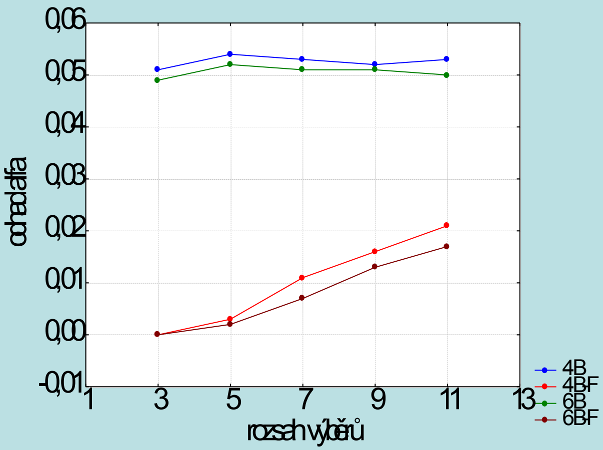
Pro všechna zkoumaná rozložení dávají F-test a Bartlettův test srovnatelné výsledky, u F-testu pozorujeme poněkud nižší odhad. Jednoznačně nejlepší výsledky jsou dosahovány při použití B-F testu, který i pro výběry z exponenciálního rozložení poskytuje odhad pravděpodobnosti chyby 1. druhu dostatečně hluboko pod 0,05.

Případ více než dvou nezávislých náhodných výběrů

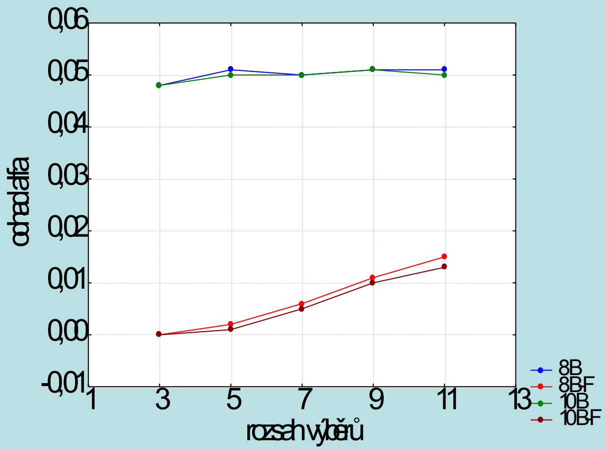
Dále jsme se zabývali srovnáním Bartlettova testu s Brownovým – Forsytheovým testem pro 4, 6, 8 a 10 nezávislých náhodných výběrů, jejichž rozsahy byly 3, 5, 7, 9, 11. Kvůli větší přehlednosti jsou grafy závislosti odhadu na rozsahu výběrů uvedeny zvlášť pro 4 a 6 výběrů a poté pro 8 a 10 výběrů. V grafech se modrá a zelená barva vztahuje k Bartlettovu testu, červená a hnědá pak k Brownovu – Forsytheovu testu.

a) Normální rozložení N(10; 1)

Počet výběrů 4 a 6



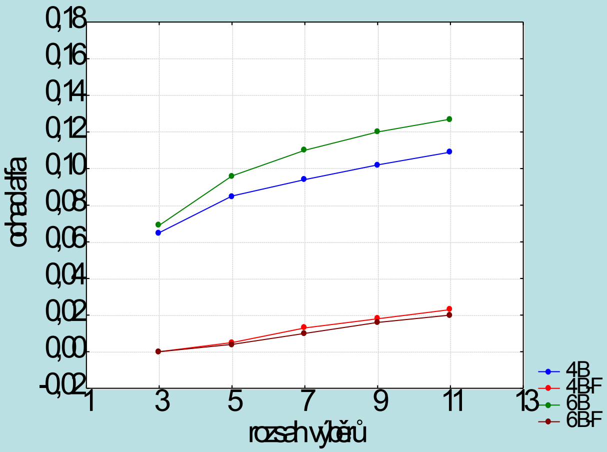
Počet výběrů 8 a 10



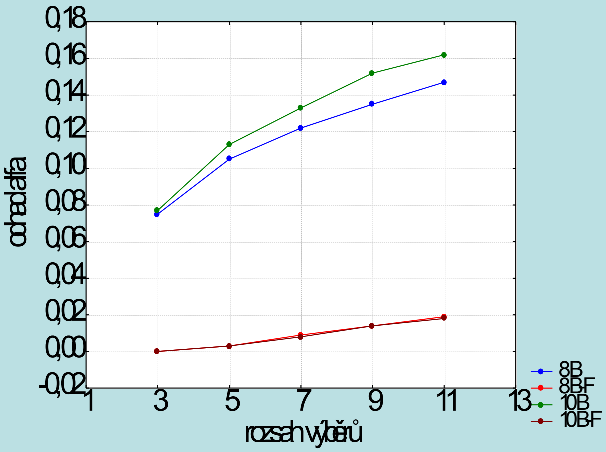
Pro výběry z normálního rozložení dává Bartlettův test odhady velmi blízké hladině významnosti 0,05. Není zde pozorovatelná závislost na rozsahu výběrů. Brownův – Forsytheův test neoprávněně zamítá nulovou hypotézu s podstatně menší relativní četností, která nepřesáhne 0,021.

b) Studentovo rozložení t(10)

Počet výběrů 4 a 6



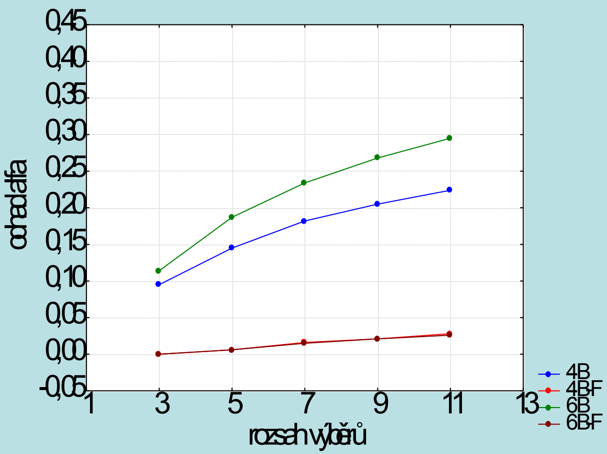
Počet výběrů 8 a 10



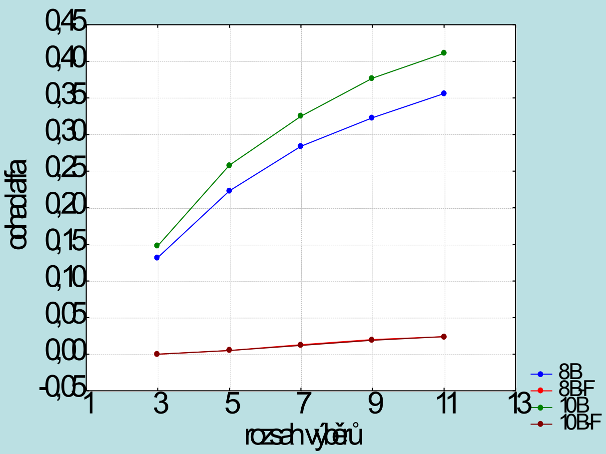
Pro výběry ze Studentova rozložení jsou výsledky Bartlettova testu již ovlivněny porušením předpokladu normality. Získané odhady narůstají se zvětšujícím se rozsahem výběrů a v nejméně příznivém případě, tj. pro 10 nezávislých náhodný výběrů o rozsahu 11, odhad pravděpodobnosti chyby 1. druhu převyšuje 0,16. Brownův – Forsytheův test neoprávněně zamítá nulovou hypotézu s relativní četností, která nepřesáhne 0,023. Rozdíly mezi odhady pro různé počty výběrů jsou u B-F testu zanedbatelně malé.

c) Logaritmicko – normální rozložení LN(1; 0,4)

Počet výběrů 4 a 6



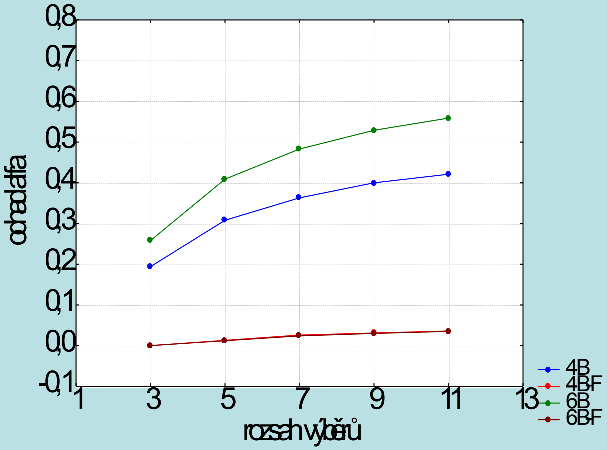
Počet výběrů 8 a 10



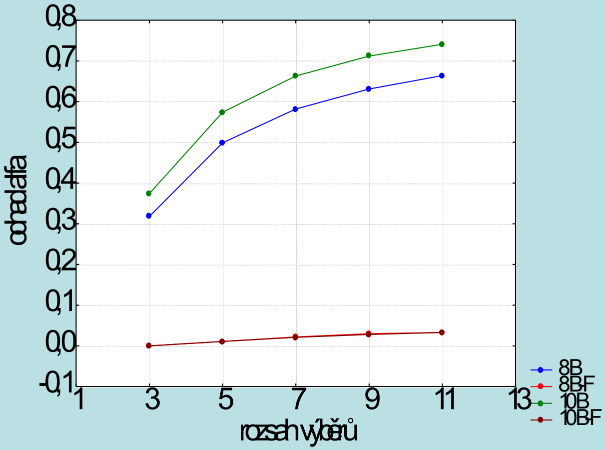
Pro výběry z logaritmicko - normálního rozložení odhad pravděpodobnosti chyby 1. druhu získaný Bartlettovým testem velmi výrazně narůstá, zvláště pro větší počet rozsáhlejších výběrů. Zde je dokonce o něco vyšší než 0,42, tudíž použití Bartlettova testu skutečně nelze doporučit. Daleko lepší výsledky poskytuje Brownův – Forsytheův test, kde odhady zůstávají pod 0,03.

d) Exponenciální rozložení Ex(0,85)

Počet výběrů 4 a 6



Počet výběrů 8 a 10



Vidíme, že použití Bartlettova testu pro výběry z exponenciálního rozložení nelze vůbec doporučit. Odhad pravděpodobnosti chyby 1. druhu je neúnosně velký, v nejméně příznivém případě – pro 10 nezávislých náhodných výběrů o rozsahu 11 - se tento odhad blíží 0,75. Naproti tomu odhady získané Brownovým – Forsytheovým testem jsou nanejvýš 0,035, což ještě zdaleka nedosahuje hladiny významnosti 0,05.

Komentář

Výsledky našich simulačních studií vedou k závěru, že pro testy homogenity rozptylů je vhodné používat Brownův – Forsytheův test, a to jak pro dva, tak pro více nezávislých náhodných výběrů. Ukazuje se, že tento test lze aplikovat i na výběry, které pocházejí z výrazně nenormálních rozložení. To lze vysvětlit tím, že při jeho konstrukci jsou použity výběrové mediány jednotlivých výběrů, přičemž medián – na rozdíl od průměru – je robustní vůči odlehlým či extrémním hodnotám.

U Brownova – Forsytheova testu odhad pravděpodobnosti chyby 1. druhu ve všech případech zůstal pod hladinou významnosti 0,05, nejhorší výsledek byl 0,036 pro 4 nezávislé výběry z exponenciálního rozložení. Bartlettův test zcela selhává pro výběry z nesymetrických rozložení. Např. pro 10 nezávislých výběrů z exponenciálního rozložení, jejichž rozsah byl 11, se odhad pravděpodobnosti chyby 1. druhu blížil číslu 0,8.

Výhodou Brownova – Forsytheova testu je rovněž skutečnost, že velikosti odhadů vykazují jen velmi nepatrnou závislost na počtu výběrů.

Brownův – Forsytheův test je implemenován např. v systémech STATISTICA či MINITAB, Bartlettův test najdeme v systému MINITAB, F-test pak v obou zmíněných systémech.

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 $\frac{|M_k - M_l|}{\frac{S_*}{\sqrt{p}}}$. Rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

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

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

$$|M_k - M_l| \geq S^* \sqrt{(r-1)\left(\frac{1}{n_k} + \frac{1}{n_l}\right) F_{1-\alpha}(r-1, n-r)}.$$

Výhodou Schefféhо 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 Schefféhо 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.

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.

Příklad: U čtyř odrůd brambor (označených symboly A, B, C, D) se zjišťovala celková hmotnost brambor vyrostlých vždy z jednoho trsu. Výsledky (v kg):

odrůda	hmotnost
A	0,9 0,8 0,6 0,9
B	1,3 1,0 1,3
C	1,3 1,5 1,6 1,1 1,5
D	1,1 1,2 1,0

Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota hmotnosti trsu brambor nezávisí na odrůdě. Zamítnete-li nulovou hypotézu, zjistěte, které dvojice odrůd se liší na hladině významnosti 0,05.

Řešení:

Data považujeme za realizace čtyř nezávislých náhodných výběrů ze čtyř normálních rozložení se stejným rozptylem. Testujeme hypotézu, že všechny čtyři střední hodnoty jsou stejné.

Vypočítáme výběrové průměry v jednotlivých výběrech: $M_{1.} = 0{,}8$, $M_{2.} = 1{,}2$, $M_{3.} = 1{,}4$, $M_{4.} = 1{,}1$,
celkový průměr: $M_{..} = 1{,}14$,
výběrové rozptyly: $S_1^2 = 0{,}02$, $S_2^2 = 0{,}03$, $S_3^2 = 0{,}04$, $S_4^2 = 0{,}01$,
vážený průměr výběrových rozptylů: $S_*^2 = \dfrac{\sum_{i=1}^{r}(n_i-1)S_i^2}{n-r} = \dfrac{3\cdot 0{,}02 + 2\cdot 0{,}03 + 4\cdot 0{,}04 + 2\cdot 0{,}01}{11} = \dfrac{0{,}3}{11} = 0{,}027$,
reziduální součet čtverců: $S_E = (n-r)S_*^2 = 11\cdot\dfrac{0{,}3}{11} = 0{,}3$,
skupinový součet čtverců: $S_A = \sum_{i=1}^{r} n_i (M_{i.} - M_{..})^2 = 4(0{,}8-1{,}14)^2 + 3(1{,}2-1{,}14)^2 + 5(1{,}4-1{,}14)^2 + 3(1{,}1-1{,}14)^2 = 0{,}816$
celkový součet čtverců: $S_T = S_A + S_E = 0{,}816 + 0{,}3 = 1{,}116$,
testová statistika $F_A = \dfrac{S_A/f_A}{S_E/f_E} = \dfrac{0{,}816/3}{0{,}3/11} = 9{,}97$,

Kritický obor $W = \langle F_{0,95}(3,11), \infty) = \langle 3{,}59, \infty)$. Protože testová statistika se realizuje v kritickém oboru, H_0 zamítáme na hladině významnosti 0,05.

Vypočteme poměr determinace: $P^2 = \dfrac{S_A}{S_T} = \dfrac{0{,}816}{1{,}116} = 0{,}731$

Výsledky zapíšeme do tabulky ANOVA:

Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F_A
skupiny	$S_A = 0{,}816$	3	$S_A/3 = 0{,}272$	$\frac{S_A/(r-1)}{S_E/(n-r)} = 9{,}97$
reziduální	$S_E = 0{,}3$	11	$S_E/11 = 0{,}02727$	-
celkový	$S_T = 1{,}116$	14	-	-

Nyní pomocí Scheffého metody zjistíme, které dvojice odrůd se liší na hladině významnosti 0,05.

Srovnávané odrůdy	Rozdíly $\lvert M_k - M_l \rvert$	Pravá strana vzorce
A, B	0,4	0,41
A, C	0,67	0,36
A, D	0,3	0,41
B, C	0,2	0,40
B, D	0,1	0,44
C, D	0,3	0,40

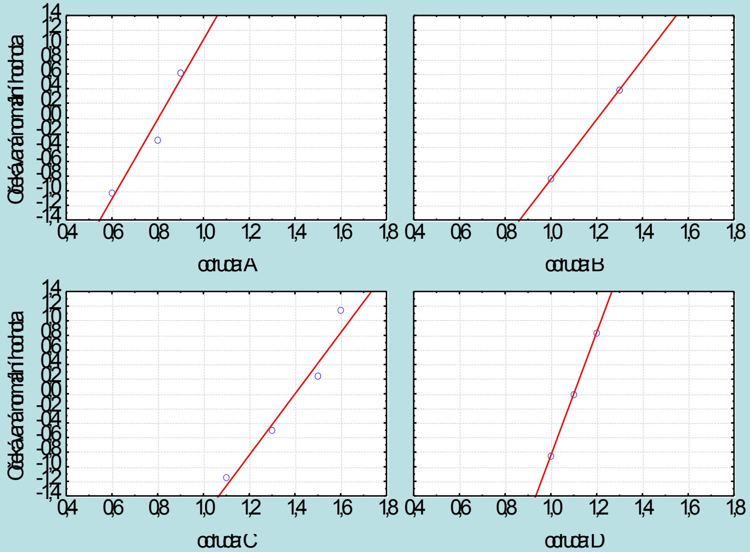
Na hladině významnosti 0,05 se liší odrůdy A a C.

Řešení pomocí systému STATISTICA

Otevřeme nový datový soubor o dvou proměnných X a odrůda a 15 případech. Do proměnné X zapíšeme zjištěné hmotnosti, do proměnné odrůda kódy pro dané odrůdy (1 pro A, 2 pro B, 3 pro C a 4 pro D).

	1 X	2 odruc
1	0,	A
2	0,	A
3	0,	A
4	0,	A
5	1,	B
6	1	B
7	1,	B
8	1,	C
9	1,	C
10	1,	C
11	1,	C
12	1,	C
13	1,	D
14	1,	D
15	1	D

Ověříme normalitu daných čtyř náhodných výběrů pomocí N-P plotu:



Odchylky od normality jsou jen nepatrné.

Vypočteme výběrové průměry a výběrové rozptyly:
Statistiky – Základní statistiky a tabulky – Rozklad & jednofakt. ANOVA – OK – Proměnné – Závislé – X, Grupovací - odrůda – OK – Skupiny tabulek - zaškrtneme Rozptyly - Výpočet.

Rozkladová tabulka popisných s
N=15 (V seznamu záv. prom. ne

odruc	X průme	X N	X Sm.od	X Rozpt
A	0,800	4	0,141	0,020
B	1,200	3	0,173	0,030
C	1,400	5	0,200	0,040
D	1,100	3	0,100	0,010
Vš.sk	1,140	1	0,282	0,079

Nyní ověříme předpoklad shody rozptylů.
Na záložce Skupiny tabulek zaškrtneme Levenův test – Výpočet.

Leveneův test homogenity rozpylů (priklad8301
Označ. efekty jsou význ. na hlad. p < ,05000

Promě	SČ efekt	SV efe	PČ efekt	SČ chyba	SV chyb	PČ chyba	F	p
X	0,018	3	0,006	0,065	1	0,005	1,047	0,410

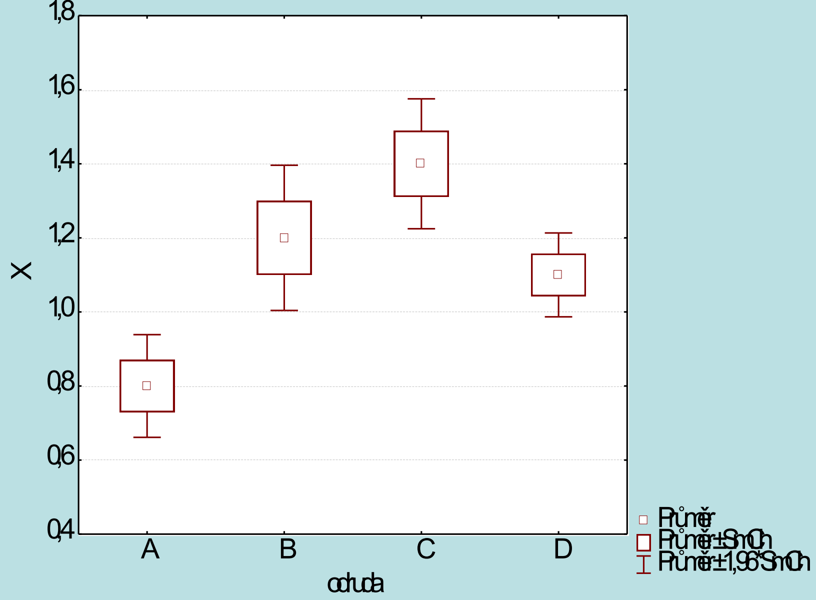
Vidíme, že p-hodnota Levenova testu je 0,41, tedy větší než hladina významnosti 0,05. Hypotézu o shodě rozptylů nezamítáme na hladině významnosti 0,05.

Přistoupíme k testu hypotézy o shodě středních hodnot.
Na záložce Skupiny tabulek zaškrtneme Analýza rozptylu – Výpočet.

	Analýza rozptylu (priklad8301) Označ. efekty jsou význ. na hlad. p < ,05000							
Proměi	SC efekt	SV efel	PC efekt	SC chyba	SV chyt	PC chyba	F	p
X	0,816	3	0,272	0,300	1	0,027	9,973	0,001

Jelikož p-hodnota = 0,001805 je menší než hladina významnosti 0,05, hypotézu o shodě středních hodnot zamítáme na hladině významnosti 0,05.

Výpočet doplníme krabicovými diagramy:



Nyní aplikujeme Schefféhο metodu mnohonásobného porovnávání, abychom zjistili, které dvojice odrůd se liší na hladině významnosti 0,05. Na záložce Post – hoc zvolíme Schefféův test.

	Scheffého test; promen.:X (prik Označ. rozdíly jsou významné r			
odruda	{1} M=,80	{2} M=1,2	{3} M=1,4	{4} M=1,1
A		0,059	0,001	0,190
B	0,059		0,464	0,905
C	0,001	0,464		0,163
D	0,190	0,905	0,163	

Tabulka obsahuje p-hodnoty pro vzájemné porovnání středních hodnot hmotnosti všech čtyř odrůd. Vidíme, že na hladině významnosti 0,05 se liší odrůdy A, C.

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 Kruskalův – Wallisův test. Test shody rozptylů má smysl provádět až po ověření předpokladu normality.

II. Případ r ≥ 3 nezávislých náhodných výběrů z alternativních rozložení

Test homogenity binomických rozložení

Nechť máme $r \geq 3$ nezávislých náhodných výběrů o rozsazích $n_1, \ldots, n_r$, přičemž j-tý náhodný výběr pochází z alternativního rozložení $A(\vartheta_j)$, $j = 1, 2, \ldots, r$.

Testujeme hypotézu H_0: $\vartheta_1 = \ldots = \vartheta_r$ proti alternativní hypotéze H_1: aspoň jedna dvojice parametrů je různá.

Označme

$n = \sum_{j=1}^{r} n_j$ celkový rozsah všech r výběrů,

$M_* = \frac{1}{n}\sum_{j=1}^{r} n_j M_j$ vážený průměr výběrových průměrů.

Testové kritérium: $Q = \frac{1}{M_*(1-M_*)}\sum_{j=1}^{r} n_j (M_j - M_*)^2 \approx \chi^2(r-1)$, když H_0 platí.

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

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

Podmínka dobré aproximace: $n_j M_* > 5$ pro všechna $j = 1, \ldots, r$.

Brandtův – Snedecorův výpočetní tvar: $Q = \frac{1}{M_*(1-M_*)}\sum_{j=1}^{r} n_j M_j^2 - n\frac{M_*}{1-M_*}$.

Test homogenity založený na arkussinusové transformaci

Není-li splněna podmínka $n_j M_* > 5$ pro všechna j = 1, ..., r, doporučuje se následující postup: označme

$A_j = \arcsin\sqrt{M_j}$, j = 1, ..., r,

$B = \frac{1}{n}\sum_{j=1}^{r} n_j A_j$.

Pak statistika $Q = 4\sum_{j=1}^{r} n_j \left(A_j - B\right)^2 \approx \chi^2(r-1)$.

H_0 tedy zamítáme na asymptotické hladině významnosti α, když $Q \geq \chi^2_{1-\alpha}(r-1)$.

Mnohonásobné porovnávání

Zamítneme-li nulovou hypotézu na asymptotické hladině významnosti α, chceme zjistit, které dvojice parametrů ϑ_k, ϑ_l se liší. Platí-li nerovnost $|A_k - A_l| \geq \sqrt{\frac{1}{8}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}\, q_{1-\alpha}(r, \infty)$, pak na hladině významnosti α zamítáme hypotézu o shodě parametrů ϑ_k, ϑ_l. (Hodnoty $q_{1-\alpha}(r, \infty)$ najdeme v tabulkách.)

Příklad: Na gymnázium bylo přijato 142 studentů. Ti byli náhodně rozděleni do čtyř tříd A, B, C, D. V každé třídě byla matematika vyučována jinou metodou. Na konci školního roku psali všichni studenti stejnou písemnou práci a byl zaznamenán počet těch studentů, kteří vyřešili všechny zadané úkoly.

Třída	A	B	C	D
Počet studentů	35	36	37	34
Počet úspěšných studentů	5	8	17	15

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že rozdíly mezi třídami jsou způsobeny pouze náhodnými vlivy.

Řešení:

Máme čtyři nezávislé náhodné výběry, j-tý pochází z rozložení A(ϑ_j), j = 1, 2, 3, 4.

Testujeme hypotézu H$_0$: $\vartheta_1 = \vartheta_2 = \vartheta_3 = \vartheta_4$.

$n_1 = 35$, $n_2 = 36$, $n_3 = 37$, $n_4 = 34$, $n = 142$

$m_1 = 5/35$, $m_2 = 8/36$, $m_3 = 17/37$, $m_4 = 15/34$, $m_* = (5+8+17+15)/142 = 45/142$.

Podmínky dobré aproximace:

$$35 \cdot \frac{45}{142} = 11,09,\ 36 \cdot \frac{45}{142} = 11,41,\ 37 \cdot \frac{45}{142} = 11,73,\ 34 \cdot \frac{45}{142} = 10,77$$

Testová statistika

$$Q = \frac{1}{m_*(1-m_*)}\left(\sum_{j=1}^{r} n_j m_j^2 - \frac{M_*^2}{n}\right) = \frac{1}{\frac{45}{142}\left(1-\frac{45}{142}\right)}\left[35\cdot\left(\frac{5}{35}\right)^2 + 36\cdot\left(\frac{8}{36}\right)^2 + 37\cdot\left(\frac{17}{37}\right)^2 + 34\cdot\left(\frac{15}{34}\right)^2 - 142\cdot\left(\frac{45}{142}\right)^2\right] = 12,28$$

Kritický obor: $W = \langle \chi^2_{0,95}(3), \infty) = \langle 7,81, \infty)$.

Protože testové kritérium se realizuje v kritickém oboru, H$_0$ zamítáme na asymptotické hladině významnosti 0,05.

Nyní metodou mnohonásobného porovnávání zjistíme, které dvojice parametrů se od sebe liší na hladině významnosti 0,05.

Pomocí arkussinusové transformace vypočteme hodnoty $A_j = \arcsin\sqrt{M_j}$:

$A_1 = 0{,}3876$, $A_2 = 0{,}4909$, $A_3 = 0{,}7448$, $A_4 = 0{,}7264$

Platí-li nerovnost $|A_k - A_l| \geq \sqrt{\frac{1}{8}\left(\frac{1}{n_k}+\frac{1}{n_l}\right)}\, q_{1-\alpha}(r,\infty)$, pak na hladině významnosti α zamítáme hypotézu o shodě parametrů

ϑ_k, ϑ_l.

Kvantil studentizovaného rozpětí najdeme v tabulkách: $q_{0,95}(4,\infty) = 3{,}63$

Srovnávané třídy	Rozdíly $\lvert A_k - A_l \rvert$	Pravá strana vzorce
A, B	0,1033	0,30
A, C	0,3572	0,30
A, D	0,3388	0,31
B, C	0,2539	0,30
B, D	0,2356	0,31
C, D	0,0184	0,30

Na hladině významnosti 0,05 se liší třídy A, C a A, D.

Řešení pomocí systému STATISTICA

Vytvoříme nový datový soubor se dvěma proměnnými a 142 případy.
Proměnná USPECH obsahuje hodnotu 1, pokud student vyřešil všechny zadané úkoly, jinak obsahuje hodnou 0.
Proměnná TRIDA má hodnotu 1, pokud student pochází z třídy A, hodnotu 2 pro třídu B, hodnotu 3 pro třídu C a hodnotu 4 pro třídu D.
Nejprve zjistíme podíly úspěšných studentů v jednotlivých třídách.
Statistiky – Základní statistiky a tabulky – Rozklad – OK – Proměnné – Závislé – USPECH, Grupovací - TRIDA – OK – Skupiny tabulek - odškrtneme Směrovat. odchylka - Výpočet.

TRID	USPE Průmě	USPE N
A	0,142	3
B	0,222	3
C	0,459	3
D	0,441	3
Vš.sk	0,316	14

Vidíme, že nejslabší výkony podávali studenti ze třídy A, úspěšných bylo pouze 14,3% studentů, ve třídě B 22,2%, ve třídě C 45,9% a ve třídě D 44,1%. Třídy C a D se z hlediska úspěchu v písemce z matematiky liší jen nepatrně

Dále provedeme testování hypotézy o shodě parametrů čtyř alternativních rozložení. Nejprve ověříme splnění podmínek dobré aproximace: $n_j m_* > 5$ pro všechna j = 1, ..., r. Vážený průměr m_* se nachází v posledním řádku výstupní tabulky procedury Rozklad. Jeho hodnotu okopírujeme do políček pro průměry tříd A, B, C, D, poslední řádek odstraníme a k tabulce přidáme jednu novou proměnnou, do jejíhož Dlouhého jména napíšeme =v2*v3.

TRID	USPE(Průmě	USPE N	NProi =v2*v
A	0,316	3:	11,09
B	0,316	3(	11,40
C	0,316	3	11,72
D	0,316	3·	10,77

Vidíme, že podmínky dobré aproximace jsou splněny.

Statistiky – Základní statistiky/tabulky – Kontingenční tabulky - OK - Specif. tabulky – List 1 USPECH, List 2 TRIDA, OK– Možnosti – Statistiky dvourozměrných tabulek - zaškrtněte Pearson & M-L Chi –square – Detailní výsledky - Detailní 2-rozm. tabulky.

Statist.	Chi-kva	sv	p
Pearsonův c	12,28	df=	p=,00
M-V chi-kva	12,80	df=	p=,00

Testová statistika Q se realizuje hodnotou 12,2876, počet stupňů volnosti je 3, odpovídající p-hodnota = 0,00646, tedy na asymptotické hladině významnosti 0,05 hypotézu H_0 zamítáme. S rizikem omylu nejvýše 0,05 jsme tedy prokázali, že rozdíly v podílech úspěšných studentů v jednotlivých třídách nelze vysvětlit náhodnými vlivy.

Upozornění: Systém STATISTICA neumožňuje provedení metody mnohonásobného porovnávání pro náhodné výběry z alternativního rozložení. Pro orientaci lze použít Schefféhο metodu. V našem případě:

TRIDA	{1} M=,14	{2} M=,22	{3} M=,45	{4} M=,44
A		0,907	0,034	0,060
B	0,907		0,173	0,253
C	0,034	0,173		0,998
D	0,060	0,253	0,998	

Na asymptotické hladině významnosti 0,05 se liší třídy A a C.