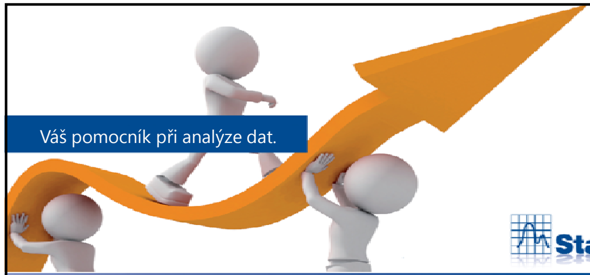


Váš pomocník při analýze dat.

StatSoft

Analýza síly testu ve *STATISTICA*

Mgr. Lenka Blažková

StatSoft

Modul Analýza síly testu

Power Analysis and Interval Estimation

- *Analýza síly testu*
- *Odhad velikosti vzorku*
- *Pokročilé techniky pro odhad intervalu spolehlivosti*
- *Rozdělení pravděpodobnosti*

StatSoft

STATISTICA: Analýza síly testu

- Pro danou velikost výběru, pravděpodobnost α a typické parametry konkrétního testu spočte **sílu testu**.
- Vytvoří **grafy závislosti síly testu** na
  a)Velikosti výběru
  b)Velikosti efektu
  c)Hodnotě *α*

- Jeden průměr, t-test
- Dva průměry, t-test, nezáv. vzorky
- Dva průměry, t-test, závislé vzorky
- Několik průměrů, Plán. kontrast
- Několik průměrů, 1-faktor. ANOVA
- Několik průměrů, 2-faktor. ANOVA
- Jeden rozptyl, Chí-kvadrát test
- Dva rozptyly, F-test
- Jedna korelace, t-test
- Dvě korelace, Z-test
- Koeficient determinace
- Jeden podíl, Z, Chí-kvadrát test
- Dva podíly, Z-test
- Dva podíly, párové vzorky
- Přežívání - Ln-pořadový test
- Přežívání - Exponenciál., "Accrual"
- Přežívání - Exp., Accrual/Dropouty
- Modelování strukturálních rovnic

StatSoft

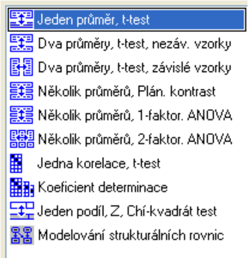
STATISTICA: Odhad velikosti vzorku

- Pro požadovanou sílu testu, pravděpodobnost α a typické parametry konkrétního testu spočte **minimální rozsah výběru**.
- Vytvoří grafy **závislosti velikosti vzorku** na
  a)Minimální požadované hodnotě síly testu
  b)Velikosti efektu
  c)Hodnotě *α*

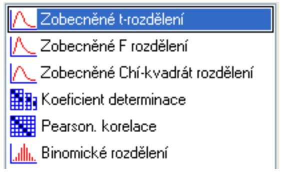
- Jeden průměr, t-test
- Dva průměry, t-test, nezáv. vzorky
- Dva průměry, t-test, závislé vzorky
- Několik průměrů, Plán. kontrast
- Několik průměrů, 1-faktor. ANOVA
- Několik průměrů, 2-faktor. ANOVA
- Jeden rozptyl, Chí-kvadrát test
- Dva rozptyly, F-test
- Jedna korelace, t-test
- Dvě korelace, Z-test
- Koeficient determinace
- Jeden podíl, Z, Chí-kvadrát test
- Dva podíly, Z-test
- Dva podíly, párové vzorky
- Přežívání - Ln-pořadový test
- Přežívání - Exponenciál., "Accrual"
- Přežívání - Exp., Accrual/Dropouty
- Modelování strukturálních rovnic

STATISTICA: Odhad intervalu spolehlivosti

Pro daný rozsah výběru, zvolenou spolehlivost 1-α a hodnotu příslušné statistiky získané na základě výběru, vypočte **(1-α)% interval spolehlivosti** pro odhadovaný parametr.



STATISTICA: Rozdělení pravděpodobnosti



Doplňuje pravděpodobnostní kalkulátor o **necentrální rozdělení t**, **F** a **chí-kvadrát**, a také o **binomické rozdělení**.

Umožňuje výpočet teoretických hodnot **korelačního koeficientu** a **koeficientu determinace** pro zadaný rozsah výběru a pozorovanou hodnotu parametru a zvolenou pravděpodobnost.

Kde hledat informace?

Jacob Cohen

Statistical power analysis for the behavioral sciences

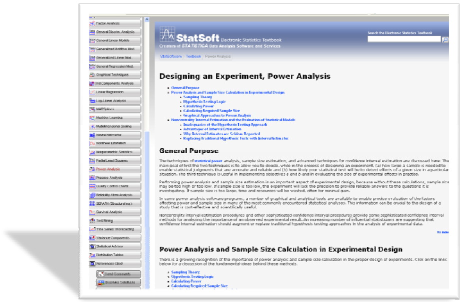
P. D. Ellis

The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Result

Christopher L. Aberson

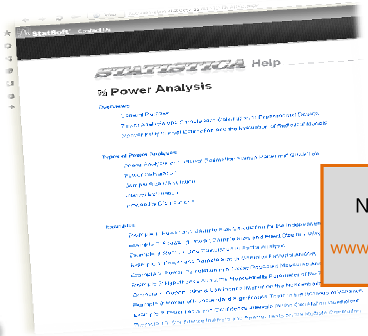
Applied Power Analysis for the Behavioral Sciences

Kde hledat informace?



Učebnice *STATISTICA*
www.statsoft.com/textbook/power/analysis/

Kde hledat informace?



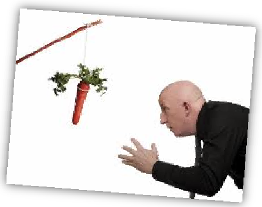
Nápověda *STATISTICA*

www.documentation.statsoft.com

Co nás dnes čeká?

- Úvod do analýzy síly testu
- T-testy
- ANOVA jednoduchého a dvojného třídění
- Pearsonův korelační koeficient
- Testy homogenity (χ^2, McNemarův)
- Koeficient determinace R^2
- Zobecněná rozdělení F, χ^2 a t

Motivace



Statistické šetření

Pro **konečnou populaci** (=soubor **jednotek**) chceme ověřit **pravdivost** nějaké **hypotézy** (výroku, tvrzení).

Šetření lze zpravidla podrobit jen část populace kvůli faktu, že

- populace je značně rozsáhlá,
- měření jsou časově a finančně náročná,
- jednotky nemají zájem na účasti ve výzkumu.

Náhodný výběr (Random sampling)

- Na základě analýzy vybraných jednotek, chceme vyslovit závěr pro celou populaci.
- Výběr by měl být
  - **reprezentativní**,
  - **náhodný** (stanovení výběrového plánu – existuje více typů náhodných výběrů).
  - **dostatečně rozsáhlý** pro dosažení dostatečné spolehlivosti odvozených závěrů.

Velikost vzorku (Sample size)

Příliš malý vzorek → Nedostatečná přesnost. Nespolehlivé závěry.

Příliš rozsáhlý vzorek → Plýtvání časem a dalšími zdroji výměnou za často pouze minimální zpřesnění.

Co je síla testu?

Testování hypotéz

Test: Pravidlo, které na základě výsledků zjištěných z náhodného výběru předepisuje rozhodnutí o zamítnutí nebo nezamítnutí nulové hypotézy týkající se celé populace z níž výběr pochází.

Závěry jsou platné vždy jen s určitou pravděpodobností.

Testová hypotéza

- **Nulová hypotéza H$_0$**:

Parametr θ je z množiny Θ_H.

Průměr je roven dvěma.

(Průměr je menší/větší nebo roven dvěma.)

- **Alternativní hypotéza** H$_1$:

Parametr θ je z množiny Θ_A.

  - **Oboustranná**

    Průměr není roven dvěma.

  - **Jednostranná**

    a) *Průměr je větší než dvě.*

    b) *Průměr je menší než dvě.*

Závěr: Zamítnutí či nezamítnutí testové hypotézy H$_0$ (ve prospěch alternativy) na základě testové statistiky spočtené z dat.

Chyba prvního druhu

Hypotézu zamítneme, ačkoli platí.

Pravděpodobnost chyby 1.druhu omezujeme hodnotou **α** (nejčastěji α =0.05), které říkáme **hladina významnosti testu (significance level)**.

***p*-hodnota (*p*-value)** – pravděpodobnost, že získáme stejné nebo extrémnější testové kritérium než je vypočítané, za předpokladu, že ve skutečnosti platí nulová hypotéza.

Chyba druhého druhu

Hypotézu nezamítneme, ačkoli neplatí.

Pravděpodobnost chyby 2.druhu označujeme **β**.

Síla testu (Power): Pravděpodobnost, že **správně zamítneme** hypotézu, která ve skutečnosti neplatí, **1-β**.

Minimalizovat chybu druhého druhu znamená **maximalizaci** síly testu.

Je-li k dispozici více různých statistických testů pro testování stanovené hypotézy se stejnou hladinou chyby prvního druhu, volíme ten z nich, který má největší sílu.

Vtah mezi chybami 1. a 2. druhu a silou testu

- Ve většině případů **snižování chyby jednoho druhu** vede ke **zvyšování chyby druhého druhu** a naopak.
- Vzájemný vztah je ovlivněn **velikostí výběru** a **velikostí efektu**.

Velikost výběru

Pravděpodobnost chyby I.druhu

Pravděpodobnost chyby II.druhu

Velikost efektu

Praktická versus statistická významnost

Test	Statisticky významný (dle *p*-hodnoty)	Statisticky nevýznamný (dle *p*-hodnoty)
Prakticky důležitý	**rozsah výběru OK**	rozsah výběru příliš **malý**
Prakticky nedůležitý	rozsah výběru příliš **velký**	**rozsah výběru OK**

→ **Apriorní analýza síly testu**

Faktory ovlivňující sílu testu

- Odchylka (**velikost efektu**, ES effect size)
  - **čím větší ES, tím je síla testu vyšší**.
- **Variabilita** (směrodatná odchylka) základního souboru.
  - **Čím je menší variabilita, tím je vyšší síla testu.** Variabilitu odhadujeme na základě náhodného výběru.
- **Rozsah *n* výběru.**
  - **Čím větší je rozsah souboru, tím vyšší je síla testu.**
- **Velikost chyby 1.druhu α**
  - **Čím je vyšší α, tím je nižší β** a tedy tím je **vyšší síla testu.**
- **Typ statistického testu**
  - Některé testy mají přirozeně větší sílu testu než jiné alternativní testy.

Určení **velikosti výběru**

Apriorní analýza (před provedením pokusu)	
Zjišťujeme	Známe (zadáváme)
Potřebnou velikost výběru n.	Hladinu významnosti testu pro chybu prvního druhu α. Požadovanou sílu testu 1-β. Velikost efektu, kterou potřebujeme detekovat.

Určení **síly testu**

Aposteriorní analýza (po provedení pokusu)	
Zjišťujeme	Známe (zadáváme)
Skutečnou sílu testu.	Hladinu významnosti testu pro chybu prvního druhu α. Velikost výběru n. Velikost efektu, kterou potřebujeme detekovat.

Síla testu

- Sílu testu **1-β** je třeba zkoumat pro všechny možné hodnoty parametru θ z množiny Θ_A.
- Jde vlastně o analýzu **silofunkce 1-β(θ)**.

Typy analýzy síly testu

- **Apriorní**

  Může zajistit, že neplýtváme časem a zdroji na výzkum, který má jen malou naději na prokázání signifikantního efektu a také zabrání zahrnutí zbytečně mnoha jednotek.
- **Post hoc**

  Pomáhá správně interpretovat výsledky testování, kde nevyšel průkazný efekt (nedošlo k zamítnutí nulové hypotézy).

Konkrétní aplikace

Test o proporci znaku v populaci

Ilustrační příklad: volební preference

Otázka: Má politická strana mezi voliči většinu?

Opora analýzy: anketa náhodného vzorku 100 lidí (odpověď ANO-NE).

π ... skutečné procento voličů p...odhad hodnoty π

Nulová hypotéza: $\pi \leq 0.5$

Alternativa: $\pi > 0.5$ (tzv. ***reject-support*** přístup)

Testové pravidlo: zamítá H_0 pro $p \geq 0.58$.

Ilustrační příklad: volební preference

- Počet voličů strany se řídí binomickým rozdělením

$$P(X=k)=\binom{n}{k}\pi^k(1-\pi)^{n-k}$$

- Pokud ve vzorku $n=100$ napočítáme více něž $k=58$ voličů strany, zamítáme nulovou hypotézu a prohlásíme, že strana má většinu hlasů.

Test hypotézy o parametru π alternativního rozdělení (normální aproximace)

$$H_0: \pi \leq \pi_0$$

$$H_1: \pi > \pi_0$$

Testová statistika $U=\dfrac{p-\pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}}$

má za platnosti nulové hypotézy asymptoticky normální rozdělení N(0, 1).

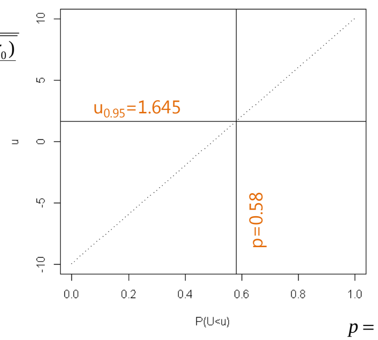
Kritický obor

$$W_\alpha=\{U \geq u_{1-\alpha}\}$$

10.12.2012 Analýza síly testu ve STATISTICA 31

Ilustrační příklad: volební preference

$$U=\frac{p-\pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}}$$



$$p=\min_{\pi\in(0,1)}\left(\frac{\pi-0.5}{0.05}\geq 1.645\right)$$

Situace A: π=0.5

- Ve skutečnosti *nulová hypotéza platí*. Pravděpodobnost, že ji zamítneme, je rovna hodnotě

$$\alpha^* = \sum_{k=59}^{100} \binom{100}{k} 0.5^k \, (1-0.5)^{100-k} = 0.044.$$

Chyba prvního druhu má tedy pravděpodobnost 0.044 nebo menší, vyhovuje podmínce 5 %.

Pozn. Hraniční hodnota pro testové kritérium p=0.58 je nejnižší možná, která zaručuje pravděpodobnost chyby prvního druhu menší než 5 %.

Situace B: π=0.55

- Ve skutečnosti *nulová hypotéza neplatí*.

Pravděpodobnost, že ji zamítneme, je rovna hodnotě

$$1-\beta = 1 - \sum_{k=1}^{58} \binom{100}{k} 0.55^k \, (1-0.55)^{100-k} = 0.241.$$

Síla testu je tedy velmi malá. – Naše, byť správná, domněnka, že politickou stranu preferuje většina voličů, bude potvrzena pouze ve 24.1 % případů analýz náhodného výběru.

Tento test je proto nevhodný pro zodpovězení otázky ze zadání.

Ilustrační příklad: volební preference

- Jaká je **vhodná velikost vzorku *N*** abychom dosáhli rozumně veliké síly testu při zachování nízké hladiny testu v této situaci?

STATISTICA předpokládá v podobných případech použití spíše χ^2 testu než přesného binomického testu.

Chí-kvadrát test dobré shody

$H_0 : \pi_i = \pi_{i0}, \quad i = 1, \dots, k$

$H_1 :$ non H_0

Testová statistika $G = \sum_{i=1}^{k} \frac{(n_i - n\pi_{i0})^2}{n\pi_{i0}}$

má za platnosti nulové hypotézy asymptoticky chí-kvadrát rozdělení s (k-1) stupni volnosti.

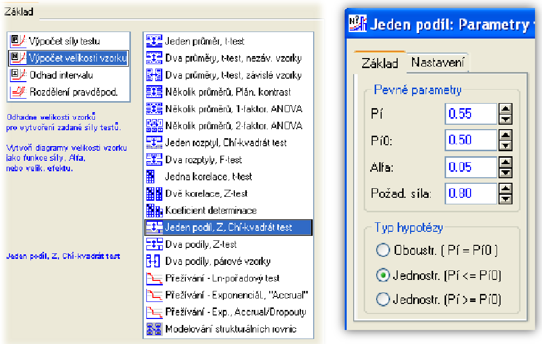
$n\pi_{i0} \geq 5$

Kritický obor

$$W_\alpha = \{G \geq \chi^2_{1-\alpha}\}$$

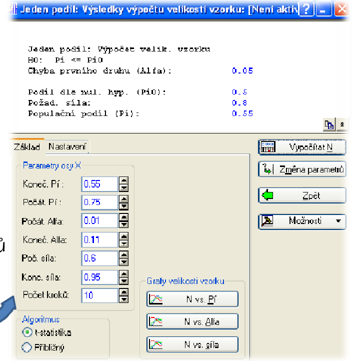
StatSoft

Ilustrační příklad: volební preference



StatSoft

Ilustrační příklad: volební preference



Volby se týkají grafických výstupů

StatSoft

Aproximace normálním rozdělením

- Pro velká N a hodnoty α, které nejsou blízké 0 nebo 1, lze binomické rozdělení aproximovat normálním rozdělením

$$\mathrm{Bi}(N,\pi) \approx \mathrm{N}\left(\pi, \pi\,(1-\pi)/N\right).$$

- Volba ***t-statistika*** používá pro výpočet přesné hodnoty pravděpodobnosti zamítnutí H_0 binomické rozdělení.
- Volba ***Přibližný*** používá normální aproximaci binomického rozdělení a počítá přibližnou hodnotu pravděpodobnosti zamítnutí H_0.

StatSoft

Aproximace normálním rozdělením

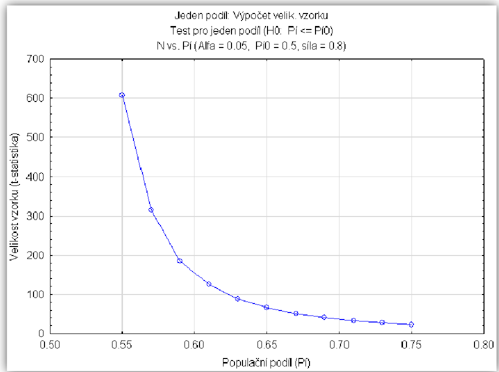
Testová statistika pro testování hypotézy H_0: $\pi=\pi_0$ je rovna

$$Z = \frac{p-\pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{N}}}.$$

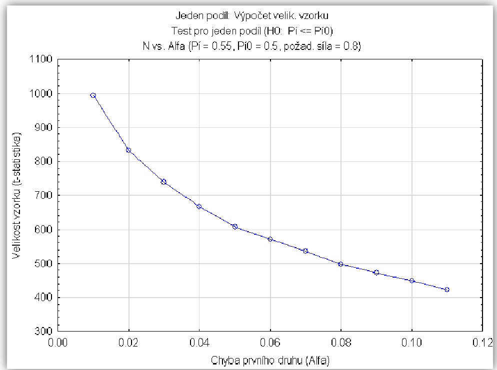
U malých vzorků se uplatňuje oprava na spojitost:

$$Z_C = \frac{(p+C)-\pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{N}}} \qquad C = \begin{cases} \dfrac{0.5}{N} & \text{pro } p \ge \pi_0 \\ -\dfrac{0.5}{N} & \text{pro } p < \pi_0 \end{cases}$$

Ilustrační příklad: volební preference

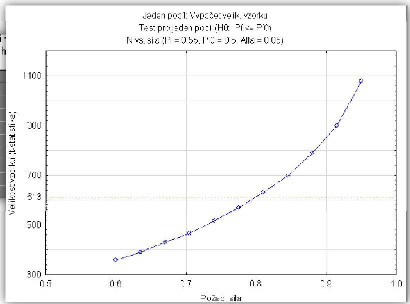


Ilustrační příklad: volební preference



Ilustrační příklad: volební preference

	Výpočet velikosti Jeden podíl, Z, Ch H0: Pí <= Pí0
	Hodnota
Podíl dle nul. hyp. (Pí0)	0.5000
Populační podíl (Pí)	0.5500
Alfa (nominál.)	0.0500
Aktuál. Alfa (přesně)	0.0522
Požad. síla	0.8000
Aktuální síla (normální aprox.)	0.7947
Aktuální síla (přesně)	0.8009
Požad. velikost vzorku (N)	607.0000



Vypočtená **velikost vzorku je 607**, dosažená **síla testu** bude přesně **0.8009** a pravděpodobnost chyby prvního druhu bude **0.0522**. Při této velikosti vzorku jsme schopni detekovat efekt 0.55.

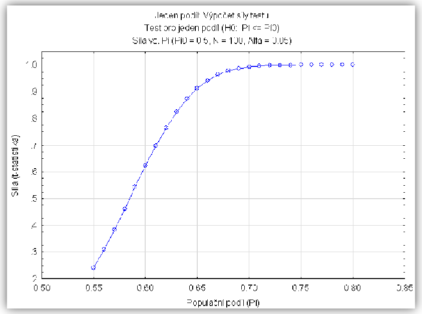
Ilustrační příklad: volební preference

Stanovili jsme rozsah výběru pro detekci efektu 0.05, **je tento rozsah výběru dostatečný i pro efekt jiné velikosti?**

V praxi neprovádíme výpočet síly testu pouze pro jedinou hodnotu efektu, ale zabýváme se rozsahem výběru pro různě velké efekty…

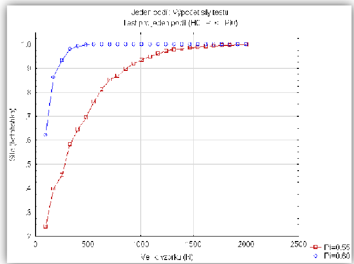
Ilustrační příklad: volební preference

- **Síla testu** v závislosti na **velikosti efektu** při pevném rozsahu výběru.



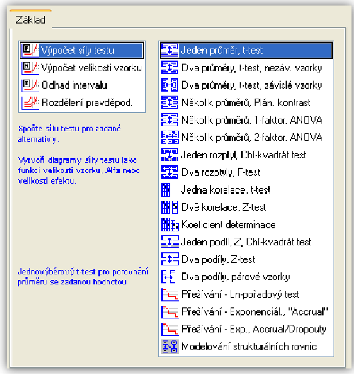
Ilustrační příklad: volební preference

- **Síla testu** v závislosti na **rozsahu výběru** pro několik velikostí efektu.



T-testy

Jednovýběrový t-test



Test hypotézy o střední hodnotě

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

Testová statistika $T = \dfrac{\bar{x} - \mu_0}{\frac{s_n}{\sqrt{n}}}$

má za platnosti nulové hypotézy Studentovo t-rozdělení s (n-1) stupni volnosti.

Kritický obor

$W_\alpha = \{|T| \geq t_{1-\alpha/2}\}$

Obecně má testová statistika **necentrální t rozdělení s (n-1) stupni volnosti** a parametrem **necentrality $\delta = n^{1/2} E_S$**, kde

$$E_S = \frac{\mu - \mu_0}{\sigma}.$$

Index BMI

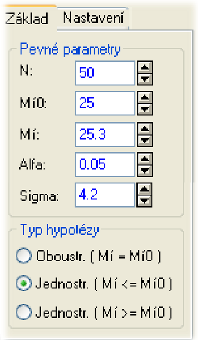
- BMI je podíl váhy[kg] a druhé mocniny výšky[m].
- Hodnocení BMI:

20-25	optimum
26-30	mírná nadváha
31-35	obezita 1.stupně
36-40	obezita 2.stupně

Trpí v průměru populace nadváhou nebo obezitou?

Jaká je **síla testu**, je-li skutečná průměrná hodnota BMI v populaci rovna 25.3 a populační směrodatná odchylka BMI je 4.2?

BMI Index



$\mathrm{H}_0 : \bar{X} \leq 25$

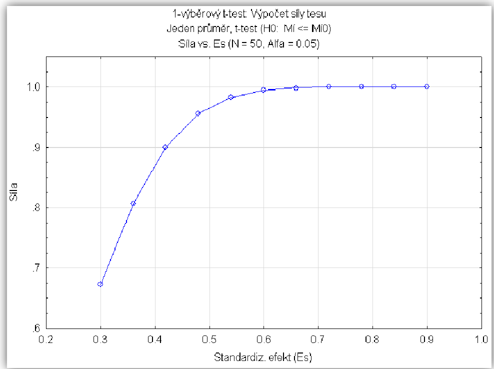
$\mathrm{H}_1 : \bar{X} > 25$

Vaha.sta

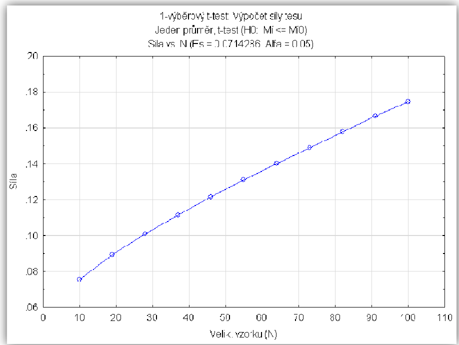
Výpočet síly testu ([Není aktivní tabulka])
Jeden průměr, t-test
H0: Mí <= Mí0

	Hodnota
Průměr dle nul. hypotézy (Mi0)	25.0000
Populační průměr (Mi)	25.3000
Populač. sm.odch. (sigma)	4.2000
Standardiz. efekt (Es)	0.0714
Velikost vz. ve skup. (N)	50.0000
Chyba prvního druhu (Alfa)	0.0500
Kritická hodnota t	1.6766
Síla	0.1257

BMI index - síla testu

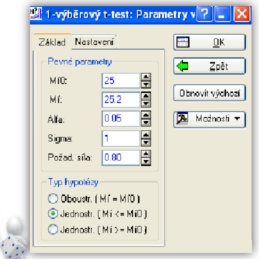


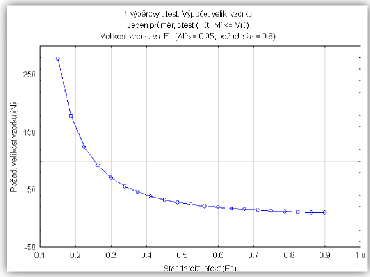
BMI index – síla testu



BMI index – velikost výběru

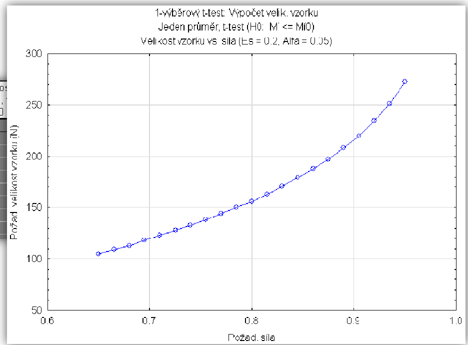
- Jak velký výběr zaručí sílu testu 0.8 i pro malý efekt? (E_s=0.2)



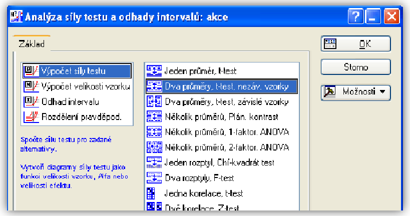


BMI – optimální rozsah výběru

	Výpočet velikos… Jeden průměr, … H0: Mi <= Mi0
	Hodnota
Průměr dle nul. hypotézy (Mi0)	25.0000
Populační průměr (Mi)	25.2000
Populač. sm.odch. (sigma)	1.0000
Standardiz. efekt (Es)	0.2000
Chyba prvního druhu (Alfa)	0.0500
Požad. síla	0.8000
Aktuální síla pro požad. N	0.8002
Požad. velikost vzorku (N)	156.0000



Dvouvýběrový t-test



Test hypotézy o **shodě středních hodnot** dvou **nezávislých** výběrů

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Testová statistika $T = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

má za platnosti nulové hypotézy Studentovo t-rozdělení s (n_1+n_2-2) stupni volnosti.

Kritický obor

$$W_\alpha = \{ |T| \geq t_{1-\alpha/2} \}$$

Obecně má testová statistika **necentrální t rozdělení s (n-2) stupni volnosti** a parametrem **necentrality** $\boldsymbol{\delta = (n_1 n_2/(n_1+n_2))^{1/2} E_S}$, kde $E_S = \dfrac{\mu_1 - \mu_2}{\sigma}$.

Příklad: t-test, nezávislé vzorky

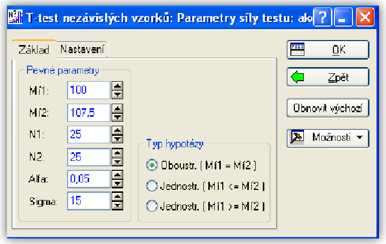
Plánované rozsahy skupin jsou 25 pro každý výběr.

Populační směrodatná odchylka u obou skupin nechť je **15**.

Nechť skupina 2 je kontrolní, a lze předpokládat, že **populační průměr sledované charakteristiky je roven 100**.

Populační průměr u skupiny 1 je předmětem experimentu, nicméně z hlediska prováděného experimentu nebude provedené ošetření u této skupiny shledáno efektivním, pokud nezvýší populační průměr alespoň na hodnotu **107.5**.

Příklad: t-test, nezávislé vzorky



Mí1, Mí2 – populační průměry pro první a druhou skupinu
N1, N2 – rozsahy vzorků pro první a druhou skupinu
Sigma – populační směrodatná odchylka (shodné rozptyly skupin)

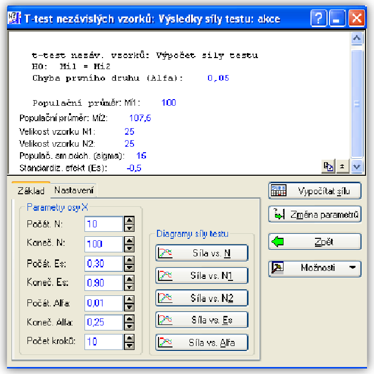
Příklad: t-test, nezávislé vzorky

V praxi většinou **neznáme populační průměry skupin** ani hodnotu jejich společné **populační směrodatné odchylky**.

Velikost síly testu závisí na tzv. velikosti efektu (**effect size**)

$$E_s = \frac{\mu_1 - \mu_2}{\sigma}$$

Informace o konkrétních hodnotách populačních charakteristik lze převést na velikost tohoto efektu.



Velikost efektu

$$E_s = \frac{\mu_1 - \mu_2}{\sigma}$$

V literatuře (např. *Cohen, 1983: **Statistical Power Analysis for the Behavioral Sciences***) jsou pro tento test doporučována následující pravidla:

1. Malý efekt $E_s = 0.2$
2. Středně velký efekt $E_s = 0.5$
3. Velký efekt $E_s = 0.8$

Příklad: t-test, nezávislé vzorky

$$E_s = \frac{\mu_1 - \mu_2}{\sigma}$$

Při zachování N1, N2 a α je síla testu stejná, pokud se nezmění velikost efektu.

	Výpočet síly testu (akce) Dva průměry, t-test, nezáv. H0: Mí1 = Mí2
	Hodnota
Populační průměr: Mí1	100,0000
Populační průměr: Mí2	107,5000
Populač. sm.odch. (sigma)	15,0000
Standardiz. efekt (Es)	-0,5000
Velikost vzorku N1	25,0000
Velikost vzorku N2	25,0000
Chyba prvního druhu (Alfa)	0,0500
Kritická hodnota t	2,0106
Síla	0,4101

	Výpočet síly testu (akce) Dva průměry, t-test, nezáv. H0: Mí1 = Mí2
	Hodnota
Populační průměr: Mí1	40,0000
Populační průměr: Mí2	50,0000
Populač. sm.odch. (sigma)	20,0000
Standardiz. efekt (Es)	-0,5000
Velikost vzorku N1	25,0000
Velikost vzorku N2	25,0000
Chyba prvního druhu (Alfa)	0,0500
Kritická hodnota t	2,0106
Síla	0,4101

Příklad: t-test, nezávislé vzorky

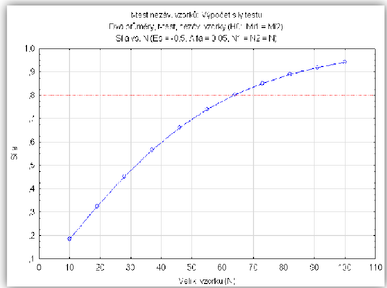
- V našem příkladě je dosažená síla rovna 0.8.

  Obvyklá nejmenší akceptovaná hodnota je kolem 0.8.

- Pro detekci středně velkého efektu na hladině α=5 % je tedy velikost skupin 25 nedostatečná.

Grafická analýza velikosti síly testu

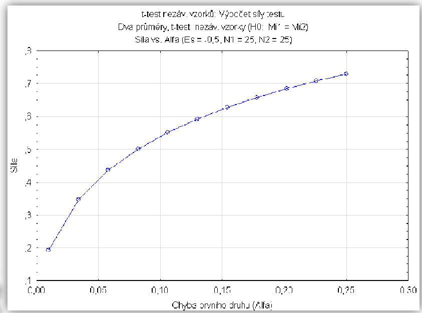
- Diagramy síly testu: **Síla vs. N**



Minimální potřebná velikost skupin pro dosažení síly testu alespoň 0.8 je přibližně 64.

Grafická analýza velikosti síly testu

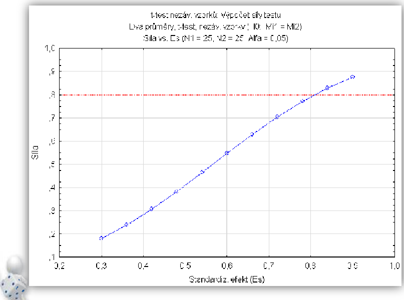
- Diagramy síly testu: **Síla vs. α**



Síla testu roste s rostoucí chybou prvního druhu. Malá změna α sílu zvýší jen minimálně, navíc již používáme největší přípustnou hodnotu pro chybu prvního druhu.

Grafická analýza velikosti síly testu

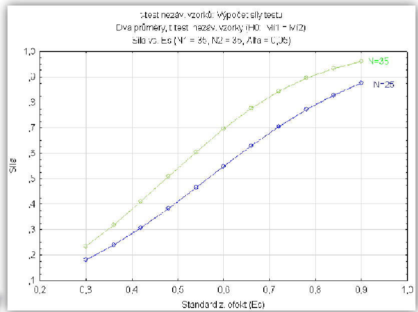
- Diagramy síly testu: **Síla vs. Es** (středně velký efekt)



Minimální potřebná velikost skupin pro dosažení síly testu alespoň 0.8 je přibližně 64.

Grafická analýza velikosti síly testu

- Diagramy síly testu: **Síla vs. Es** (středně velký efekt)

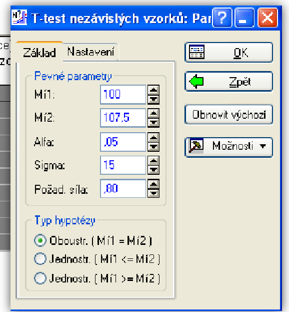


Vykreslíme i graf pro N1=N2=35.

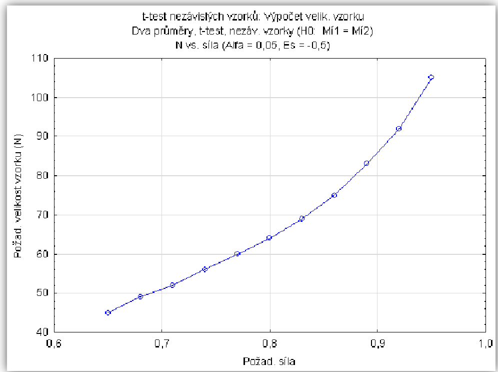
Zvýšením rozsahu skupin o 10 získáme test se silou vyšší o 0.10 až 0.15.

Výpočet velikosti vzorku

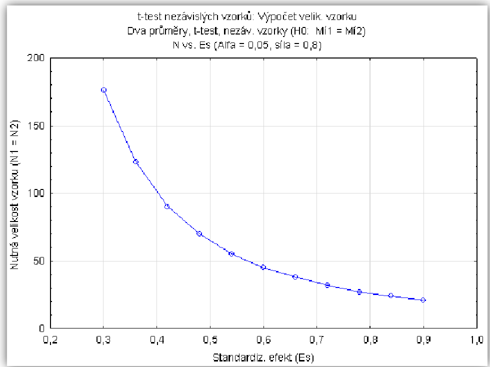
	Výpočet velikosti vzorku (akce… Dva průměry, t-test, nezáv. vz… H0: Mí1 = Mí2
	Hodnota
Populační průměr: Mí1	100,0000
Populační průměr: Mí2	107,5000
Populač. sm.odch. (sigma)	15,0000
Standardiz. efekt (Es)	-0,5000
Chyba prvního druhu (Alfa)	0,0500
Kritická hodnota t	1,9790
Požad. síla	0,8000
Aktuální síla pro požad. N	0,8015
Požadov. N (ve skup.)	64,0000



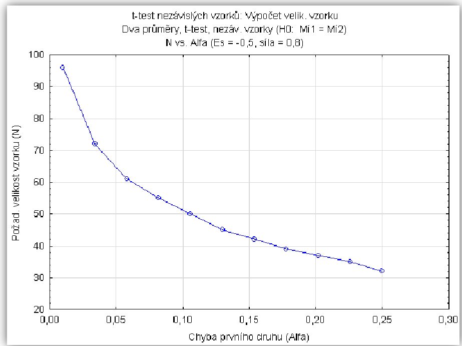
Výpočet velikosti vzorku



Výpočet velikosti vzorku



Výpočet velikosti vzorku



Jednofaktorová ANOVA

StatSoft

Jednofaktorová ANOVA

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k \qquad H_1: \exists\, i, j: \mu_i \neq \mu_j$$

Porovnáváme průměrnou úroveň spojité veličiny u ***k*** skupin. Celkovou variabilitu zkoumané proměnné rozdělíme na **meziskupinovou** a **vnitroskupinovou**.

$$s_T^2 = s_A^2 + s_e^2$$

Testová statistika

$$F = \frac{n-k}{k-1}\frac{s_A^2}{s_e^2}$$

má za platnosti nulové hypotézy **centrální F rozdělení s *k*-1 a *n*-*k* stupni volnosti**.

Za platnosti alternativní hypotézy, má testová statistika **necentrální F rozdělení s parametrem necentrality δ**.

StatSoft

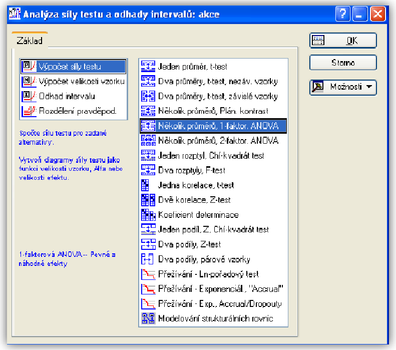
Jednofaktorová ANOVA – síla testu

- **Síla testu:** pravděpodobnost oprávněného zamítnutí nulové hypotézy

$$P(F \geq F_{1-\alpha}(k-1, n-k)) = 1 - F_\delta(F_{1-\alpha}(k-1, n-k)).$$

StatSoft

Jednofaktorová ANOVA



StatSoft

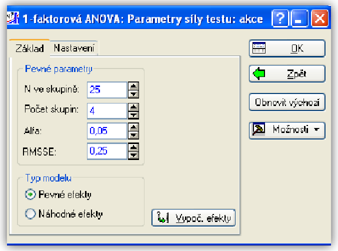
Jednofaktorová ANOVA

Zajímá nás efekt nového léku, který je vylepšenou verzí léku testovaného před dvěma lety.

Jakou sílu testu požadovat?

Jaká velikost vzorku je potřeba pro dosažení této síly?

Jednofaktorová ANOVA



Pevné efekty – porovnáváme ošetření, které jsme pozorovali v experimentu.

Náhodné efekty – konkrétní ošetření dosažená v experimentu (hodnoty faktorové proměnné) jsou náhodným výběrem z nějaké větší množiny možných hodnot.

RMSSE (*Root Mean Square Standardized Effect*)

$$Y_i = \mu + \alpha_j + e_{ij}$$

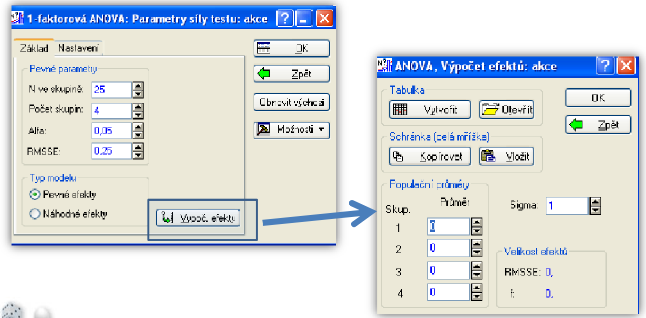
$$RMSSE = \sqrt{\frac{\sum_{j=1}^{J}\left(\frac{\alpha_j}{\sigma}\right)^2}{J-1}} \qquad f = \sqrt{\frac{\sum_{j=1}^{J}\left(\frac{\alpha_j}{\sigma}\right)^2}{J}}$$

Jednofaktorová ANOVA

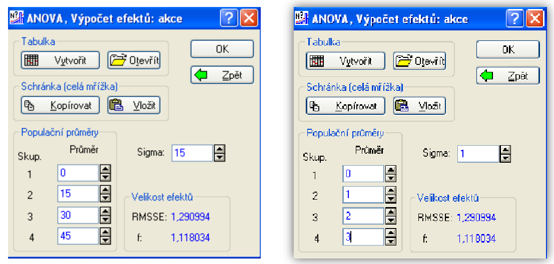
RMSSE – nemění se, pokud ke všem skupinovým průměrům přičteme stejnou konstantu.

Nicméně je těžké určit hodnoty efektu představující malý, střední a velký rozdíl porovnávaných populačních průměrů.

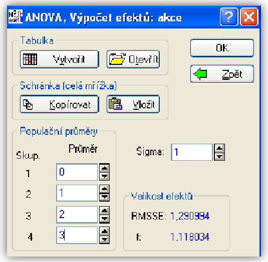
Jednofaktorová ANOVA

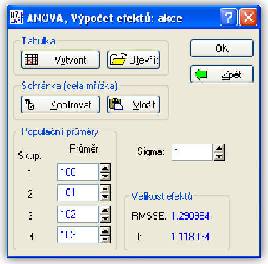


Jednofaktorová ANOVA



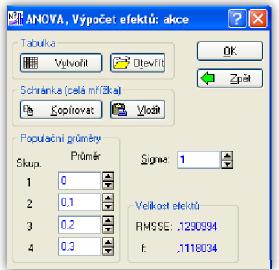
Jednofaktorová ANOVA





RMSSE a f jsou invariantní vůči lineární transformaci skupinových průměrů.

Jednofaktorová ANOVA



Průměrný efekt je v obou případech 0.2, ale RMSSE se podstatně liší.

Jednofaktorová ANOVA

- Cohen – doporučení

  f=0.1 **malý efekt**

  f=0.25 **středně velký efekt**

  f=0.40 **velký efekt**

$$RMSSE = \sqrt{\frac{J}{J-1}}\, f$$

Pro 4 sk. je odpovídající RMSSE rovno 0,2886.

  - Pouze hrubé vodítko. Je třeba zkoumat nějaký širší rozsah hodnot RMSSE, resp. f.

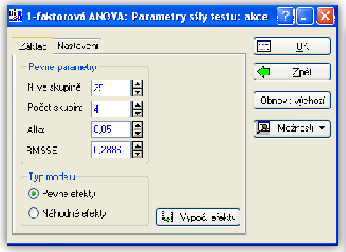
- StatSoft – doporučení pro RMSSE

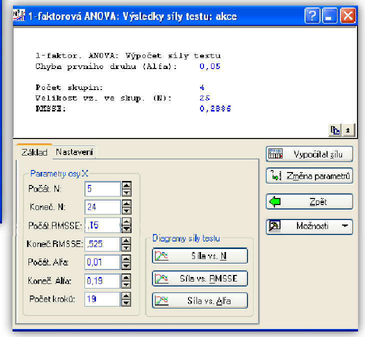
  RMSSE=0.15 **malý efekt**

  RMSSE=0.3 **středně velký efekt**

  RMSSE=0.5 **velký efekt**

Výpočet síly testu



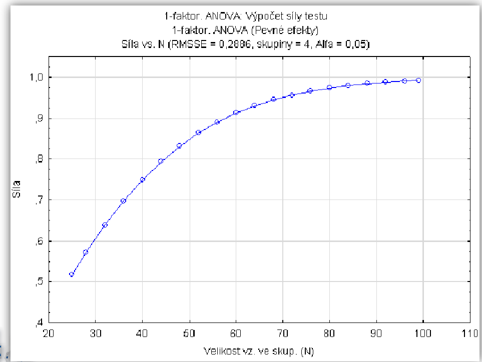


Výpočet síly testu

	Výpočet síly testu (akce) ANOVA, 1-faktorová Pevné efekty
	Hodnota
Počet skupin	4,0000
Velikost vz. ve skup. (N)	25,0000
RMSSE	0,2886
Parametr zobec. t-rozdělení (Delta)	6,2467
Chyba prvního druhu (Alfa)	0,0500
SV efektu	3,0000
Chyba SV	96,0000
Kritická hodnota F	2,6994
Síla	0,5179

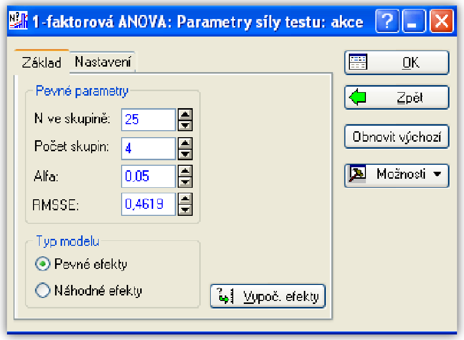
Síla testu je příliš malá, zvětšíme počáteční N na 25 a konečné na 100.

Grafická analýza síly testu ANOVA

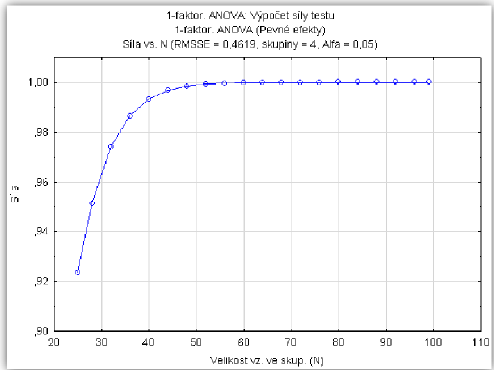


Nejstrmější nárůst síly testu pro N mezi 25 a 50.

Velký efekt



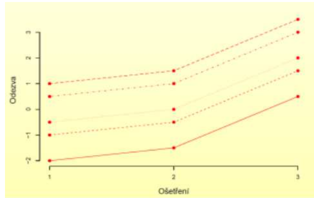
Velký efekt



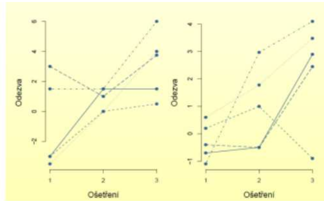
Dvoufaktorová ANOVA

Model s interakcí

Odezva se pro hodnoty faktorů liší jen posunutím.



Odezva se pro hodnoty faktorů liší i jinak než posunutím



Model dvojného třídění

2 kvalitativní znaky: *k* kategorií prvního faktoru

l kategorií druhého faktoru

$$Y_{jgp} = \mu + \alpha_j + \beta_g + \lambda_{jg} + e_{jgp}$$

H_{0A}: $\alpha_1 = \alpha_2 = \ldots = \alpha_k$

H_{0B}: $\beta_1 = \beta_2 = \ldots = \beta_l$

H_{0AB}: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{kl}$

řádkové efekty sloupcové efekty interakce

$$s_T^2 = s_A^2 + s_B^2 + s_{AB}^2 + s_e^2$$

$$F_A = \frac{n-kl}{k-1}\frac{s_A^2}{s_e^2} \quad F_B = \frac{n-kl}{l-1}\frac{s_B^2}{s_e^2} \quad F_{AB} = \frac{n-kl}{(k-1)(l-1)}\frac{s_{AB}^2}{s_e^2}$$

Síla testu

- Předpoklad stejně velkých skupin o rozsahu ***N***.
- Další parametry:
  - počet kategorií obou faktorů
  - RMSSE pro řádky
  - RMSSE pro sloupce
  - RMSSE pro interakci

Steiger and Fouladi (1997)

Síla testu

– RMSSE pro řádky

$$RMSSE_{\alpha} = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{\alpha_i}{\sigma}\right)^2}{k-1}}$$

– RMSSE pro sloupce

$$RMSSE_{\beta} = \sqrt{\frac{\sum_{j=1}^{l}\left(\frac{\beta_j}{\sigma}\right)^2}{l-1}}$$

– RMSSE pro interakci

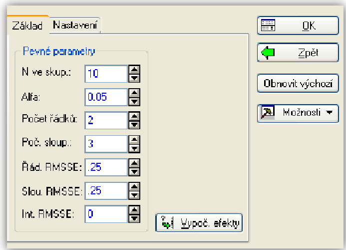
$$RMSSE_{\gamma} = \sqrt{\frac{\sum_{i=1}^{k}\sum_{j=1}^{l}\left(\frac{\gamma_{ij}}{\sigma}\right)^2}{(k-1)(l-1)}}$$

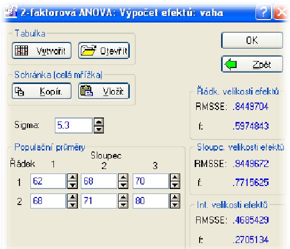
Parametr necentrality a RMSSE

F statistiky pro test řádkového, sloupcového efektu a efektu interakcí mají obecně **necentrální *F* rozdělení**. Parametr necentrality δ úzce souvisí s RMSSE, např. *F* statistika pro řádkový efekt má necentrální *F* rozdělení s ***k*-1** a ***kl*(*n*-1)**stupni volnosti a parametr necentrality je dán předpisem:

$$\delta_{\alpha} = nl\sum_{i=1}^{k}\left(\frac{\alpha_i}{\sigma}\right)^2 = nl(k-1)(RMSSE_{\alpha})^2$$

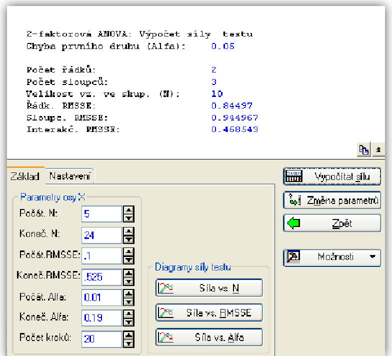
Síla testu





2-faktorová (2 X 3) ANOVA Cell Means (váha)	B1	B2	B3
A1	62.0000	68.0000	70.0000
A2	68.0000	71.0000	80.0000
Sigma (vnitř.)	5.3000		

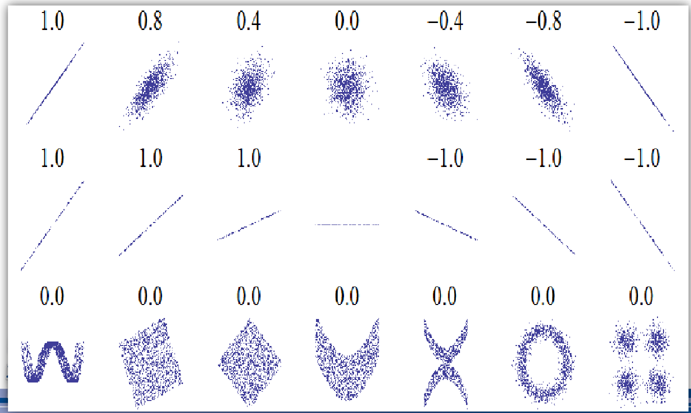
Síla testu



	Výpočet síly testu (vah ANOVA, 2-faktorová Hodnota
Počet řádků	2.0000
Počet sloupců	3.0000
Velikost vz. ve skup. (N)	10.0000
Chyba prvního druhu (Alfa)	0.0500
Řádk. efekt	
RMSSE	0.8450
SV efektu	1.0000
Chyba SV	54.0000
Kritická hodnota F	4.0195
Síla	0.9951
Sloupc. efekt	
RMSSE	0.9450
SV efektu	2.0000
Chyba SV	54.0000
Kritická hodnota F	3.1682
Síla	0.9998
Interakč. efekt	
RMSSE	0.4685
SV efektu	4.0000
Chyba SV	54.0000
Kritická hodnota F	3.1682
Síla	0.4294

Test nulovosti korelace

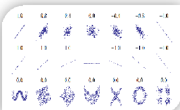
Pearsonův korelační koeficient



Pearsonův korelační koeficient

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var} X \operatorname{var} Y}} = \frac{\mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y)}{\sqrt{\operatorname{var} X \operatorname{var} Y}}$$

- 1	• Silná negativní závislost • Y = -k X
1	• Silná pozitivní závislost • Y = k X
0	• Mezi veličinami není ***lineární*** závislost



Pearsonův korelační koeficient

- Nezávisí na jednotkách

$$\rho_{X,Y} = \operatorname{sign}(ac)\rho_{aX+b,cY+d}$$

- **Výběrový korelační koeficient**

$$r_{X,Y} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}$$

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})$$

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

Test nulovosti korelace

H_0: ρ=0 proti H_1: |ρ|>0

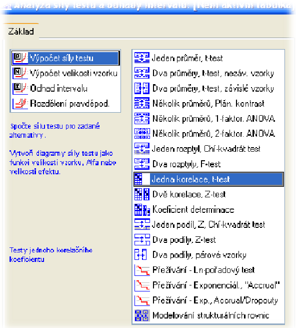
$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

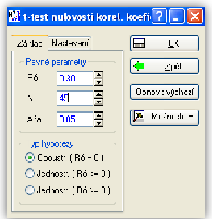
Jestliže je (**X**, **Y**) výběr z **dvourozměrného norm. rozdělení** a **ρ=0**, potom má testová statistika T ***t*-rozdělení** s ***n*-2** stupni volnosti.

Příklad

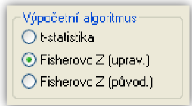
- Jaká je síla testu nulovosti korelace pro rozsah výběru 45, pokud skutečná korelace dosahuje hodnoty 0.30?
- V případě, že je síla testu příliš malá, jaká velikost výběru zaručí sílu 0.90?

Síla testu





Síla testu



t-statistika: přesný výpočet, pomalý, ale pro rozsahy vzorků typické pro výzkumné práce je dostatečný.

Fisherovo Z (uprav.): metoda užívající Fisherovu transformaci s upravenými vzorci pro výpočet průměru a rozptylu (Fouladi, 1991).

Fisher Z (původ.): používá tradiční aproximaci založenou na Fisherově transformaci. Předpokládá, že Fisherova aproximace *r* je v průměru rovna skutečné hodnotě *ρ* a rozptyl Fisherovy transformace r je roven $(N-3)^{1/2}$.

Fisherova Z-transformace

R. A. Fisher – postup pro testování nulovosti korelačního koeficientu

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

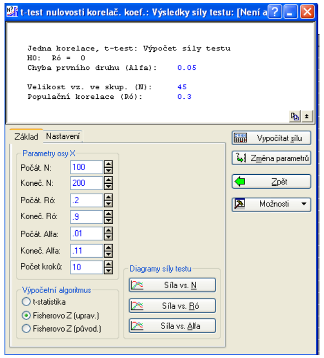
S rostoucím rozsahem výběru *n* se rozdělení náhodné veličiny Z blíží normálnímu rozdělení:

$$N\left(\frac{1}{2}\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{2(n-1)};\ \frac{1}{n-3}\right) \approx N\left(\frac{1}{2}\ln\frac{1+\rho}{1-\rho};\ \frac{1}{n-3}\right)$$

Za platnosti nulové hypotézy (nulovost korelačního koeficientu) má Z rozdělení N(0; 1/(n-3)) a veličina U má přibližně rozdělení N(0,1)

$$U = \frac{\sqrt{n-3}}{2}\ln\frac{1+r}{1-r}$$

Test významnosti korelace

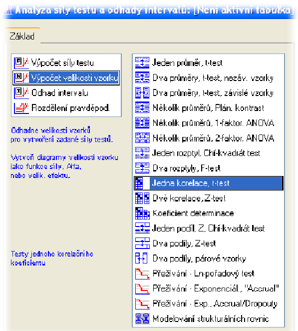


Síla testu

	Výpočet síly testu ([Není aktivní tabulka]) Jedna korelace, t-test H0: Ró = 0 Hodnota
Velikost vz. ve skup. (N)	45.0000
Korelace dle nulové hypotézy (Ró0)	0.0000
Populační korelace (Ró)	0.3000
Chyba prvního druhu (Alfa)	0.0500
Síla (Fisher. Z upravené)	**0.5275**

- Dosažená síla testu je **0.5275**, tato hodnota je nedostatečná.

Výpočet vhodného rozsahu výběru

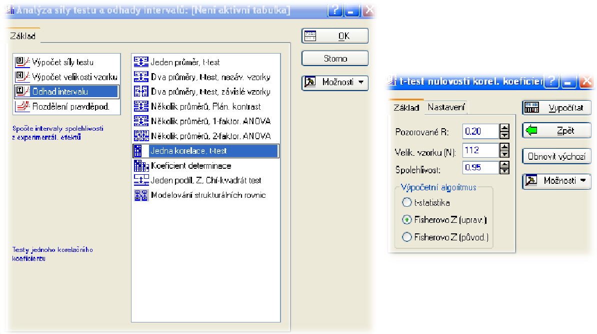


Výpočet vhodného rozsahu výběru

	Výpočet velikosti vzorku ([Není aktivní tabulka]) Jedna korelace, t-test H0: Ró = 0
	Hodnota
Populační korelace (Ró)	0.3000
Chyba prvního druhu (Alfa)	0.0500
Požad. síla	0.9000
Aktuální síla pro požad. N	0.9008
Požadov. N (Fisher. Z upravené)	112.0000

Vhodná velikost výběru je 112 místo 45.

Přesný interval spolehlivosti pro korelační koeficient



Přesný interval spolehlivosti pro korelační koeficient

	Odhad intervalu ([Není aktivní tabulka]) Jedna korelace, t-test
	Hodnota
Pozorovaný korel. koef. R	0.2000
Korelace dle nulové hypotézy (Ró0)	0.0000
Oboustranná p-hodnota	0.0343
Velikost vz. ve skup. (N)	112.0000
Interval spolehlivosti	0.9500
Meze spolehlivosti (Fisher. Z upravené):	
Ró:	
Dolní mez	0.0149
Horní mez	0.3703

Závěr:
- Za předpokladu, že skutečná hodnota korelace v populaci je 0.30 jsme stanovili rozsah výběru ***N*=112**, který **zaručí sílu testu 0.9008** (hladina testu α=5 %).
- Pokud je v tomto výběru hodnota **pozorovaného korelačního koeficientu 0.20**, jsme schopni kromě testu nulovosti korelace (***p*=0.0343**) stanovit i **95% interval spolehlivosti** pro odhad korelačního koeficientu, který nám poskytne přesnější informaci než uvedená *p*-hodnota.

Přesný interval spolehlivosti pro korelační koeficient

	Odhad inter Jedna korel t-test
	Hodnota
Pozorovaný korel. koef. R	0.2000
Korelace dle nulové hypotézy (Ró0)	0.0000
Oboustranná p-hodnota	0.0343
Velikost vz. ve skup. (N)	112.0000
Interval spolehlivosti	0.9500
Meze spolehlivosti (Fisher. Z upravené):	
Ró:	
Dolní mez	0.0149
Horní mez	0.3703

	Odhad inter Jedna korela t-test
	Hodnota
Pozorovaný korel. koef. R	0.2000
Korelace dle nulové hypotézy (Ró0)	0.0000
Oboustranná p-hodnota	0.1645
Velikost vz. ve skup. (N)	50.0000
Interval spolehlivosti	0.9500
Meze spolehlivosti (Fisher. Z upravené):	
Ró:	
Dolní mez	-0.0821
Horní mez	0.4491

Můžeme porovnat vhodnost rozsahu výběru i podle šířky získaného intervalu spolehlivosti…

Přesný interval spolehlivosti pro korelační koeficient

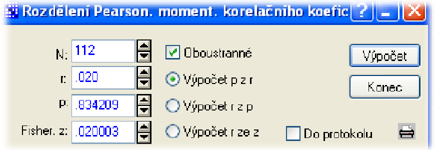
Použijeme-li **přesný** výpočet intervalu spolehlivosti, bude takto získaný interval spolehlivosti pro korelační koeficient obsahovat nulu právě tehdy, když (oboustranný) test nulovosti korelace zamítne nulovou hypotézu.

Pearsonův korelační koeficient

Interval spolehlivosti

asymptotický interval spolehlivosti $\left(\frac{C_L-1}{C_L+1}, \frac{C_U-1}{C_U+1}\right)$

$$C_{L,U} = \exp\left\{2Z \pm \frac{2u(\frac{\alpha}{2})}{\sqrt{n-3}}\right\}$$



Test pro konkrétní nenulovou hodnotu korelace

Příklad

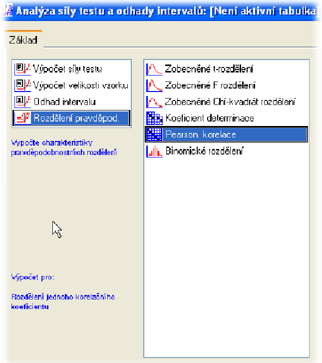
- Chceme testovat hypotézu, že populační korelační koeficient je roven 0.6 proti oboustranné alternativě.

  H_0: *ρ=a=0.60* proti H_1: |*ρ*-*a*|= |*ρ*-0.6| >0

- Pro experiment jsme použili výběr o rozsahu *N*=100 a hodnota výběrového korelačního koeficientu byla 0.721.
- Jaká je síla testu, je-li skutečná hodnota korelace v populaci 0.45?

StatSoft

Pravděpodobnostní kalkulátor *STATISTICA*



StatSoft

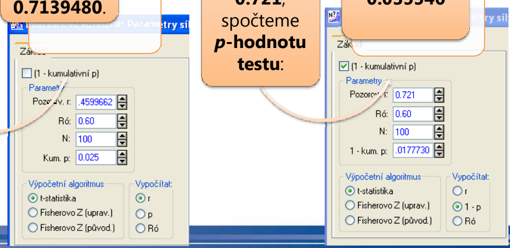
Pravděpodobnostní kalkulátor *STATISTICA*

Pro výpočet **kritických hodnot *r*** pro oboustranný test na hladině **α=0.05** spočteme ***r*** pro hodnoty **kum. p. 0.025** a **0.975**:

Kritické hodnoty jsou tedy **0.4599662** a **0.7139480**.

Pro oboustranný test, kdy pozorovaná hodnota korelace je **0.721**, spočteme ***p*-hodnotu testu**:

2*P(T>0.721) =2*0.017773= **0.035546**



StatSoft

Výpočet síly testu

- Předpokládáme, že skutečná hodnota korelace je 0.60, spočteme kritické hodnoty testu:

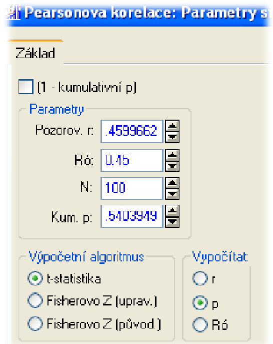
Kritické hodnoty jsou **0.4599662** a **0.7139480**.

- Ponecháme vždy kritickou hodnotu jako hodnotu pozorovaného ***r*** a hodnotu ***Ró*** nastavíme na **0.45**.

StatSoft

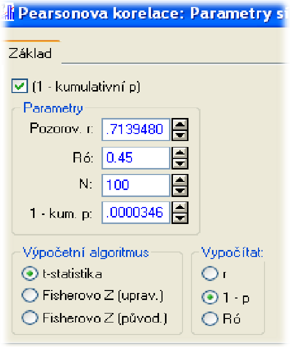
Výpočet síly testu

- Pravděpodobnost, že výběrová korelace bude nižší než dolní kritická hodnota, je za předpokladu ***ρ*=0.45** rovna **0.5403949**.



Výpočet síly testu

- Pravděpodobnost, že výběrová korelace bude za předpokladu ***ρ*=0.45** vyšší než horní kritická hodnota, je rovna **0.0000346**.



Výpočet síly testu

- Je-li hodnota populační korelace 0.60 a výběrová korelace dosahuje hodnoty 0.45, pak pravděpodobnost, že zamítneme nulovou hypotézu, která neplatí, (tedy **síla testu**) je rovna součtu obou získaných podmíněných pravděpodobností:

$$P(r < 0.4599662) + P(r > 0.7139480) =$$
$$= 0.5403949 + 0.0000346 =$$
$$= 0.5404295$$

Test pro porovnání dvou korelací

Příklad

- Zajímá nás síla testu, zda korelace mezi váhou a výškou je přibližně stejná ve čtvrté i osmé třídě ZŠ. Předpokládáme, že hodnoty populačních korelačních koeficientů jsou rovny $\boldsymbol{\rho_1=0.45}$, $\boldsymbol{\rho_2=0.18}$.
  Rozsahy výběrů, které máme k dispozici jsou $\boldsymbol{N_1=50}$, $\boldsymbol{N_2=60}$, výběrové korelace jsou $\boldsymbol{r_1=0.67}$, $\boldsymbol{r_2=0.30}$.

Testy rovnosti korelačních koeficientů

k nezávislých výběrů z dvojrozměrných normálních rozdělení

H_0: $\rho_1=\rho_2=\ldots=\rho_k$

$$Z_i = \frac{1}{2}\ln\frac{1+r_i}{1-r_i}$$

k=2:

Za platnosti H_0 má Z_1-Z_2 má přibližně N(0, 1/(n_1-3)+1/(n_2-3))

k>=3:

$$U = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}}$$

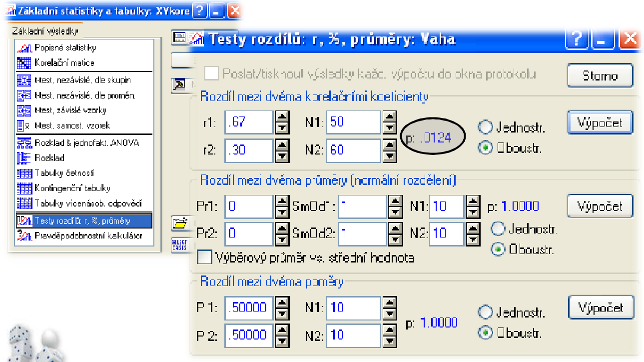
$$|U| \geq u\left(\frac{\alpha}{2}\right)$$

$$Q = \sum_{i=1}^{k}(n_i-3)(Z_i-b)^2$$

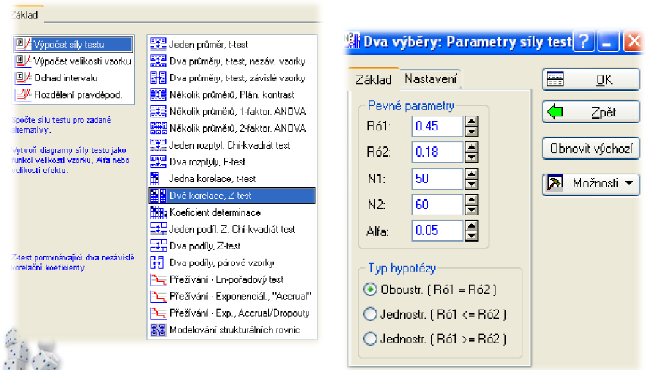
$$b = \frac{1}{n-3k}\sum_{i=1}^{k}(n_i-3)Z_i$$

$$Q \geq \chi^2_{k-1}(\alpha)$$

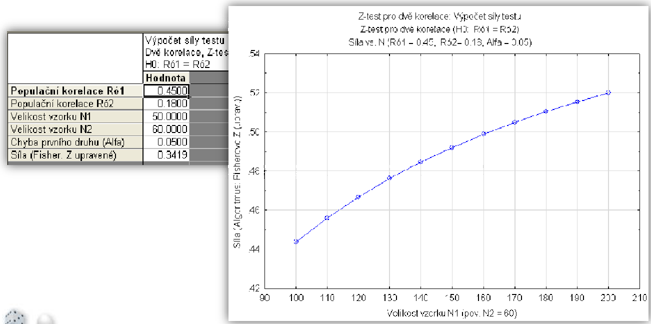
Výsledky pro oboustranný test



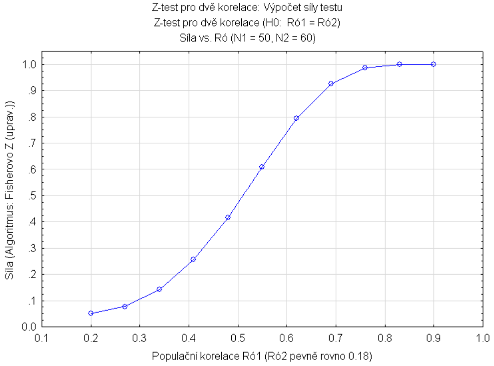
Síla testu



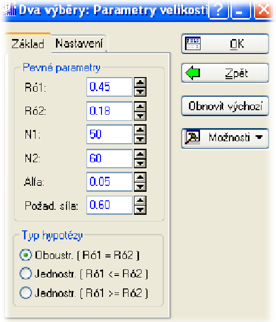
Síla testu



Síla testu v závislosti na ρ_1

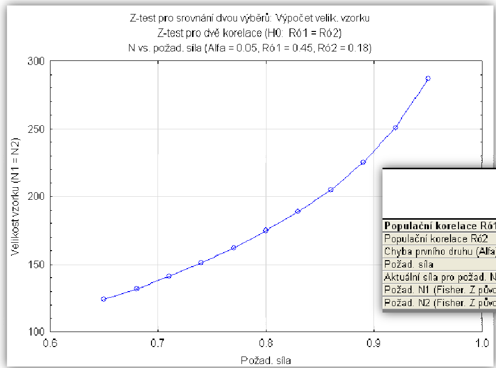


Jaké jsou vhodné rozsahy vzorků pro dosažení síly testu alespoň 0.70?



Pozn. Přípustné rozmezí pro požadovanou sílu testu je 0.65-0.999.

Rozsah výběru



	Výpočet velikos Dvě korelace, Z- H0: Ró1 = Ró2 Hodnota
Populační korelace Ró1	0.4500
Populační korelace Ró2	0.1800
Chyba prvního druhu (Alfa)	0.0500
Požad. síla	0.8000
Aktuální síla pro požad. N	0.8016
Požad. N1 (Fisher. Z původní)	175.0000
Požad. N2 (Fisher. Z původní)	175.0000

McNemarův test

StatSoft

McNemarův test

- Neparametrický test pro dvě **závislé** nominální veličiny (čtvercová kontingenční tabulka 2x2).
- Testujeme, zda jsou marginální četnosti stejné (test homogenity):

	Test B pozitivní	Test B negativní	Řádkové součty
Test A pozitivní	a	b	a+b
Test A negativní	c	d	c+d
Sloupcové součty	a+c	b+d	

StatSoft

McNemarův test

H$_0$: $p_b = p_c$ proti **H$_1$:** $|p_b - p_c| > 0$

Nulová hypotéza říká, že marginální pravděpodobnosti pro obě veličiny jsou shodné, tedy

$p_a + p_b = p_a + p_c$

a

$p_c + p_d = p_b + p_d$.

Testová statistika

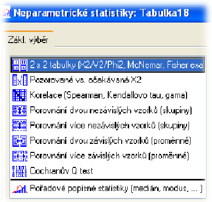
$\chi^2 = (|b-c|-1)^2/(b+c)$

	Test B pozitivní	Test B negativní	Řádkové součty
Test A pozitivní	a	b	a+b
Test A negativní	c	d	c+d
Sloupcové součty	a+c	b+d	

StatSoft

McNemarův test

- Porovnání dvou dichotomických proměnných (např. počet studentů, kteří zvládli testy základních matematických schopností na začátku semestru a na jeho ko



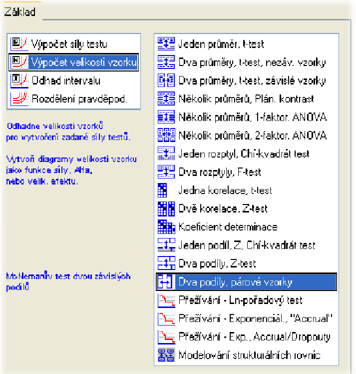
	Tabulka 2x2 (Tabulka18) Sloupec1	Sloupec2	Řádek celkem
Počet, řádek 1	25	16	41
Procent z celku	30.488%	19.512%	50.000%
Počet, řádek 2	37	4	41
Procent z celku	45.122%	4.878%	50.000%
Sloupec celkem	62	20	82
Procent z celku	75.610%	24.390%	
Chí-kvadrát (sv=1)	9.52	p= .0020	
V-kvadrát (sv=1)	9.41	p= .0022	
Yatesův korigovaný chí-kv.	8.00	p= .0047	
Fí-kvadrát	.11613		
Fisherovo p; jednostr.		p= .0020	
oboustr.		p= .0039	
McNemar. chí-kvadrát (A/D)	13.79	p= .0002	
McNemar. chí-kvadrát (B/C)	7.55	p= .0060	

StatSoft

Příklad: McNemarův test

- Jaký je vhodný rozsah výběru pro McNemarův test, který chceme použít pro testování zda spolu souvisí kouření a pití alkoholu, chceme-li sílu testu alespoň **0.80**?
- Máme-li k dispozici výběr celkem **250** jedinců, jakou můžeme očekávat sílu testu?

Rozsah výběru

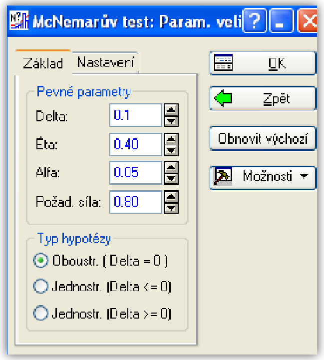


Delta
Hodnota mezi 0 a 1, je rovna rozdílu relativních četností (b-c)/n. H_0:δ=0.

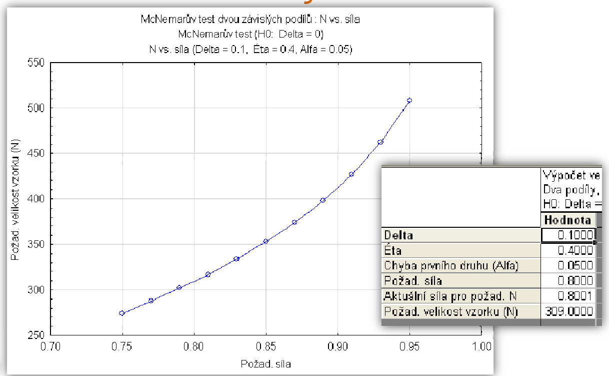
Éta
(Nuisance parameter) udává celkový součet četností (b+c)/n.

Alfa
Hladina testu, default je 0.05.

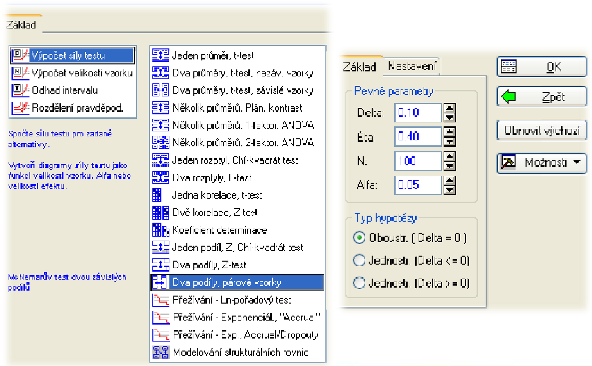
Rozsah výběru



Rozsah výběru



Síla testu

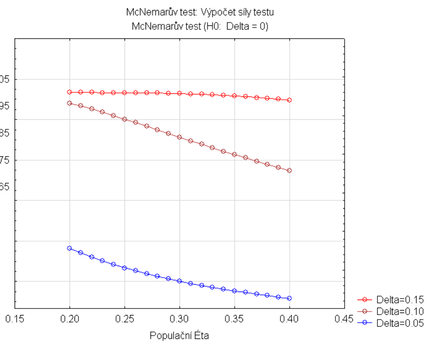


McNemarův test – síla testu

	Hodnota
Delta	0.1500
Éta	0.4000
Velikost vz. ve skup. (N)	250.0000
Chyba prvního druhu (Alfa)	0.0500
Síla	0.9718

	Hodnota
Delta	0.1000
Éta	0.4000
Velikost vz. ve skup. (N)	250.0000
Chyba prvního druhu (Alfa)	0.0500
Síla	0.7105

	Hodnota
Delta	0.0500
Éta	0.4000
Velikost vz. ve skup. (N)	250.0000
Chyba prvního druhu (Alfa)	0.0500
Síla	0.2374



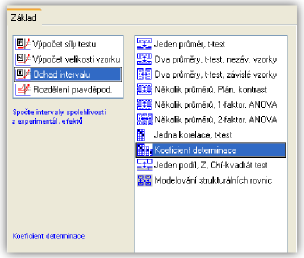
Koeficient determinace R^2

Regresní analýza

Lineární regrese modeluje závislost spojité proměnné pomocí spojitých nezávisle proměnných. Model je zpravidla hodnocen na základě ***p*-hodnot** testů nulovosti regresních koeficientů a tzv. **koeficientu determinace** (druhá mocnina vícenásobného korelačního koeficientu), který udává procento variability závisle proměnné vystižené daným modelem.

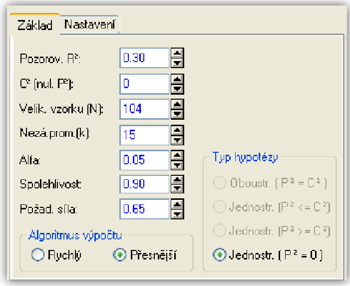
Regresní analýza

Podrobnější informaci o kvalitě modelu lze získat z **intervalu spolehlivosti** pro hodnotu **populačního koeficientu determinace P^2**.



Příklad: regrese

- Předpokládejme, že jsme vytvořili model vícenásobné regrese s **15** nezávisle proměnnými pro výběr o velikosti ***N*=104** a získali jsme koeficient determinace **R^2=0.30**.



Příklad: regrese

Interval spolehlivosti pro P^2

Dolní hranice je hodnota odhadu dolní meze intervalu spolehlivosti pro jednostrannou hypotézu.

Hranice pro sílu testu jsou post-hoc odhadem intervalu spolehlivosti pro sílu testu, požadované síle pak odpovídá odhad pro potřebný rozsah vzorku.

Koeficient determinace sice vyšel signifikantní, ale jeho interval spolehlivosti je poměrně široký a obsahuje i hodnoty blízké nule, proto lze očekávat, že i síla testu bude dosahovat velkých nebo malých hodnot.

	Odhad intervalu ([Není a Koeficient determinace Hodnota
Pozorov. R-kvadrát (R2)	0.3000
C2 (nul. P2)	0.0000
P-hodn. pro pozorov. R2	0.0039
Velik. vzorku (N)	104.0000
Nezávisle prom. (k)	15.0000
Interval spolehlivosti	0.9000
Meze spolehlivosti:	
P2:	
Dolní mez	0.0595
Horní mez	0.3257
Dolní hranice	0.0850
Síla:	
Dolní mez	0.2539
Horní mez	0.9952
Požad. N:	
Dolní mez	45.0000
Horní mez	241.0000

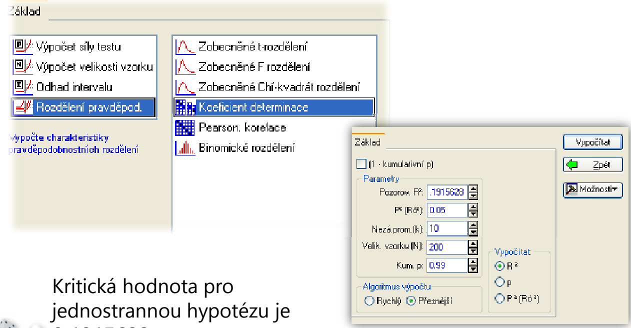
Testy o koeficientu determinace

- Zajímavým testem je test hypotézy H_0: $P^2 \geq a$.

Cíle:

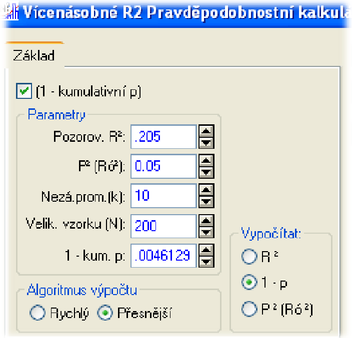
- **Kritické hodnoty** pro testy libovolných hypotéz o P^2
- **p-hodnota** pro pozorované R^2
- Stanovení **síly testu** pro testování hypotézy pro konkrétní alternativu, že P^2=0.30.

Koeficient determinace – kritické hodnoty



Kritická hodnota pro jednostrannou hypotézu je **0.1915628**.

Koeficient determinace – *p*-hodnota



Zobecněná rozdělení

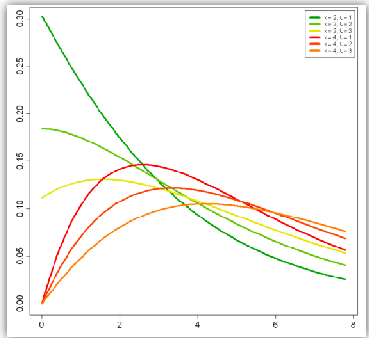
Necentrální χ^2-rozdělení

Náhodné veličiny X_i nechť jsou nezávislé a mají normální rozdělení $N(\mu_i,\sigma_i^2)$. Potom náhodná veličina

$$\sum_{i=1}^{k}\left(\frac{X_i}{\sigma_i}\right)^2$$

má **necentrální χ^2** dané počtem stupňů volnosti $\boldsymbol{k}$ a parametrem necentrality $\boldsymbol{\delta}$, definovaným často jako

$$\delta=\sum_{i=1}^{k}\left(\frac{\mu_i}{\sigma_i}\right)^2.$$



Hustota pravděpodobnosti

Příklad: Necentrální chí-kvadrát

- Chí-kvadrát test dobré shody
- Chí-kvadrát test nezávislosti pro kontingenční tabulky
- Chí-kvadrát test homogenity
  - Testové statistiky mají při splnění H_0 **chí-kvadrát rozdělení s ν stupni volnosti**
  - Pokud H0 neplatí, mají testové statistiky **necentrální chí-kvadrát** rozdělení se **stejným počtem stupňů volnosti ν** a s parametrem **necentrality δ**, který závisí na tvaru uvažované alternativní hypotézy

Příklad: Necentrální chí-kvadrát

- Uvažujme **chí-kvadrát test dobré shody** s nulovou hypotézou

$$H_0 : p_i = p_{i0}$$

a alternativou $H_1 : \exists i : p_i \neq p_{i0}$.

- **Testová statistika** $G_{k-1} = \sum_{i=1}^{k} \left(\frac{x_i - np_{i0}}{np_{i0}} \right)^2$ má asymptoticky **chí-kvadrát rozdělení** s ***k*-1** stupni volnosti.

- Nulovou hypotézu zamítáme, když $G_{k-1} > \chi^2_{k-1,1-\alpha}$.

Příklad: Necentrální chí-kvadrát

- Pro **výpočet síly testu**

$$P(G_{k-1} > \chi^2_{k-1,1-\alpha})$$

potřebujeme **parametr necentrality** $\delta = n \sum_{i=1}^{k} \frac{(p_{i1} - p_{i0})^2}{p_{i0}}$,

kde $\boldsymbol{p_{i1}}$ jsou hodnoty pravděpodobností dané konkrétní specifikovanou alternativou.

Příklad: Necentrální chí-kvadrát

Uvažujme konkrétně test hypotézy

$$H_0 : p_i = p_{i0} = \frac{1}{6}, \quad i = 1, \ldots, 6$$

Jestliže ve skutečnosti platí, že $\boldsymbol{p_{61}}$**=1/4**, a všechny ostatní pravděpodobnosti jsou homogenní.

A) najděte sílu testu pro tuto alternativu, pro hladinu spolehlivosti **α=0.05** a rozsah výběru ***n*=120**.

B) Najděte také **minimální n**, tak aby dosažená síla testu pro danou alternativu byla **alespoň 0.90**.

Příklad: Necentrální chí-kvadrát

Jestliže je $\mathbf{p_{61}}$**=1/4** a zbylé pravděpodobnosti jsou stejné, platí $\mathbf{p_{i1}}$**=3/20**, pro ***i*=1, ..., 5**.

Parametr necentrality je tedy roven

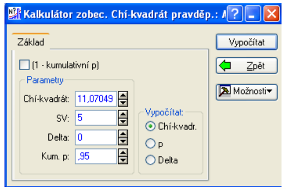
$$\delta = n \sum_{i=1}^{6} \frac{(p_{i1} - p_{i0})^2}{p_{i0}} = n \cdot \left[5 \cdot \frac{\left(\frac{3}{20} - \frac{1}{6} \right)^2}{\frac{1}{6}} + \frac{\left(\frac{1}{4} - \frac{1}{6} \right)^2}{\frac{1}{6}} \right] = \frac{n}{20} = 6$$

Příklad: Necentrální chí-kvadrát

$$P(G_{k-1} > \chi^2_{k-1,1-\alpha}) = P(G_{k-1} > 11.07) = 1 - X_{k-1,\delta}(11.07)$$

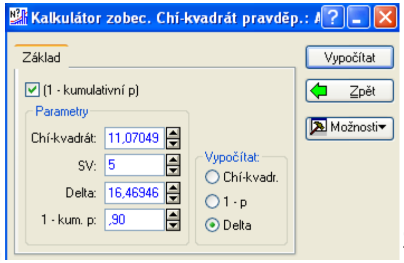
Nulovou hypotézu zamítáme pro hodnotu testové statistiky větší než 95% kvantil chí-kvadrát rozdělení s 5 stupni volnosti: **11.07**.

Síla testu je tedy 0.4329.



Příklad: Necentrální chí-kvadrát

Při změně velikosti rozsahu výběru se změní i parametr necentrality. Spočteme tedy nejprve jeho novou hodnotu pro sílu testu **1-*p*=0.90**



Nová hodnota parametru necentrality je rovna **16.5**, protože platí

$$\delta = \frac{n}{20},$$

dostáváme minimální rozsah výběru

$$20 \cdot \delta = 20 \cdot 16.5 = 330.$$

Příklad: Chí-kvadrát test nezávislosti

$$H_0 : \pi_{ij} = \pi_{i.}\pi_{.j}, \quad i = 1,\ldots,r,\ j = 1,\ldots,s$$

$$H_1 : \text{non } H_0$$

Testová statistika $G = \sum_{k=1}^{rs} \frac{(n_{ij} - n\pi_{i.}\pi_{.j})^2}{n\pi_{i.}\pi_{.j}}$

má za platnosti nulové hypotézy asymptoticky chí-kvadrát rozdělení s (*k*-1) stupni volnosti.

$$n\pi_{i0} \geq 5$$

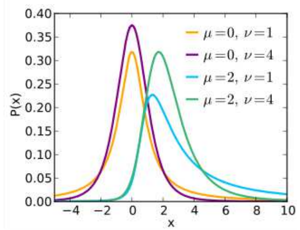
Kritický obor

$$W_\alpha = \{G \geq \chi^2_{1-\alpha}\}$$

Necentrální t-rozdělení

Parametr necentrality **delta** (v případě jednovýběrového testu je roven populačnímu průměru, v případě dvouvýběrového rozdílu populačních průměrů).

Nechť je **Z** náhodná veličina s rozdělením **N(0,1)** a **V** náhodná veličina s rozdělením **χ^2 s v stupni volnosti**, pak má **T** necentrální **t-rozdělení s parametrem necentrality δ a v stupni volnosti**.



$$T = \frac{Z + \delta}{\sqrt{\frac{V}{v}}}$$

Necentrální F-rozdělení

Má-li náhodná veličina **X necentrální χ^2** s parametrem necentrality δ a ν_1 stupni volnosti a **Y je náhodná veličina s χ^2 rozdělením** s ν_2 stupni volnosti, pak statistika

$$F = \frac{X/\nu_1}{Y/\nu_2}$$

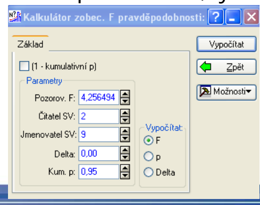
má **necentrální F rozdělení s ν_1 a ν_2 stupni volnosti a parametrem necentrality δ**.

Příklad: Vyvážená jednofaktorová ANOVA

Uvažujme test hypotézy $H_0 : \mu_1 = \mu_2 = \mu_3$

při skutečných hodnotách $\mathbf{\mu_1=59}$, $\mathbf{\mu_2=66}$ a $\mathbf{\mu_3=42}$ a dále $\mathbf{\sigma=12}$, $\mathbf{\alpha=0.05}$ a $\mathbf{n_1=n_2=n_3=4}$.

Testová statistika je porovnávána s 95% kvantilem F-rozdělení s 2 a 9 stupni volnosti, tj. s hodnotou **4.256**.



Příklad: Vyvážená jednofaktorová ANOVA

Při $\mu_1=59$, $\mu_2=66$ a $\mu_3=42$ a $\sigma=12$, $n_1=n_2=n_3=4$ je **parametr necentrality** roven

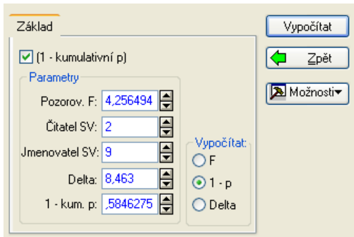
$$\delta = n\left(\sum_{j=1}^{k}\frac{1}{k}\frac{(\mu_j-\mu)^2}{\sigma^2}\right) =$$

$$= 12\left(\sum_{j=1}^{3}\frac{1}{3}\frac{(\mu_j-55.667)^2}{144}\right) = 8.463.$$

Příklad: Vyvážená jednofaktorová ANOVA

$$P(F > F_{0.95}(2;9)) = P(F > 4.256) = 1 - F(2;\, 9;\, \delta = 8.463)$$

Síla testu je **0.5846**.



Příklad: Vyvážená jednofaktorová ANOVA

Pro srovnání

$$RMSSE = \sqrt{\frac{\sum_{j=1}^{3}\left(\frac{\mu_j-\mu}{\sigma}\right)^2}{2}} = \sqrt{\frac{\sum_{j=1}^{3}\left(\frac{\mu_j-55.6667}{\sigma}\right)^2}{2}} = 1.0285$$

$$f = \sqrt{\frac{\sum_{j=1}^{3}\left(\frac{\mu_j-\mu}{\sigma}\right)^2}{3}} = 0.83979$$

	Výpočet síly testu (Ancova) ANOVA, 1-faktorová Pevné efekty
	Hodnota
Počet skupin	3,0000
Velikost vz. ve skup. (N)	4,0000
RMSSE	1,0285
Parametr zobec. t-rozdělení (Delta)	8,4625
Chyba prvního druhu (Alfa)	0,0500
SV efektu	2,0000
Chyba SV	9,0000
Kritická hodnota F	4,2565
Síla	0,5846

Příklad: Nevyvážená ANOVA

- V případě nevyvážené analýzy rozptylu (tj. rozsahy skupin se liší) je třeba pro výpočet parametru necentrality použít vzorec:

$$\delta = n\left(\sum_{j=1}^{k}\frac{n_j}{n}\frac{(\mu_j-\mu)^2}{\sigma^2}\right)$$

váhy jednotlivých skupin

Příklad: Nevyvážená ANOVA

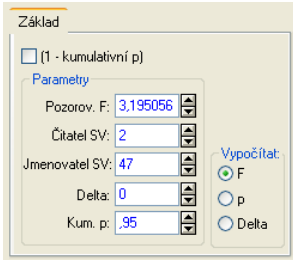
Testujeme hypotézu

$$H_0 : \mu_1 = \mu_2 = \mu_3.$$

Určete sílu testu, jsou-li skutečné hodnoty $\boldsymbol{\mu_1=3}$, $\boldsymbol{\mu_2=7}$ a $\boldsymbol{\mu_3=8}$ a dále $\boldsymbol{\sigma=4}$, $\boldsymbol{n_1=10}$, $\boldsymbol{n_2=20}$, $\boldsymbol{n_3=20}$.

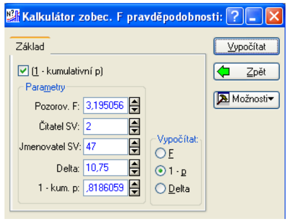
Příklad: Nevyvážená ANOVA

Testovou statistiku porovnáváme s hodnotou kvantilu $\mathbf{F_{0.95}(2; 47)=3.195}$.



Příklad: Nevyvážená ANOVA

$$\mu = \sum_{j=1}^{3} \frac{n_j}{n} \mu_j = \frac{1}{50}(10 \cdot 3 + 20 \cdot 7 + 20 \cdot 8) = 6.6$$



Parametr necentrality

$$\delta = n\left(\sum_{j=1}^{k} \frac{n_j}{n} \frac{(\mu_j - \mu)^2}{\sigma^2}\right) =$$

$$= 50\left(0.2 \cdot \frac{12.96}{16} + 0.4 \cdot \frac{0.16}{16} + 0.4 \cdot \frac{1.96}{16}\right) =$$

$$= 50 \cdot 0.215 = 10.75$$

Výsledná síla testu je **0.819**.

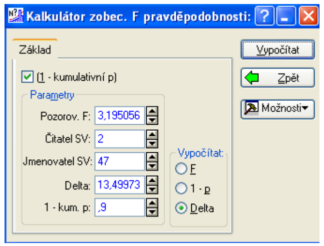
Příklad: Nevyvážená ANOVA

Jaký je **minimální rozsah výběru**, jestliže požadovaná síla je **0.9**?

Váhy jednotlivých skupin zachováme.

Příklad: Nevyvážená ANOVA

- Spočteme delta:



Ze vztahu

$$0.215n = \delta = 13.5$$

vyjádříme *n*, ***n***=62.78.

Aby rozsahy skupin byly celá čísla, vezmeme nejbližší vyšší násobek pěti, tedy n=65.

Nevyvážená ANOVA

- Návrh experimentu
  - Jednotlivé skupiny mají nějaké konkrétní **relativní četnosti** n_i v rámci **celé populace** *n*
- Výběr
  - Náhodně vybereme ***n*** jednotek z celé populace (relativní četnosti jsou zachovány)
  - Vybereme vždy $\boldsymbol{n_i = n/k}$ jednotek v každé skupině (relativní četnosti zachovány nejsou)
- Výsledek
- Předpokládejme, že v populaci je skupina malého rozsahu, $\boldsymbol{n_i = p}$, která má extrémní hodnoty (tj. průměr sledované charakteristiky je významně vyšší než v ostatních skupinách).
  - V prvním případě bude ***f***, resp. **RMSSE** menší než ve druhém případě, a tedy ve druhém případě získáme **větší sílu testu**.

Děkuji za pozornost.

lenka.blazkova@statsoft.cz

StatSoft®
STATISTICA