PSY117
Statistická analýza dat v psychologii
Přednáška 10 2017

Statistické testování hypotéz II

Přehled testů, rozdíly průměrů, velikost účinku, síla testu

The great tragedy of Science – the slaying of a beautiful hypothesis by an ugly fact

Thomas Huxley

From Phil. Trans. (1710) 27, 186-90.

II. *An Argument for Divine Providence, taken from the conftant Regularity obferv'd in the Births of both Sexes. By Dr.* John Arbuthnott, *Phyfitian in Ordinary to Her Majefty, and Fellow of the College of Phyfitians and the Royal Society.*

AMong innumerable Footfteps of Divine Providence to be found in the Works of Nature, there is a very remarkable one to be obferved in the

It is evident from what has been ſaid, that A's Lot for each Year is leſs than $\frac{1}{2}$; (but that the Argument may be ſtronger) let his Lot be equal to $\frac{1}{2}$ for one Year. If he undertakes to do the ſame thing 82 times running, his Lot will be $\overline{\frac{1}{2}}|^{82}$, which will be found eaſily by the Table of Logarithms to be $\frac{1}{4\ 8360\ 0000\ 0000\ 0000\ 0000\ 0000}$. But if A wager with B, not only that the Number of Males ſhall exceed that of Females, every Year, but that this Exceſs ſhall happen in a conſtant Proportion, and the Difference lye within fix'd limits; and this not only for 82 Years, but for Ages of Ages, and not only at *London*, but all over the World; (which 'tis highly probable is Fact, and deſigned that every Male may have a Female of the ſame Country and ſuitable Age) then A's Chance will be near an infinitely ſmall Quantity, at leaſt

leſs

less than any assignable Fraction. From whence it follows, that it is Art, not Chance, that governs.

There seems no more probable Cause to be assigned in Physicks for this Equality of the Births, than that in our first Parents Seed there were at first formed an equal Number of both Sexes.

Scholium. From hence it follows, that Polygamy is contrary to the Law of Nature and Justice, and to the Propagation of Human Race; for where Males and and Females are in equal number, if one *Man* takes Twenty Wives, Nineteen Men must live in Celibacy, which is repugnant to the Design of Nature; nor is it probable that Twenty Women will be so well impregnated by one Man as by Twenty.

Základní výzkumné otázky/hypotézy

1. Stanovení hodnoty parametru v populaci
   - **stanovení intervalu spolehlivosti** na μ, σ, ρ, b…
   - srovnání statistiky s hypotetickou hodnotou – konstantou

JEDNOVÝBĚROVÉ TESTY

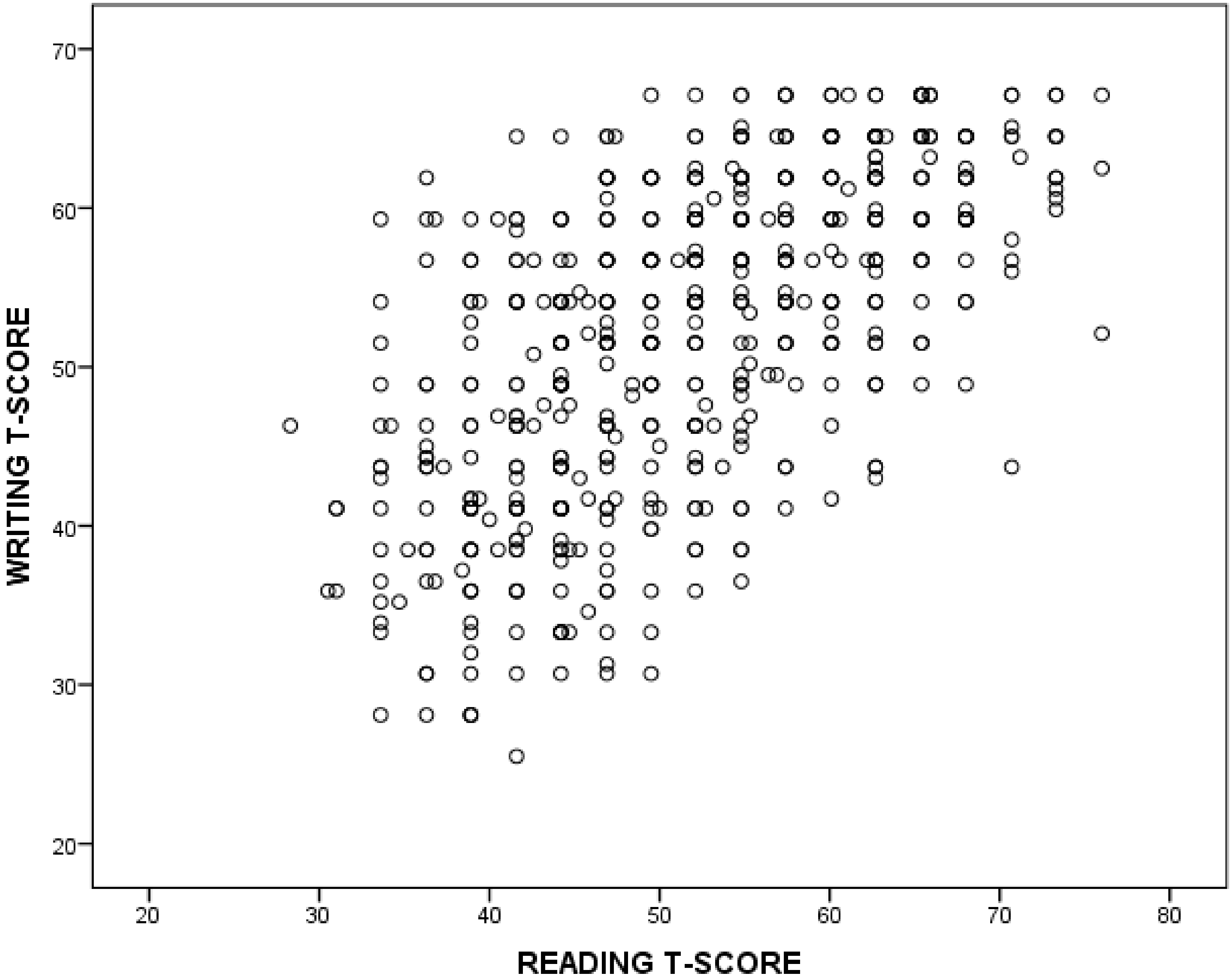
- Korelace mezi proměnnými
  - korelace, regrese, chí-kvadrát
  - H_1: $\rho \neq 0$ … H_0: $\rho = 0$
  - např. Mezi věkem a počtem návštěv lékaře za rok existuje lineární korelace.

2. Rozdíl mezi skupinami/vzorky - populacemi
   - mezi průměry, korelacemi, rozptyly, pravděpodobnostmi, pořadími….
   - lze srovnávat 2 i více skupin-populací
   - např. H_1: $\mu_1 - \mu_2 \neq 0$ … H_0: $\mu_1 - \mu_2 = 0$
   - např. Muži a ženy se liší v míře úzkostnosti.

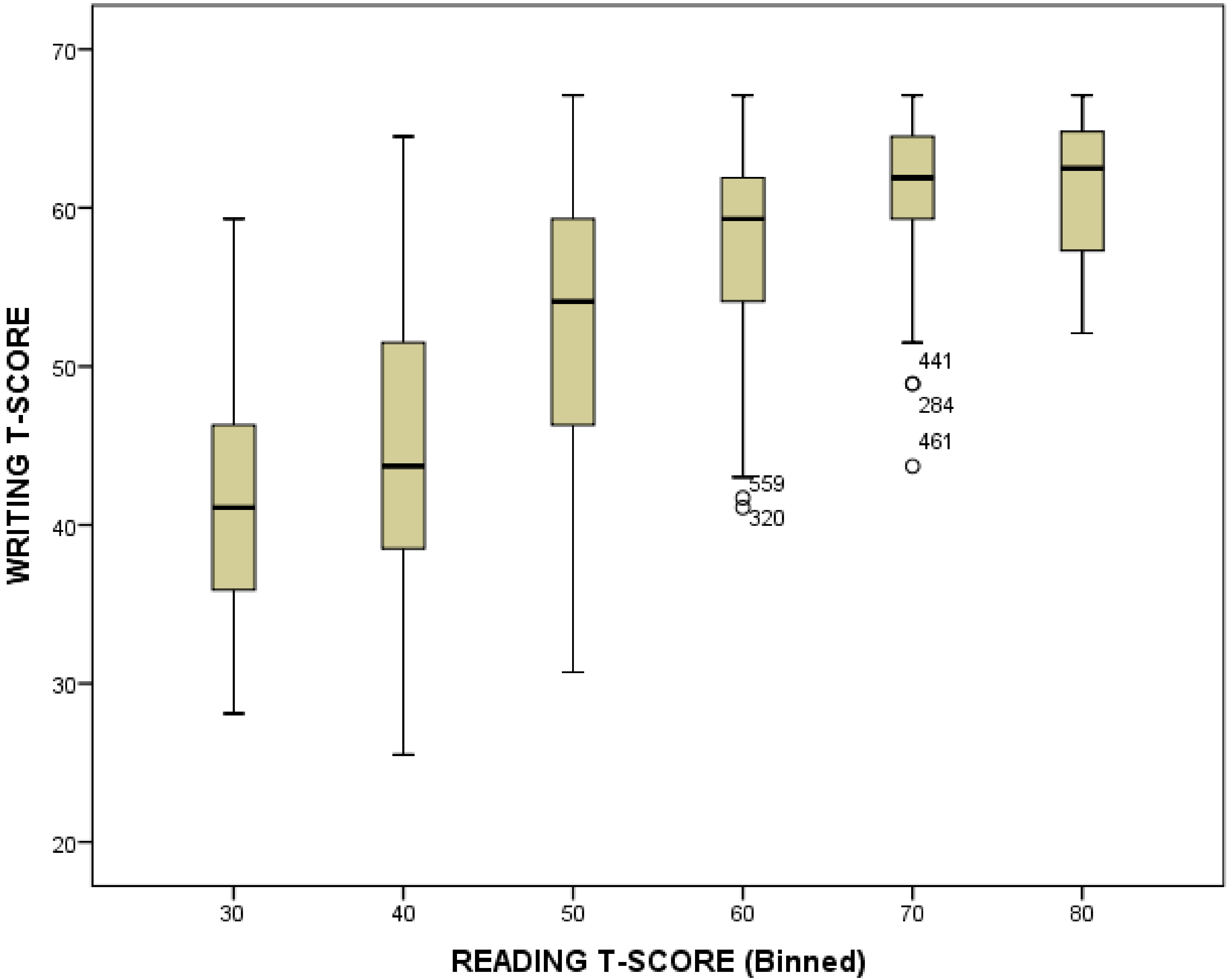
VÍCEVÝBĚROVÉ TESTY

- Rozdíl průměrů lze převést na korelaci a naopak - obecně mluvíme o **velikosti efektu/účinku**

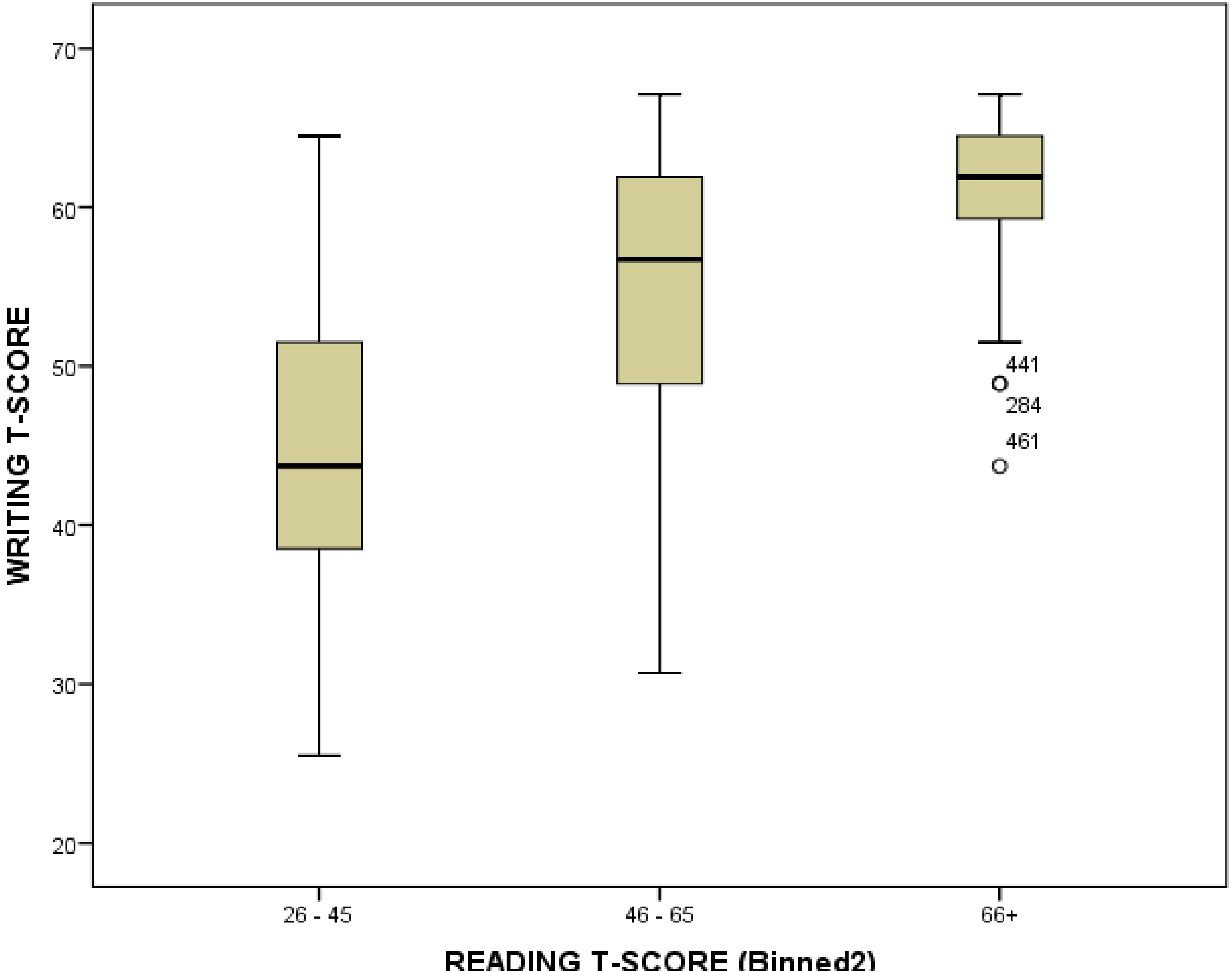
AJ: difference, association, effect size, two-tailed, one-tailed (directional)

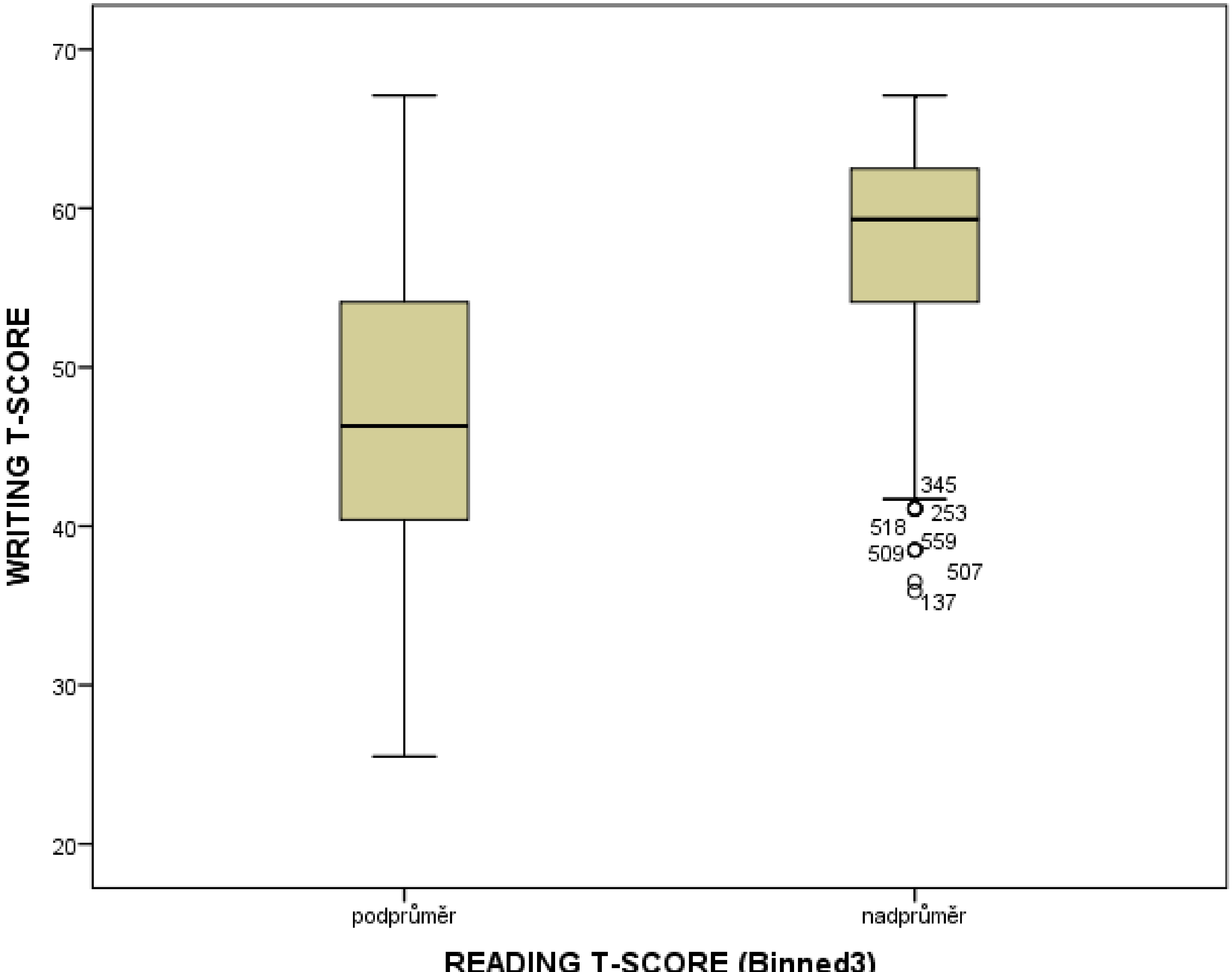


WRITING T-SCORE
70
60
50
40
30
20
20
30
40
50
60
70
80
READING T-SCORE



WRITING T-SCORE
70
60
50
40
30
20
441
284
461
559
320
30
40
50
60
70
80
READING T-SCORE (Binned)


70
60
50
40
30
20
WRITING T-SCORE
441
284
461
26 - 45
46 - 65
66+
READING T-SCORE (Binned2)


70
60
50
40
30
20
WRITING T-SCORE
345
253
518
559
509
507
137
podprůměr
nadprůměr
READING T-SCORE (Binned3)

Přehledy statistických testů

- **receptář Oseckých** třídění podle
  - počtu výběrů(skupin) – 1, 2, nebo více
  - úrovně měření – alternativní, nominální, pořadová, intervalová
  - typu procedury – interval spolehlivosti, test hypotézy, velikost potřebného výběru
- **Hendl – kapitola 12 a str. 235 (245 ve 3. vydání)**
- **online**
  - http://www.graphpad.com/www/book/Choose.htm
  - http://www.whichtest.info
  - http://www.socialresearchmethods.net/selstat/ssstart.htm
- Sheskin, D.J.: *Handbook of parametric and nonparametric statistical procedures*. CRC press, 2004.
- Kanji, G.K.: *100 statistical tests*. Sage, 2006.

Př.: Testy na rozdíly 2 středních hodnot

Intervalová závislá – rozdíly průměrů

- *párový test*: párový *t*-test
- *nezávislé skupiny*:
  - známý rozptyl v populaci: *z*-test
  - neznámý rozptyl v populaci: *t*-test pro nezávislé skupiny
    - varianta pro stejné a nestejné rozptyly mezi skupinami

Ordinální závislá – rozdíly mediánů, průměrného pořadí

- *párový test*: binomický znaménkový test, Wilcoxonovo *T* (int)
- *nezávislé skupiny*: Mann-Whitney *U*

Nominální závislá – shoda rozložení

- *párový test*: McNemarův test (dichotomie), Bowkerův test symetrie
- *nezávislé skupiny*: chí-kvadrát

AJ: sign test, chi-square, Wilcoxon *T*, Mann-Whitney *U*, paired(-samples) *t*-test (dependent, repeated measures), one-sample *t*-test, independent samples *t*-test

Co je potřeba znát?

- Testů v přehledech je mnoho…
- Pro každý je třeba znát
  - účel použití, testovaná hypotéza
  - předpoklady použití (úroveň měření, normalita)
  - interpretace výsledků (sjetiny z počítače)
- Co je třeba umět (ručně) spočítat?
  - všechny varianty *t*-testu (*z*-testu)
  - statistická významnost Pearsonova korelačního koeficientu
  - chí-kvadrát testy

Srovnání 2 nezávislých průměrů: *t* -test

Předpoklady použití ... jsou-li výrazně porušeny, volíme raději neparametrický test

- **intervalová** proměnná je v populaci **normálně rozložená** - neřeší se, je-li n_1,n_2 >30
- homogenita rozptylů (**homoscedascita**), pokud $n_1 \neq n_2$
  - řeší modifikace *t*-testu pro nestejné rozptyly (6.2.3)
  - testuje se Levenovým testem (od oka $\mathbf{s_1^2/s_2^2<2}$)
- **nezávislost pozorování** - řeší párový t-test (pro závislé výběry) (6.2.4)

- H_0: $\mu_1 - \mu_2 = 0$ (nebo roven konstantě, nebo >/< 0 či *c*) a zvolíme $\alpha = 1\%, 5\%$, nebo 10%
- **Rozdíl průměrů *d*** má

  s^2_{pooled}

  směrodatnou chybu $\boldsymbol{s_d}$ -2))]*[1/n_1+1/n_2]}

  ***t*-rozložení** s n_1+n_2-2 **stupni volnosti** (ν)
- Spočítáme **testovou statistiku *t*** = $(m_1 - m_2)/s_d = d/s_d$
- Zjistíme jaká je ***p* (*t* ≥ |zjištěná hodnota|)** - tabulky, T.DIST(*t* , ν, 1)
- Je-li $p \geq \alpha$, pak H_0 zůstává platná, je-li $p < \alpha$, H_0 zamítáme (a konstatujeme existenci statisticky významného rozdílu).
- Spočítáme Cohenovo *d* a interval spolehlivosti pro rozdíl průměrů.

Příklad: *t*-test pro nezávislé výběry

- *H*: Lidé s nízkou a vysokou depresivitou se liší v míře úzkostnosti.
  - H_0: $\delta = \mu_N - \mu_V = 0$
  - nasbíraná data: $m_N=2$; $m_V = 3$; $s_N=1{,}5$; $s_V= 1{,}6$; $n_N= n_V = 20$
  - H_0 budeme testovat na 5% hladině statistické významnosti, $\alpha = 0{,}05$
- Předpoklady splněny >> provádíme *t*-test pro nezávislé výběry (6.2.2)
- *rozdíl d* = $m_V - m_N = 3 – 2 = 1$
- $s_d= \sqrt{\{[((20-1)1{,}5^2+(20-1)1{,}6^2)/(20+20-2))]*[1/20+1/20]\}}=0{,}49$
- rozdíl má *t*-rozložení s $n_N+n_V-2 = 38$ stupni volnosti
- $t = (m_V - m_N)/s_d = 1/0{,}49 = 2{,}04$
- p $(t \geq |2{,}04|)$ je při $\nu = 38$ rovna 0,048 2*(1-T.DIST(2,04;38;1))=0,048
- $p < \alpha$, takže **zamítáme H_0.** Pokud by H_0 platila, zjištěný rozdíl by byl nepravděpodobný.
- 95% interval spolehlivosti: ${}_{0{,}025}t(38)$ = T.INV(0,025;38) = 2,02
  $d - 2{,}02*s_d < \delta < d + 2{,}02*s_d$, tj. $0{,}02 < \delta < 1{,}98$
- Cohenovo *d* = |1|/1,55 =0,65 , což je středně velký efekt.

Srovnání 2 závislých *m*: párový *t* -test

Předpoklady použití … jsou-li výrazně porušeny, volíme raději neparametrický test

- **intervalová** proměnná je v populaci **normálně rozložená** - neřeší se, je-li $N > 30$

2 ekvivalentní podoby testu, postupy:

a) pro každého člověka spočítat rozdíl $d_i = x_{1i} - x_{2i}$ a pak udělat jednovýběrový *t*-test testující H_0: $\delta = 0$

b) Nemáme-li data, je popisné statistiky pro srovnávané skupiny, pak…

- H_0: $\mu_1 - \mu_2 = 0$ (nebo roven konstantě, nebo >/< 0 či *c*) a zvolíme $\alpha = 1\%, 5\%$, nebo 10%
- **Rozdíl průměrů *d*** má **směrodatnou chybu** $s_d = \sqrt{((s_1^2 + s_2^2 - 2rs_1s_2)/N)}$

  ***t*-rozložení** s $N - 1$ **stupni volnosti** (ν)
- Spočítáme **testovou statistiku *t*** $= (m_1 - m_2)/s_d = d/s_d$
- Zjistíme jaká je ***p*** **(*t* ≥ |zjištěná hodnota|)** - tabulky, T.DIST(t , ν, 1)
- Je-li $p \geq \alpha$, pak H_0 zůstává platná, je-li $p < \alpha$, H_0 zamítáme (a konstatujeme existenci statisticky významného rozdílu).
- Spočítáme Cohenovo *d* a interval spolehlivosti pro rozdíl průměrů.

Příklad: párový *t*-test

- *H*: Lidé se liší v míře prožívané úzkosti před zkouškou a po zkoušce.
  - H_0: $\delta = \mu_{\text{PŘED}} - \mu_{\text{PO}} = 0$
  - nasbíraná data: $m_{\text{PŘED}}=2$; $m_{\text{PO}} = 3$; $s_{\text{PŘED}}=1{,}5$; $s_{\text{PO}}= 1{,}6$; $N = 20$; $r=0{,}6$
  - H_0 budeme testovat na 5% hladině statistické významnosti, $\alpha = 0{,}05$
- Předpoklady splněny >> provádíme párový *t*-test (6.2.4)
- *rozdíl d* = $m_{\text{PŘED}} - m_{\text{PO}} = 3 - 2 = 1$
- $s_d = \sqrt{(1{,}5^2+1{,}6^2-2*0{,}6*1{,}5*1{,}6)/20}=0{,}31$
- rozdíl má *t*-rozložení s $N - 1 = 19$ stupni volnosti
- $t = (m_{\text{PŘED}} - m_{\text{PO}})/s_d = 1/0{,}31 = 3{,}23$
- p ($t \geq |3{,}23|$) je při $\nu = 19$ rovna 0,004 2*(1-T.DIST(3,23;19;1))=0,0044
- $p < \alpha$, takže **zamítáme H_0.** Pokud by H_0 platila, zjištěný rozdíl by byl nepravděpodobný.
- 95% interval spolehlivosti: $_{0,025}t(19)$ = T.INV(0,025;19) = 2,09
  $d - 2{,}09*s_d < \delta < d + 2{,}09*s_d$, tj. $0{,}35 < \delta < 1{,}65$
- Cohenovo *d* = |1|/1,55 =0,65 , což je středně velký efekt.

Velikost účinku/efektu

- Možnost srovnání mezi studiemi zkoumajícími tutéž výzkumnou otázku pomocí různě operacionalizovaných proměnných
- Možnost srovnání velikosti efektu vyjádřeného různými koeficienty
- Snadnější interpretace

Pro rozdíly středních hodnot

- **Cohenovo *d*** $= |m_1 - m_2|/s_{pooled}$; $s_{pooled} = \sqrt{[((n_1-1)s_1^2 + (n_2-1)s_2^2)/(n_1+n_2-2))]}$
- varianta $d' = |m_1 - m_2|/s_{con}$; $s_{con} = s$ kontrolní skupiny

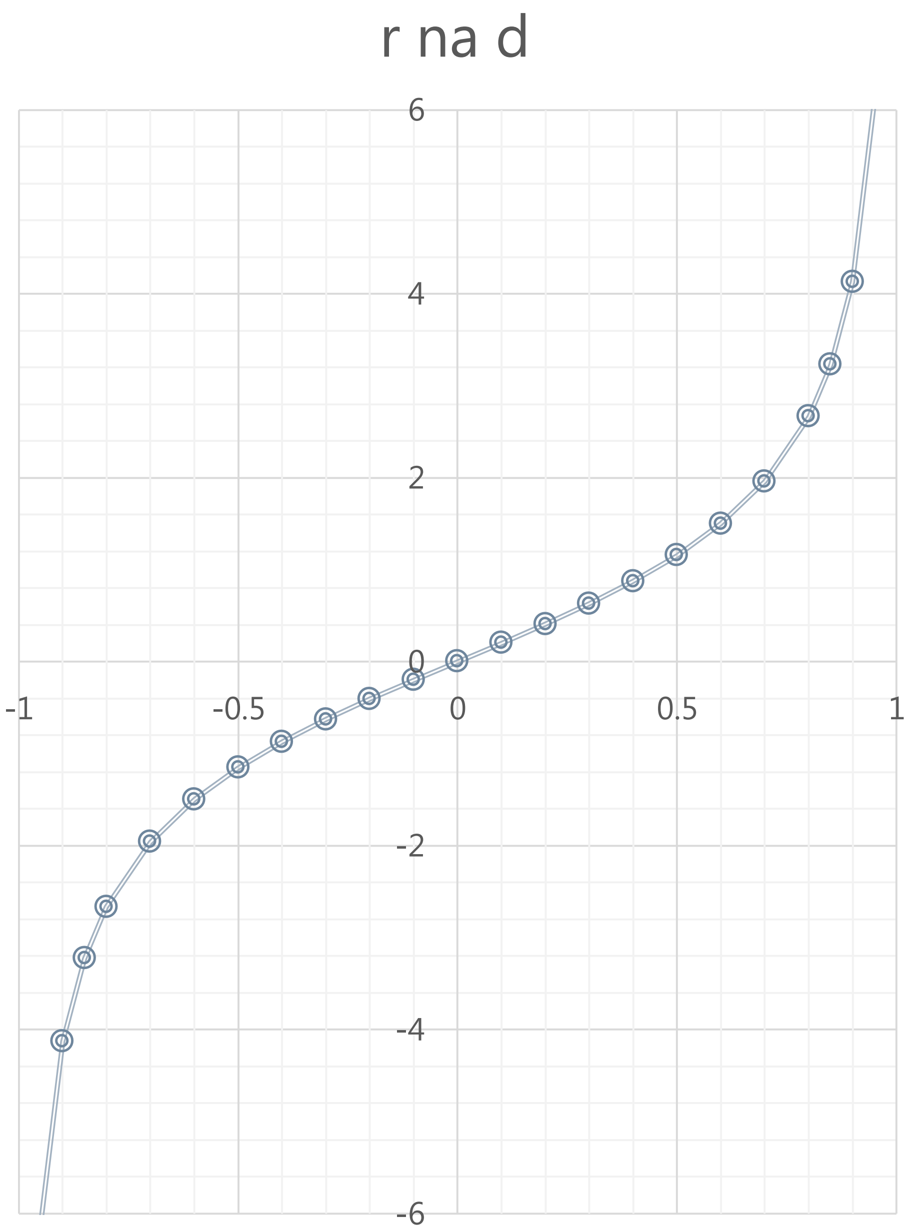
Pro těsnost vztahu (korelace)

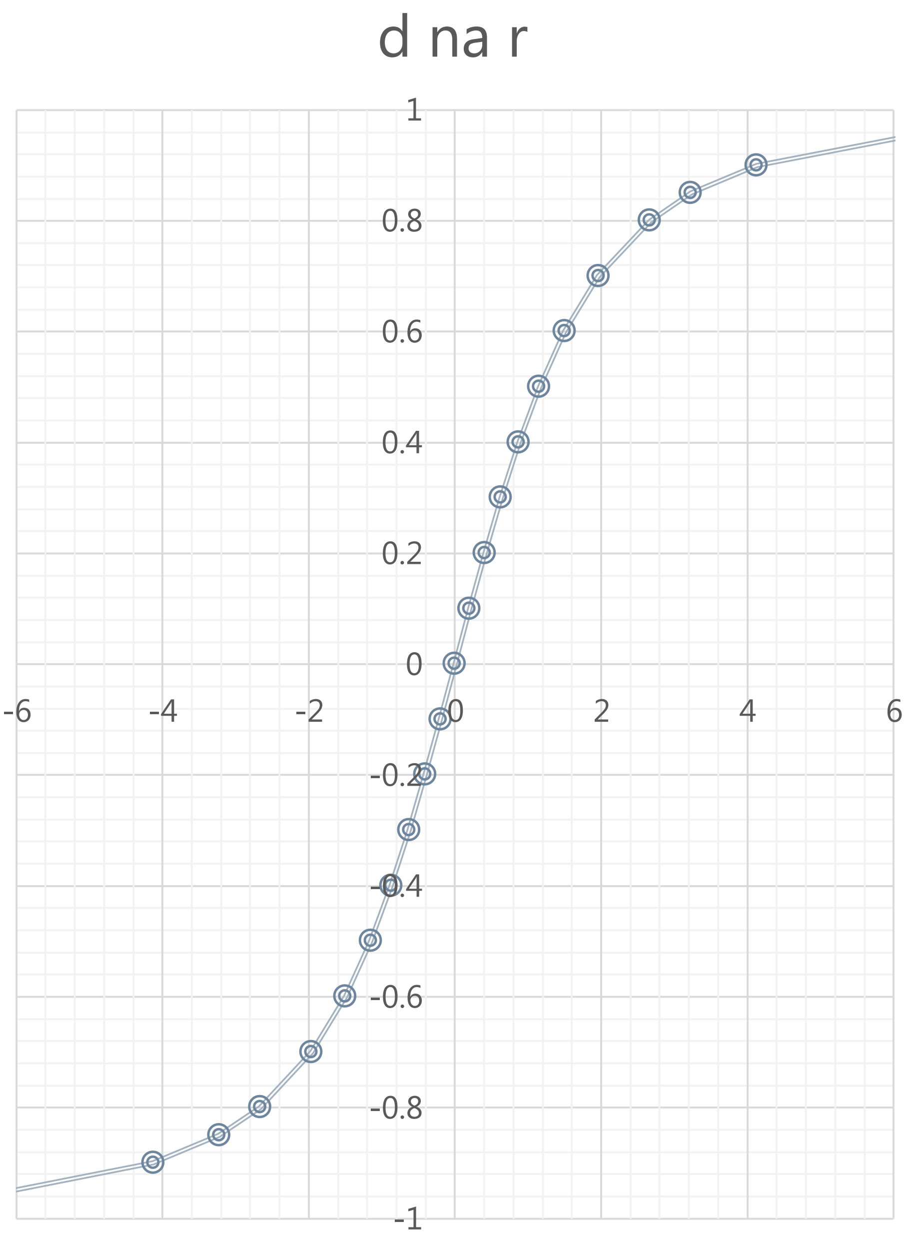
- r a r^2, R^2, η^2(eta), ω^2 – podíl vysvětleného rozptylu závislé proměnné

Indikátory velikosti efektu lze mezi sebou navzájem převádět

- Cohenovo d na r: $r = \sqrt{(d^2/(d^2 + 4))}$
- r na Cohenovo d: $d = 2r/\sqrt{(1 - r^2)}$

AJ: effect size, Cohen's d, strength of association, explained variance
https://en.wikipedia.org/wiki/Effect_size


r na d
6
4
2
0
-2
-4
-6
-1
-0.5
0
0.5
1


d na r
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-6
-4
-2
0
2
4
6

Síla testu

Síla testu (1-β) je pravděpodobnost, že existující rozdíl bude detekován, zjištěn jako statisticky významný.

Záleží na

- skutečné velikosti účinku (δ, ρ...)
- variabilitě proměnné(ých) – s,σ
- velikosti vzorku n
- zvoleném riziku chyby I. typu, α: čím nižší je α tím nižší je síla
- zvoleném testu (parametrické mají vyšší sílu)

Obvykle toužíme po co nejvyšší síle testu, cca 0,8 a výše.

- Bojujeme o ni především velikostí vzorku a kontrolou intervenujících proměnných (snižuje s).

A priori stanovení *N* pro dosažení potřebné síly testu

- Pro každý test hypotézy stanovujeme trochu jinak
- př:jednovýběrový t-test: $n > (z_{1-\alpha/2} + z_{1-\beta})^2 (1/d^2)$
- G*Power http://www.gpower.hhu.de/en.html
  - stanovení potřebné velikosti vzorku pro běžně testované hypotézy
  - manuál: http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf

Publikace výsledků testování hypotéz

- *Primárně* udáváme velikost efektu, nejlépe i s intervalem spolehlivosti
- *Sekundárně* udáváme výsledek statistického testování
  - udáváme získanou hodnotu *p* (Sig.)
  - uvádíme i testovou statistiku (i se stupni volnosti) – r, $t(\nu)$, $F\ (\nu_1, \nu_2)$, χ^2, M-W *U*...
  - Př. Průměr spokojenosti mužů je o 10 bodů vyšší než průměr spokojenosti žen, 95% CI (8;12), $t(200)=4{,}8$, $p<0{,}001$, Cohen $d=0{,}68$.
- Interpretujeme nejlépe interval spolehlivosti. Výsledek statistického testování interpretujeme vzhledem k použité nulové hypotéze.

Testy normality rozložení

- Kolmogorov-Smirnov s Lillieforsovou korekcí, Shapiro-Wilk, D'Agostino-Pearson a jiné
- Testují H_0, že rozložení proměnné se neliší od normálního rozložení
  - jsou to jedny z tzv. **testů dobré shody** (goodness-of-fit tests)
  - testovaná H_0 je shoda; tj. $p<\alpha$ = příliš velká odchylka od normality
- **Jejich užívání je kontroverzní**
  - na malých vzorcích nenormalitu nedetekují (při n=20, 1-β < 0,5)
  - na velkých vzorcích (n > 1000) jsou naopak extrémně přísné
  - *t*-testy a ANOVA jsou proti narušení normality robustní, takže nám obvykle stačí konstatovat unimodalitu bez extrémního zešikmení
  - pro rozhodování mezi použitím parametrických a neparametrických testů volíme spíše **úroveň měření** a velikost vzorku

AJ: tests of (univariate) normality, goodness-of-fit tests

Více: http://www.psy.surrey.ac.uk/cfs/p8.htm, http://www.graphpad.com/library/BiostatsSpecial/article_197.htm

„Test signifikance“ Pearsonova korelačního koeficientu

„Testem signifikance“ se míní test H_0: ρ=0

Pokud H_0: ρ=0, pak

- Z=FISHER(ρ) má normální výběrové rozložení se $s_Z=1/\sqrt{(n-3)}$
- z=FISHER(r)/s_Z ~ N(0;1)
- P(D|H_0)=2*(1 – NORM.S.DIST(Z/s_Z;1) pro oboustrannou (non-directional) H_1

Pokud H_0: ρ=c, pak

- D_Z=(FISHER(r)–FISHER(c)) má normální výběrové rozl. se $s_Z=1/\sqrt{(n-3)}$
- $z=D_Z/s_Z$ ~ N(0;1)
- P(D|H_0)=2*(1 – NORM.S.DIST(D_Z/s_Z;1) pro oboustrannou H_1

Příklady na test signifikance *r*

$r=0,5;\ N=20$

- $s_Z=1/\surd(20-3)=0,24$
- $z=\text{fisher}(0,5)/0,24=0,55/0,24=2,26$
- $P(z\geq 2,26|Z=0)=2*(1\text{-NORM.S.DIST}(2,26;1))=0,02$

$r=0,6;\ N=10$

- $s_Z=1/\surd(10-3)=0,38$
- $z=\text{fisher}(0,6)/0,38=0,69/0,38=1,83$
- $P(z\geq 1,83|Z=0)=2*(1\text{-NORM.S.DIST}(1,83;1))=0,07$

$r=0,4;\ N=20;\ H_0\text{: }\rho=0,8$

- $s_Z=1/\surd(20-3)=0,24$
- $z=(\text{fisher}(0,8)-\text{fisher}(0,4))/0,24=(1,1\text{-}0,42)/0,24=0,67/0,24=2,78$
- $P(z\geq 2,78|Z=0)=2*(1\text{-NORM.S.DIST}(2,78;1))=0,005$