

PSY117

Statistická analýza dat v psychologii

**Přednáška 10 2018**

---

# **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*

( 186 )

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

---

II. *An Argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes. By Dr. John Arbuthnott, Physician in Ordinary to Her Majesty, and Fellow of the College of Physicians and the Royal Society.*

**A**mong innumerable Footsteps of Divine Providence to be found in the Works of Nature, there is a very remarkable one to be observed 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  $\frac{1}{2}^{82}$ , which will be found easily by the Table of Logarithms to be

$\frac{1}{2}^{82}$

---

4 8360 0000 0000 0000 0000 0000 0000 0000

But if A wagers 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 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  $\mu$ ,  $\sigma$ ,  $\rho$ ,  $b$ ...
- srovnání statistiky s hypotetickou hodnotou – konstantou
- 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.

JEDNOVÝBĚROVÉ  
TESTY

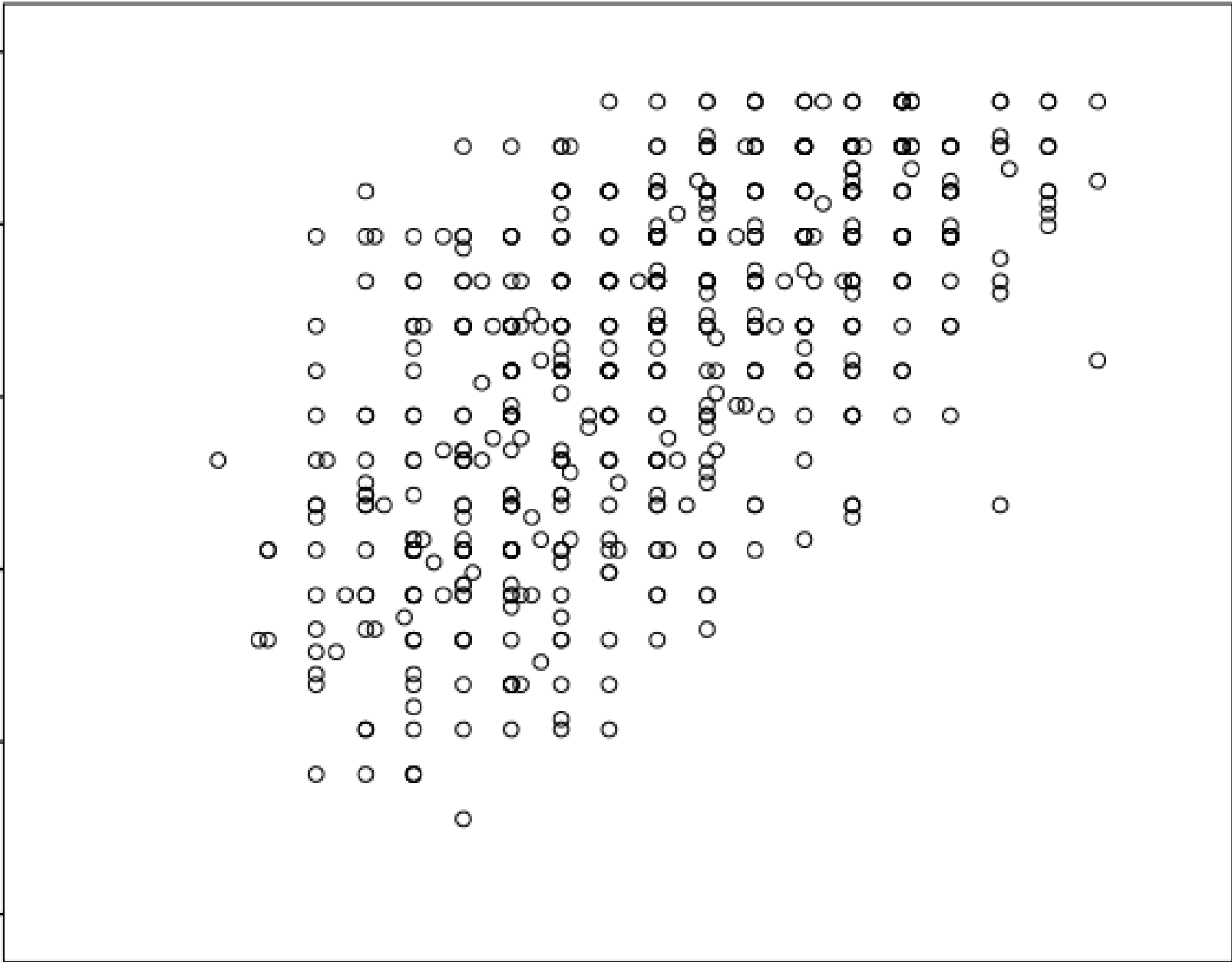
## 2. Rozdíl mezi skupinami/vzorky - populacemi

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

VÍCEVÝBĚROVÉ  
TESTY

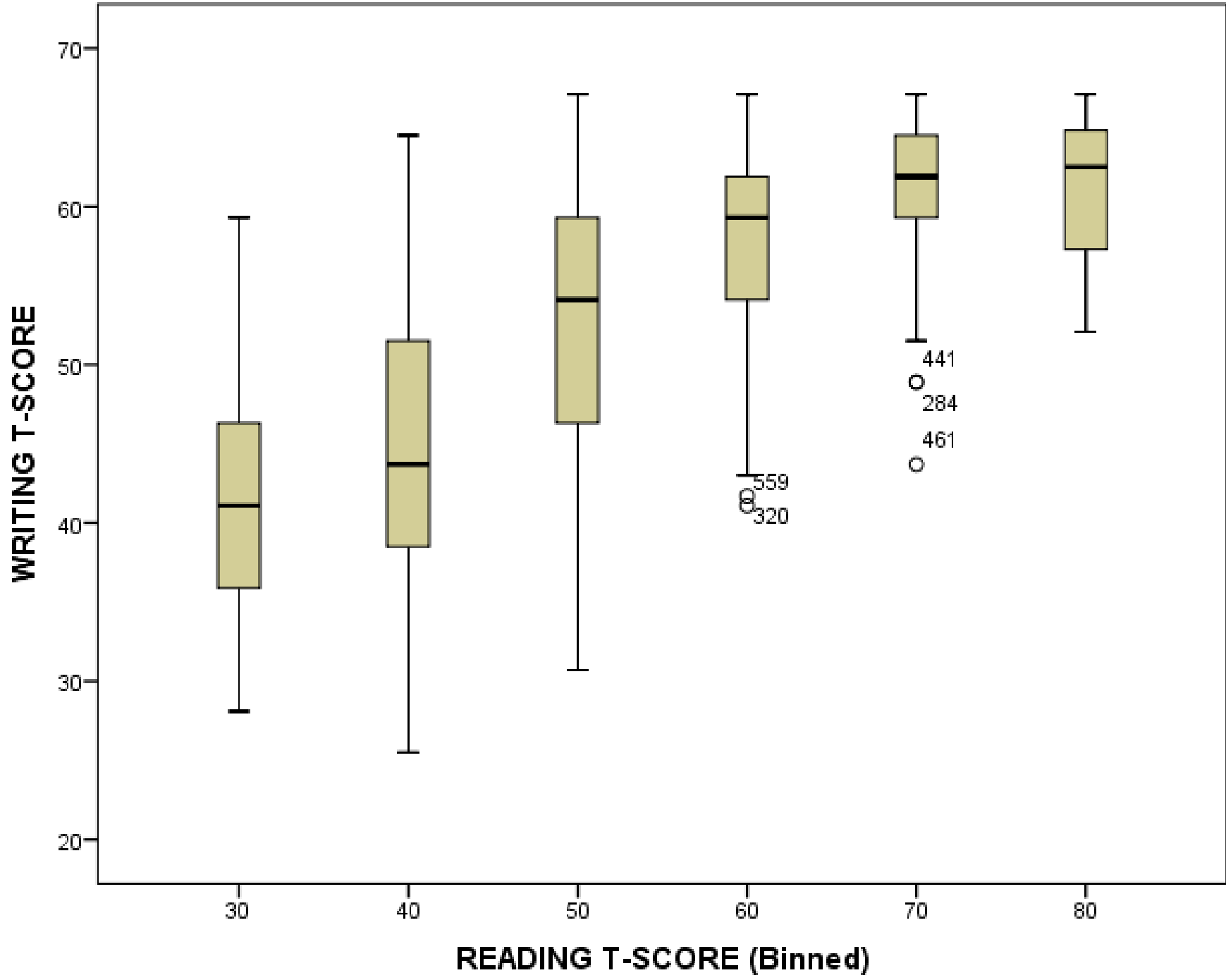
WRITING T-SCORE

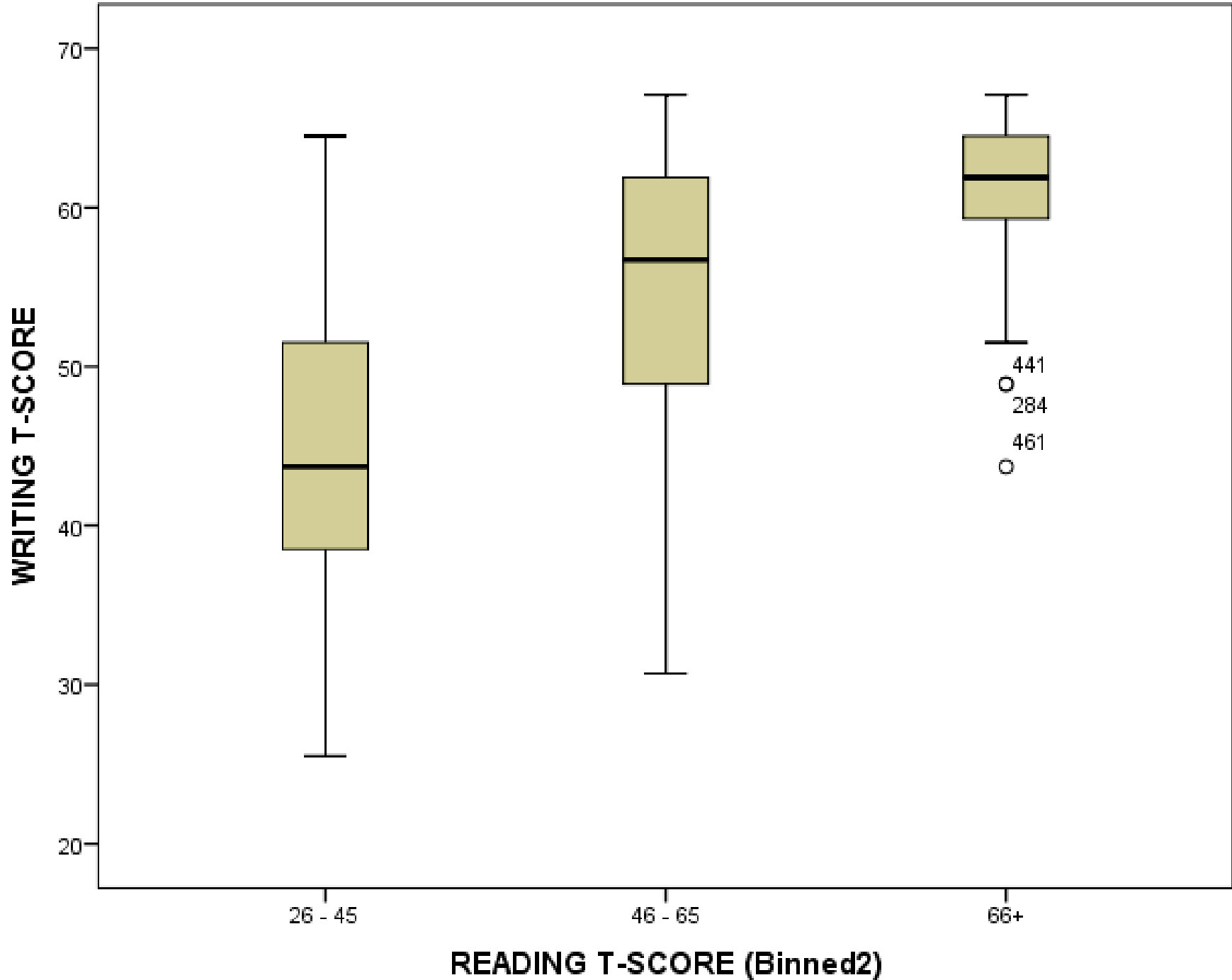
70  
60  
50  
40  
30  
20



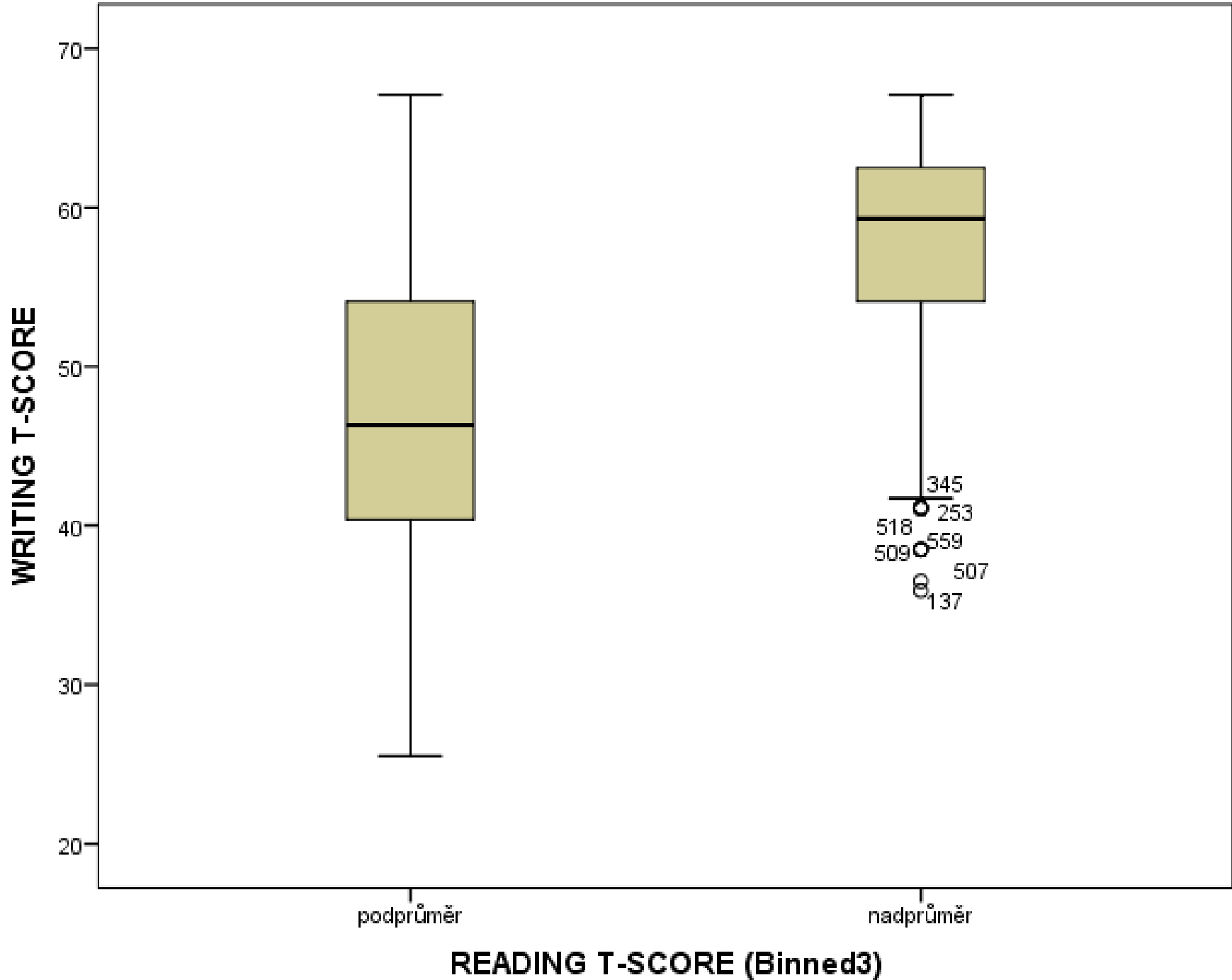
READING T-SCORE

20 30 40 50 60 70 80









# 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 nestejně 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**

# 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 nestejně rozptyly (6.2.3)
    - testuje se Levenovým testem (od oka  $s_1^2/s_2^2 < 2$ )
  - **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  $s^2_{\text{pooled}}$  směrodatnou chybu  $s_d = \sqrt{s^2_{\text{pooled}} \cdot (1/n_1 + 1/n_2)}$   **$t$ -rozložení s  $n_1 + n_2 - 2$  stupni volnosti ( $\nu$ )**
- Spočítáme **testovou statistiku  $t = (m_1 - m_2)/s_d = d/s_d$**
- Zjistíme jaká je  **$p$  ( $t \geq |\text{zjištěná hodnota}|$ )** - tabulky, T.DIST( $t, \nu, 1$ )
- 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 ( $\nu$ )**
  - Spočítáme **testovou statistiku  $t = (m_1 - m_2)/s_d = d/s_d$**
  - Zjistíme jaká je  **$p$  ( $t \geq |\text{zjištěná hodnota}|$ )** - tabulky, T.DIST( $t, \nu, 1$ )
  - 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  $d = m_{\text{PŘED}} - m_{\text{PO}} = 3 - 2 = 1$
- $s_d = \sqrt{(1,5^2 + 1,6^2 - 2 \cdot 0,6 \cdot 1,5 \cdot 1,6) / 20} = 0,31$
- 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  $2 \cdot (1 - T.DIST(3,23; 19; 1)) = 0,0044$
- $p < \alpha$ , takže **zamítáme  $H_0$** . Pokud by  $H_0$  platila, zjištěný rozdíl by byl nepravděpodobný.
- 95% interval spolehlivosti:  ${}_{0,025}t(19) = T.INV(0,025; 19) = 2,09$   
 $d - 2,09 \cdot s_d < \delta < d + 2,09 \cdot s_d$ , tj.  $0,35 < \delta < 1,65$
- Cohenovo  $d = |1| / 1,55 = 0,65$ , což je středně velký efekt.



# Standardizovaná velikost účinku/ efektu

---

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

## Pro rozdíly středních hodnot

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

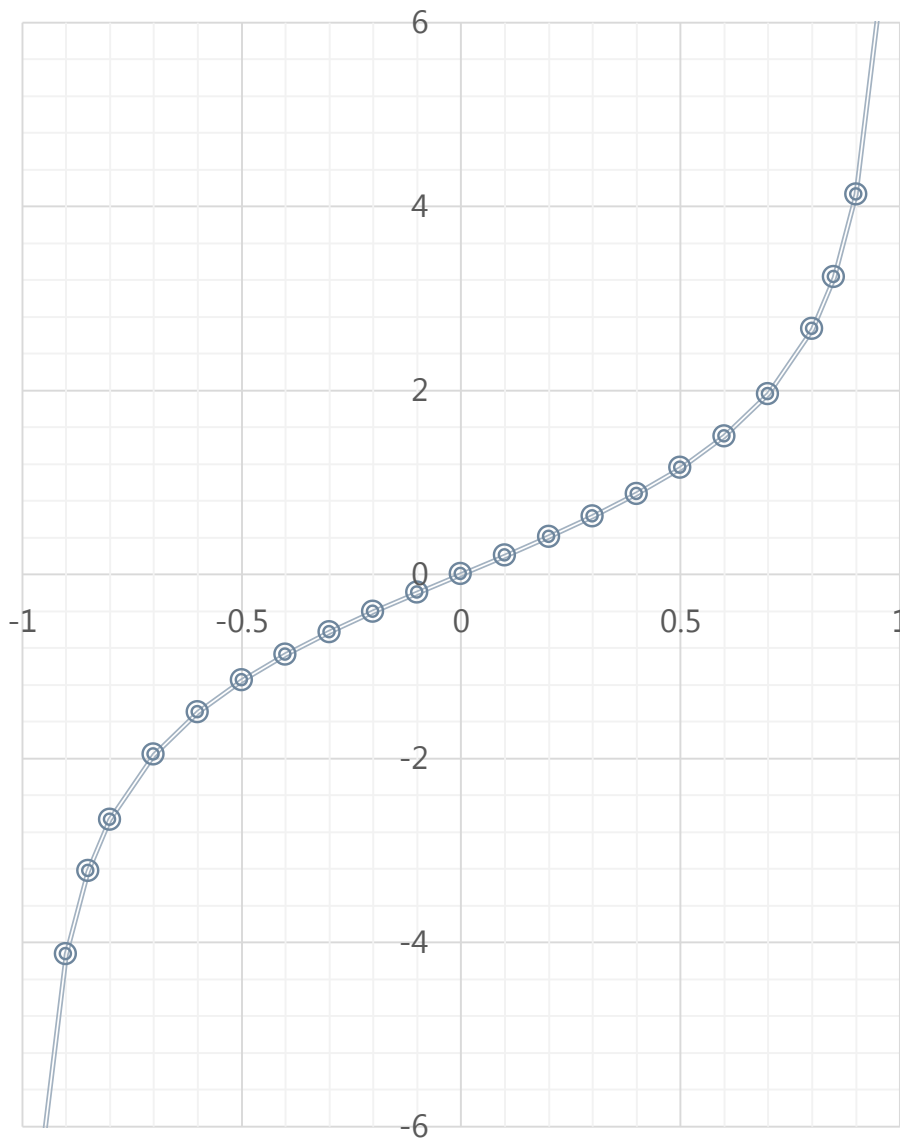
## Pro těsnost vztahu (korelace)

- $r$  a  $r^2$ ,  $R^2$ ,  $\eta^2$ (eta),  $\omega^2$  – podíl vysvětleného rozptylu závislé proměnné

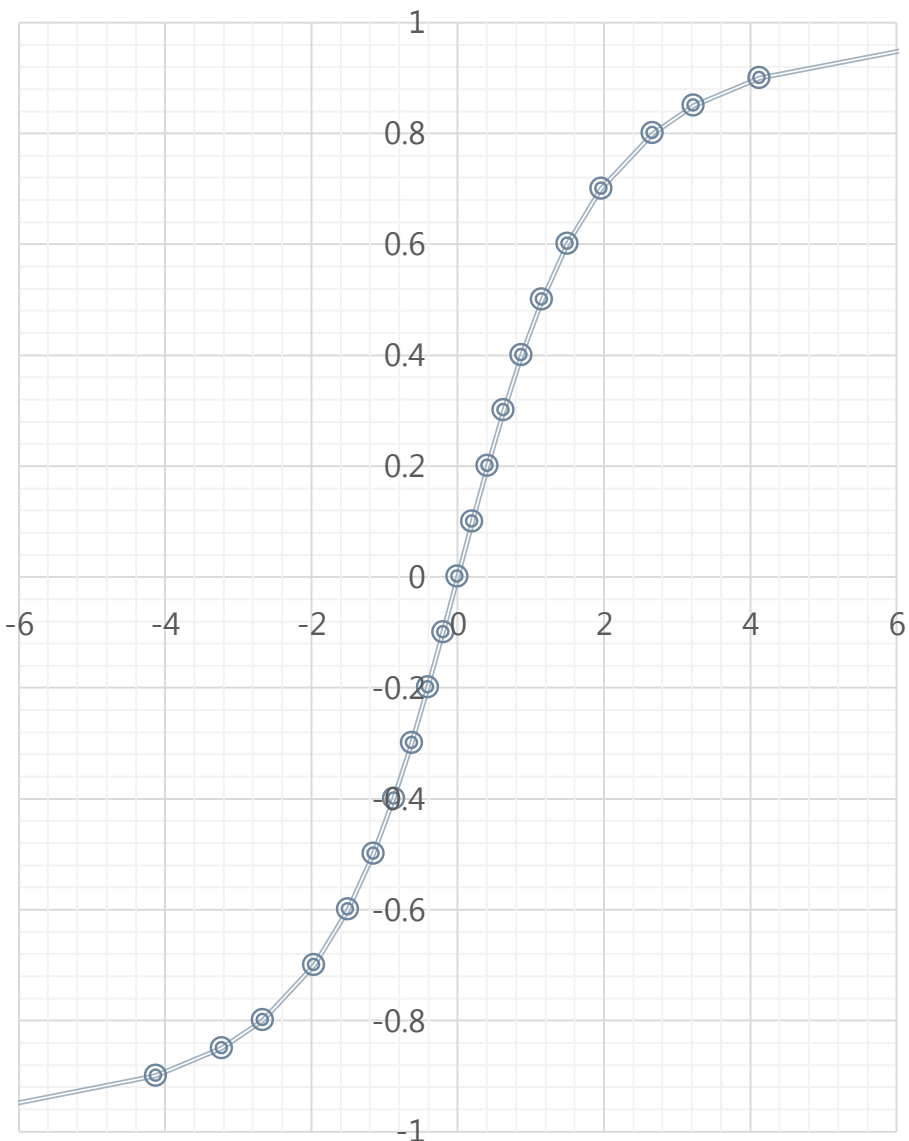
Indikátory velikosti efektu lze mezi sebou navzájem převádět

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

r na d



d na r



# Síla testu

---

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

Proč nám na ní záleží?

- Malé  $N$  -> málo informace k zamítnutí  $H_{(0)}$
- Nezamítnutí  $H_{(0)}$  – nepříliš uspokojivý závěr výzkumu.
  - Možná byla  $H$  mylná, ale taky jsme možná měli smůlu, nebo jsme něco zkazili. Třeba jsme měli moc malý vzorek.
  - Nezamítnutí  $H_{(0)}$  neznamená automaticky podporu pro  $H_{(0)}$
  - Zbytečně vynaložené náklady
- Velké  $N$  = vysoké náklady

Chceme dělat výzkum tak, abychom, pokud jsou naše výzkumné předpoklady správné, měli vysokou  $P$  podpory  $H$

$$P[ P(D|H_0) < \alpha \mid H_1 \text{ je pravdivá } ]$$

# Síla testu – co ji ovlivňuje

---

Bezprostředně:

- ❑ skutečná velikost účinku ( $\delta, \rho..$ ) – jak velký rozdíl či těsný vztah předpokládáme
- ❑ velikost vzorku  $N$
- ❑ zvolené riziko chyby I. typu,  $\alpha$ : čím nižší je  $\alpha$  tím nižší je síla

Zprostředkovaně:

- ❑ variabilita proměnné(ých) –  $s, \sigma$  ...experimentální kontrola
- ❑ zvoleném testu (parametrické mají vyšší sílu)

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

---

Simulace: <http://rpsychologist.com/d3/NHST/>

# Síla testu – jak ji spočítáme?

- Těžko, pro každý test trochu jinak
- Pro oboustranný t-test:
  - Hypotetizujeme velikost účinku  $d$
  - Stanovíme hladinu významnosti  $\alpha$
  - Budeme mít vzorek velikosti  $N$

alfa	0,05
počet "stran" testu	2
N	100
Cohenovo d	0,5
kritická hodnota t	1,98
t pro hypotetizované d	5,00
power	0,9984

$$\left[ \begin{array}{c} ( \quad ) \sqrt{\quad} \\ ( \quad ) \sqrt{\quad} \\ ( \quad ) \sqrt{\quad} \end{array} \right] \quad \left[ \begin{array}{c} \sqrt{\quad} \\ ( \quad ) \sqrt{\quad} \\ ( \quad ) \sqrt{\quad} \end{array} \right]$$

# A priori stanovení $N$ pro dosažení potřebné síly testu

---

- Jak velké  $N$  potřebujeme, abychom měli sílu testu vyšší než 0,8 při velikosti účinku  $d$ ?
- Pro každý test hypotézy stanovujeme trochu jinak
- př: jednovýběrový oboustranný t-test:  
$$N > [(z_{1-\alpha/2} + z_{\beta})/d]^2$$
- 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](http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf)

# Komunikace výsledků testování hypotéz

---

- *Primárně* udáváme velikost efektu, nejlépe i s intervalem spolehlivosti
  - *Sekundárně* udáváme výsledek statistického testování
    - udáváme **testovou statistiku** (vč. df) –  $r$ ,  $t(\nu)$ ,  $F(\nu_1, \nu_2)$ ,  $\chi^2$ , M-W  $U$ ... **a získanou hodnotu  $p$**
    - 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  $n=20$ ,  $1-\beta < 0,5$ )
  - 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



# „Test signifikance“ Pearsonova korelačního koeficientu

---

„Testem signifikance“ se míní test  $H_0: \rho=0$

Pokud  $H_0: \rho=0$ , pak

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

Pokud  $H_0: \rho=c$ , pak

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

# Příklady na test signifikance $r$

---

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

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

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

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

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

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