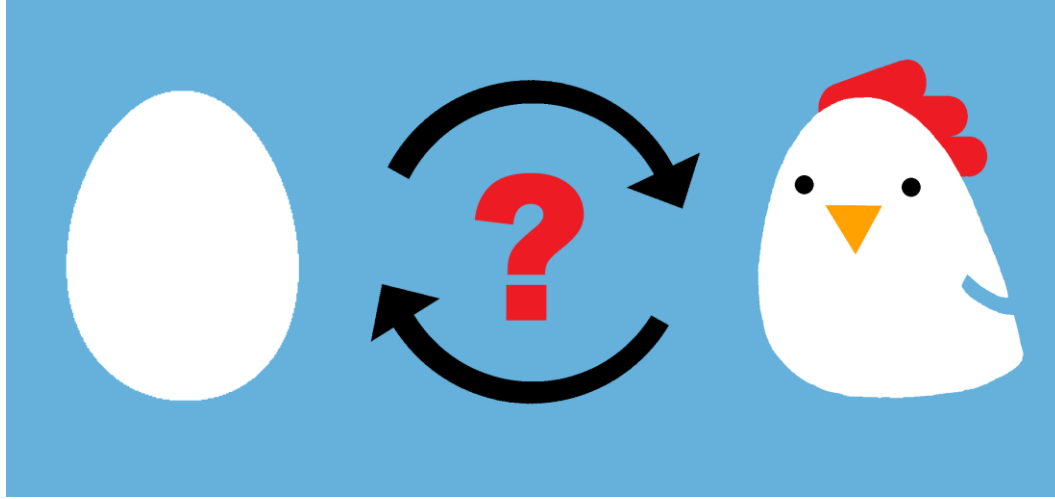


HYPOTÉZY



PROČ TO DĚLÁME?

***DVOUROZMĚRNÁ ANALÝZA JE VSTUPEM NA POLE
KAUZÁLNÍCH ÚSUDKŮ***

***KAŽDÁ VĚDA ASPIRUJE NA VÝPOVĚDI O
KAUZALITĚ***

***HLEDÁME TEDY PŘÍČINY ČI DŮSLEDKY JEVŮ,
KTERÉ STUDUJEME***

PŘIPOMENUTÍ:

KRITÉRIA KAUZÁLNÍHO USUZOVÁNÍ:

- 1) SOUBĚŽNÉ ZMĚNY V PROMĚNNÝCH
- 2) ČASOVÁ NÁSLEDNOST PŘÍČINY ADŮSLEDKU
- 3) VYLOUČENÍ VNĚJŠÍHO (JINÉHO VLIVU)

TEDY NAPŘ:

STUDUJI VLIV ROZVODU RODIČŮ NA ŠKOLNÍ PROSPĚCH DĚTÍ

MĚNÍ SE STRUKTURA RODINY -> MĚNÍ SE PROSPĚCH

(TO JSME SCHOPNI CELKEM SNADNO STATISTICKY MĚŘIT)

NEJPRVE ROZVOD RODIČŮ, POTOM ZMĚNY PROSPĚCHU

(ZDE UŽ BYCHOM POTŘEBOVALI DLOUHODOBÁ DATA – OBVYKLE NEMÁME)

ROZVOD RODIČŮ JE HLAVNÍM FAKTOREM ZMĚNY PROSPĚCHU

(ZDE UŽ JE TO VELMI SLOŽITÉ... OBVYKLE ALESPON NĚKOLIK KONTROLNÍCH PROMĚNNÝCH))



Jak testovat hypotézy na datech

VŠEOBECNĚ PLATNÝ ZÁKON GIGO: garbage in garbage out

Problém nejasnosti testování hypotéz často netkví v tom, že neznáme nějakou zázračnou techniku analýzy, ale v celém řetězci výzkumný problém – analýza.

Důležité je kladení korektních hypotéz, z nichž lze mnoho vytěžit . Hypotéza musí nabízet co nejširší pohyb mezi úrovněmi abstrakce od teorie až k indikátorům. Musíme být schopni formulovat vějíř hypotéz na různé úrovni abstrakce a uvědomovat si hierarchický vztah mezi nimi. (pokud nelze, bude někde chyba)

Princip falzifikace v testování hypotéz

Formulace nulových hypotéz:

Nulová hypotéza – neexistence vztahu/rozdílu, nevhodnost modelu (negace našeho předpokladu – snaha vyvrátit jej)

Alternativní hypotéza – vztah/rozdíl existuje, model je vhodný

-> falzifikovatelnost (alespoň v principu) – hlavní kritérium použitelnosti hypotéz

Do statistické analýzy vstupujeme vždy s nulovou hypotézou
(netřeba ji však formulovat přímo v textu, je to samozřejmé)

Neexistuje rozdíl v průměrech -> t-test

Neexistuje rozdíl v distribucích -> chí-kvadrát

Model není vhodný -> F-test, Anova, chí-kvadrát (podle povahy modelu)

Dvě roviny testování hypotéz

Na úrovni výzkumného vzorku:

Vyvrácení H_0 na základě rozdílu

Věcný význam velikosti rozdílu

-> vyhněme se mechanickému zamítání/přijímání hypotéz na základě testů

Na úrovni cílové populace (statistická inference):

Vyvrácení H_0 na základě statistických testů

Statistická významnost

(Ize rozdíl dané velikosti s danou pravděpodobností nalézt v cílové populaci, nebo vznikl náhodným rozptylem – výběrovou chybou?)

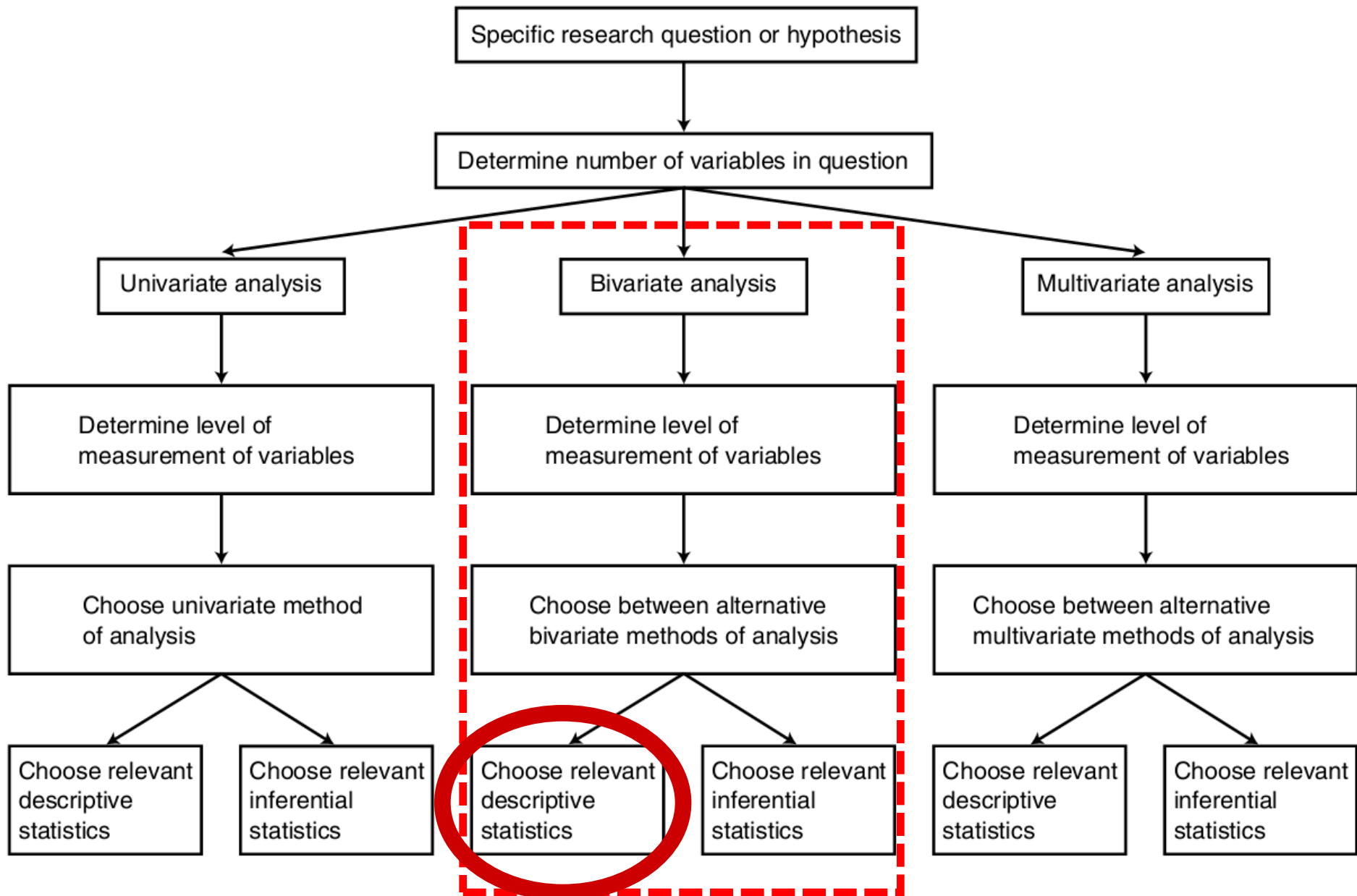
Statistická významnost nevypovídá vždy o věcné významnosti – Je závislá na velikosti vzorku

**DVOUROZMĚRNÁ A VÍCEROZMĚRNÁ ANALÝZA JE NÁSTROJEM,
JEHOŽ PROSTŘEDNICTVÍM SE KVANTITATIVNÍ VÝZKUM SNAŽÍ
POUKAZOVAT NA KAUZÁLNÍ VZTAHY**

**ČINÍ TAK VŽDY NA ZÁKLADĚ SROVNÁNÍ VARIABILITY
PROMĚNNÝCH**

**TEDY SROVNÁVÁ, ZDA ROZLOŽENÍ HODNOT JEDNÉ PROMĚNNÉ
JE ODLIŠNÉ, POKUD JE TŘÍDÍME PODLE DRUHÉ PROMĚNNÉ**

- > ODLIŠNÝ PRŮMĚR**
- > ODLIŠNÝ ROZPTYL**
- > ODLIŠNÉ ROZLOŽENÍ HODNOT**



DVOUROZMĚRNÁ ANALÝZA

**KATEGORIE
VS
KONTINUUM**

**KATEGORIE
VS
KATEGORIE**

**KONTINUUM
VS
KONTINUUM**

**SROVNÁVÁNÍ
PRŮMĚRŮ**

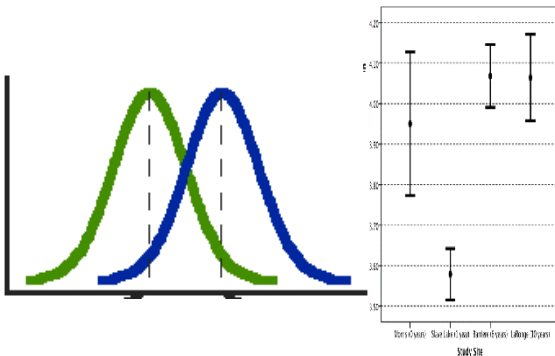
(např. příjem podle
pohlaví)

**KONTINGENČNÍ
TABULKA**

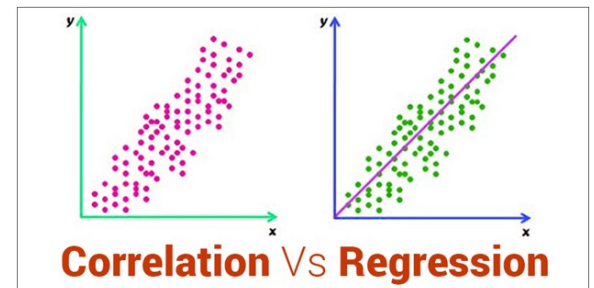
(např. vzdělanostní
kategorie podle
pohlaví)

**KORELACE,
REGRESE**

(např. příjem podle
věku)



| | Dog | Cat | Total |
|--------|-----|-----|-------|
| Male | 42 | 10 | 52 |
| Female | 9 | 39 | 48 |
| Total | 51 | 49 | 100 |



Correlation Vs Regression

A KONKRÉTNÍ VOLBA PROCEDUR V SPSS...

DVOUROZMĚNÁ
ANALÝZA

KATEGORIE
VS
KONTINUUM

KATEGORIE
VS
KATEGORIE

KONTINUUM
VS
KONTINUUM

SROVNÁVÁNÍ
PRŮMĚRŮ
ANALYZE ->
COMPARE MEANS ->
MEANS
ANALYZE ->
DESCRIPTIVE STAT.->
EXPLORE

KONTINGENČNÍ
TABULKA
ANALYZE ->
DESCRIPTIVE STAT.->
CROSSTABS

KORELACE,
REGRESE
ANALYZE ->
CORRELATE ->
BIVARIATE
REGRESION ->
LINEAR

DVOUROZMĚRNÁ ANALÝZA

**KATEGORIE
VS
KONTINUUM**

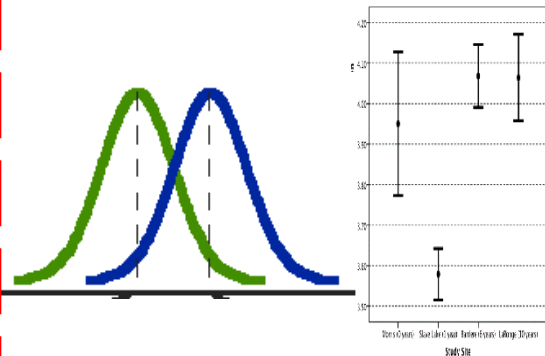
**KATEGORIE
VS
KATEGORIE**

**KONTINUUM
VS
KONTINUUM**

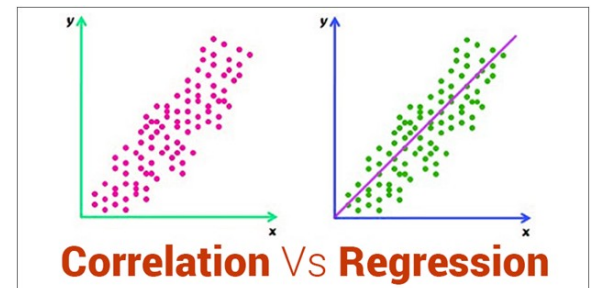
**SROVNÁVÁNÍ
PRŮMĚRŮ**
(např. příjem podle
pohlaví)

**KONTINGENČNÍ
TABULKA**
(např. vzdělanostní
kategorie podle
pohlaví)

**KORELACE,
REGRESE**
(např. příjem podle
věku)



| | Dog | Cat | Total |
|--------|-----|-----|-------|
| Male | 42 | 10 | 52 |
| Female | 9 | 39 | 48 |
| Total | 51 | 49 | 100 |



Correlation Vs Regression



... TAKŽE KONKRÉTNĚ BYCHOM MOHLI POSTUPOVAT NAPŘÍKLAD TAKTO:

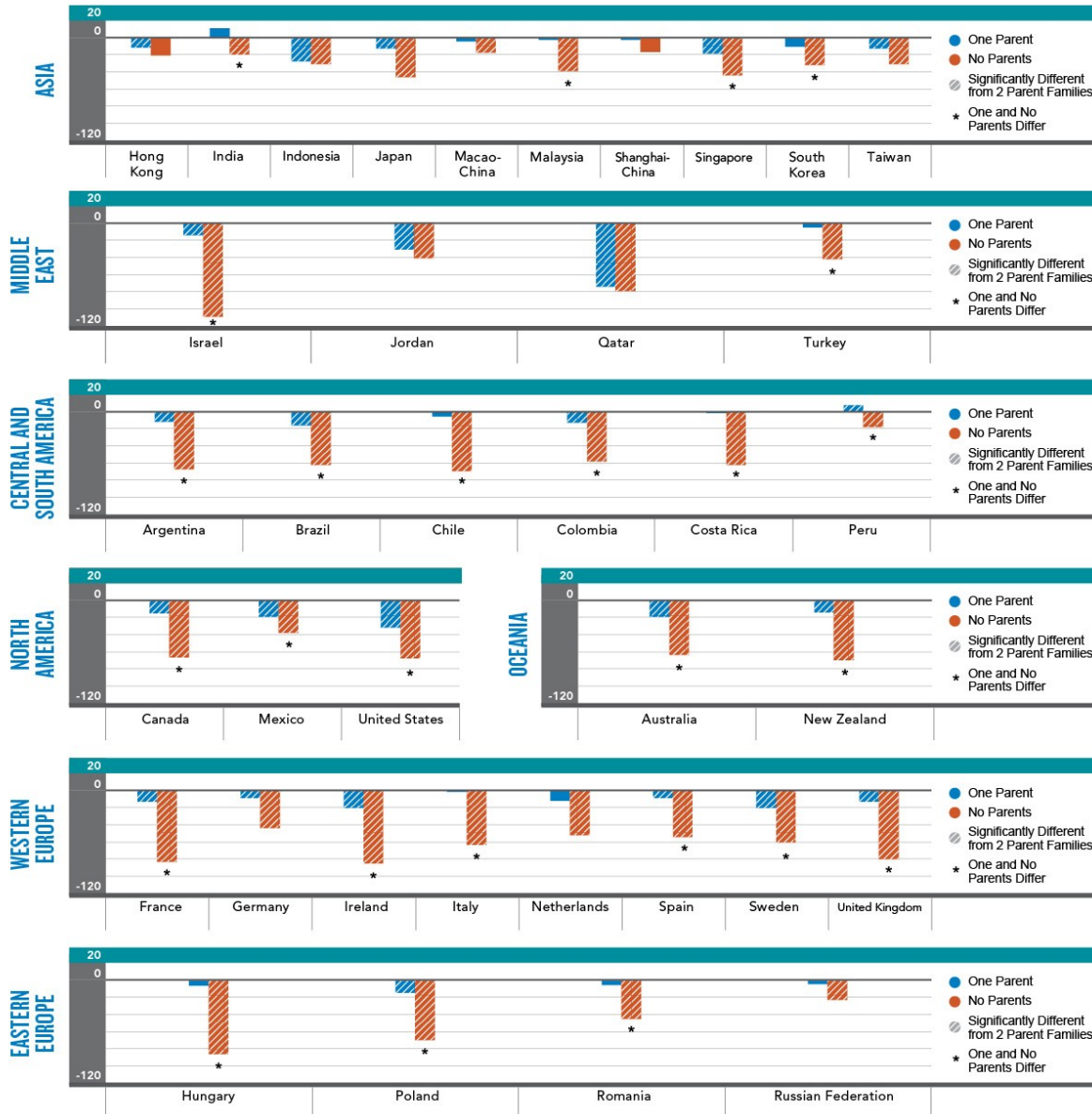
X = NEZÁVISLE PROMĚNNÁ = ROZVEDLI SE RODIČE? (2 kategorie)

Y = ZÁVISLE PROMĚNNÁ = VÝSLEDEK V TESTU PISA (škála)

(obvykle bychom do výzkumu zavedli také kontrolní proměnné např. dobu uplynulou od rozvodu rodičů, s kým nyní žije, nové partnerství rodiče, ekonomickou situaci domácnosti... ale o tom až v lekci „elaborace“)

Figure 15 DIFFERENCES IN READING LITERACY BY NUMBER OF PARENTS, PISA 2009

DIFFERENCES IN READING LITERACY OF STUDENTS, ACCOUNTING FOR BACKGROUND FACTORS:
ONE AND NO PARENTS COMPARED TO TWO

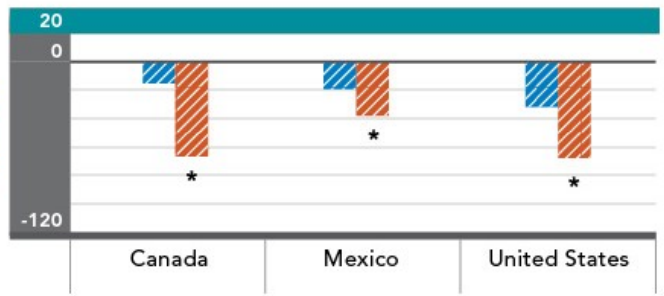


SROVNÁNÍ NA ZÁKLADĚ PRŮMĚRŮ

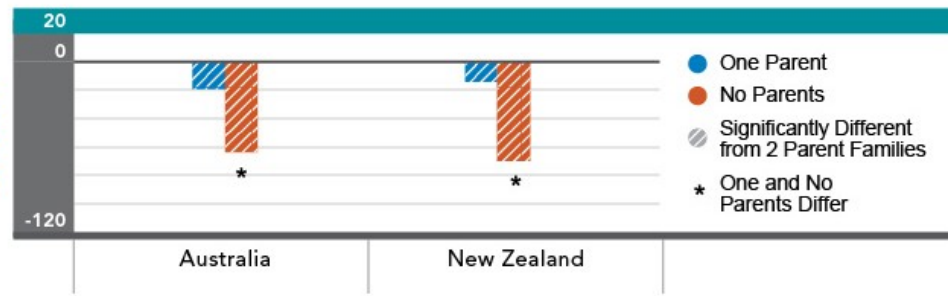
REFERENČNÍ ÚROVEŇ = RODINY SE DVĚMA RODIČI = V MĚŘÍTKU ŠKÁLY „0“

*Převzato z:
<http://worldfamilymap.ifstudies.org/2013/articles/essay/two-one-or-no-parents>*

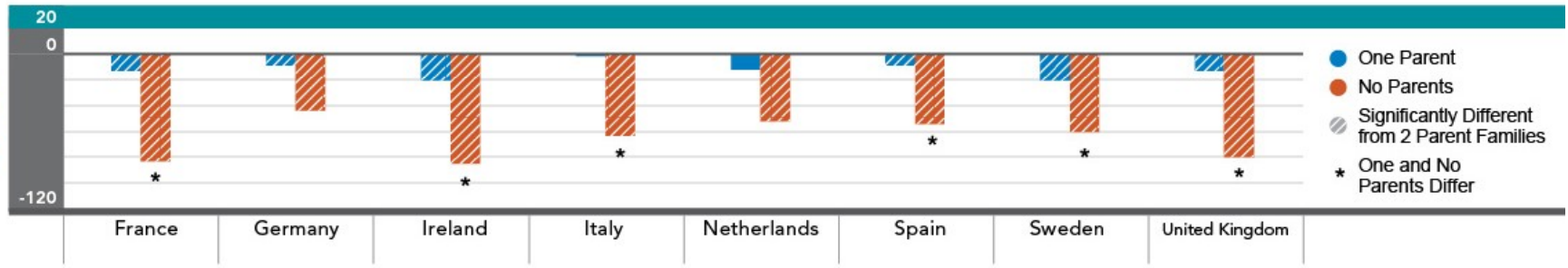
NORTH AMERICA



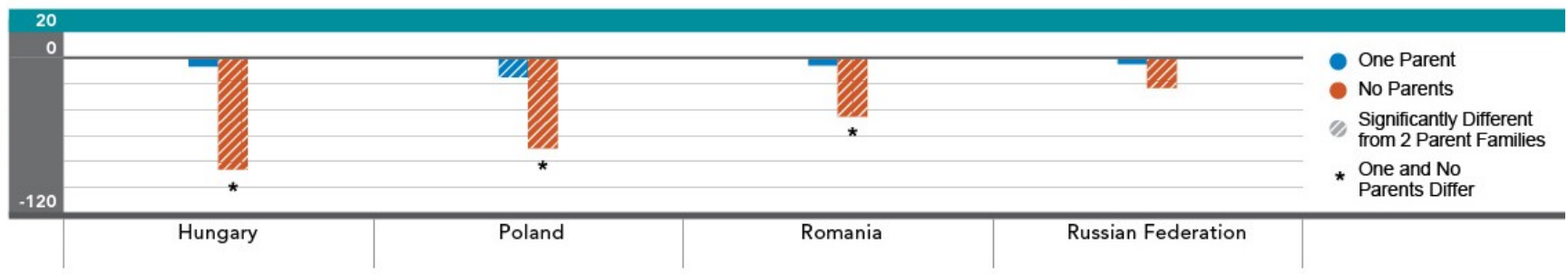
OCEANIA



WESTERN EUROPE

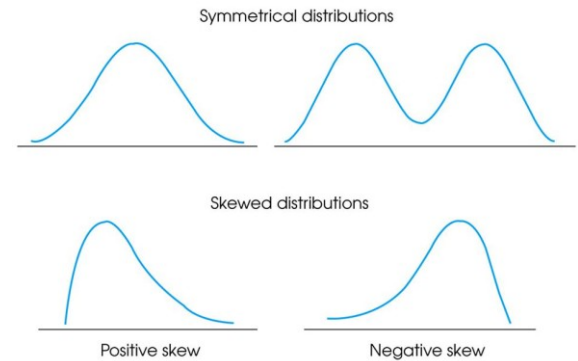
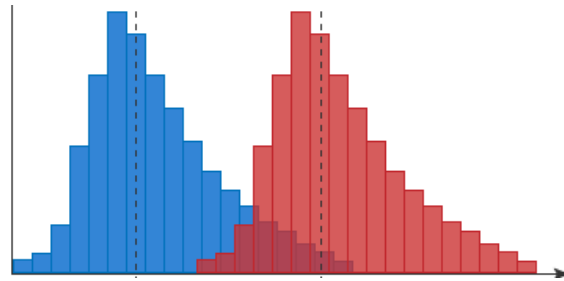
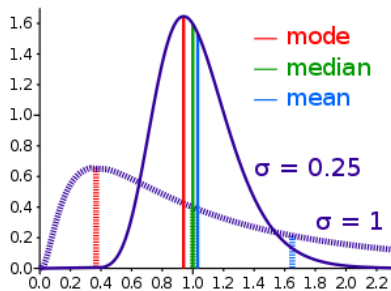


EASTERN EUROPE

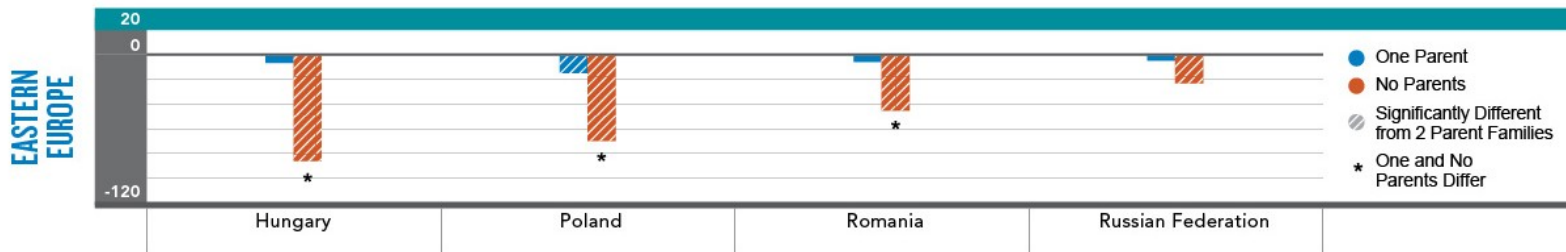


Pozor ale na důležité okolnosti:

A) Věnujme pozornost rozložení závisle proměnné v daných skupinách



B) Zvažme, jak veliký rozdíl v průměrech je věcně důležitý

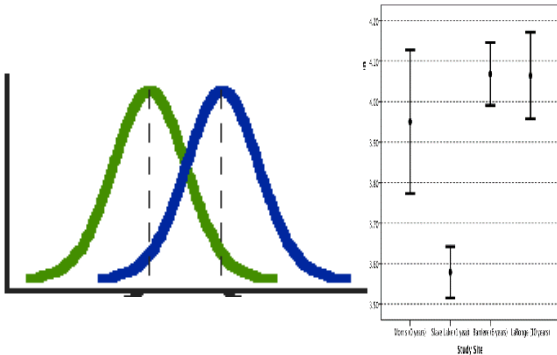


Např. 100bodová škála, rozdíl 5 bodů vs. 10bodová škála, rozdíl 5 bodů

DVOUROZMĚRNÁ ANALÝZA

**KATEGORIE
VS
KONTINUUM**

**SROVNÁVÁNÍ
PRŮMĚRŮ**
(např. příjem podle
pohlaví)



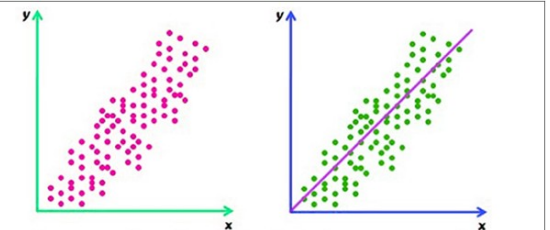
**KATEGORIE
VS
KATEGORIE**

**KONTINGENČNÍ
TABULKA**
(např. vzdělanostní
kategorie podle
pohlaví)

| | Dog | Cat | Total |
|--------|-----|-----|-------|
| Male | 42 | 10 | 52 |
| Female | 9 | 39 | 48 |
| Total | 51 | 49 | 100 |

**KONTINUUM
VS
KONTINUUM**

**KORELACE,
REGRESE**
(např. příjem podle
věku)



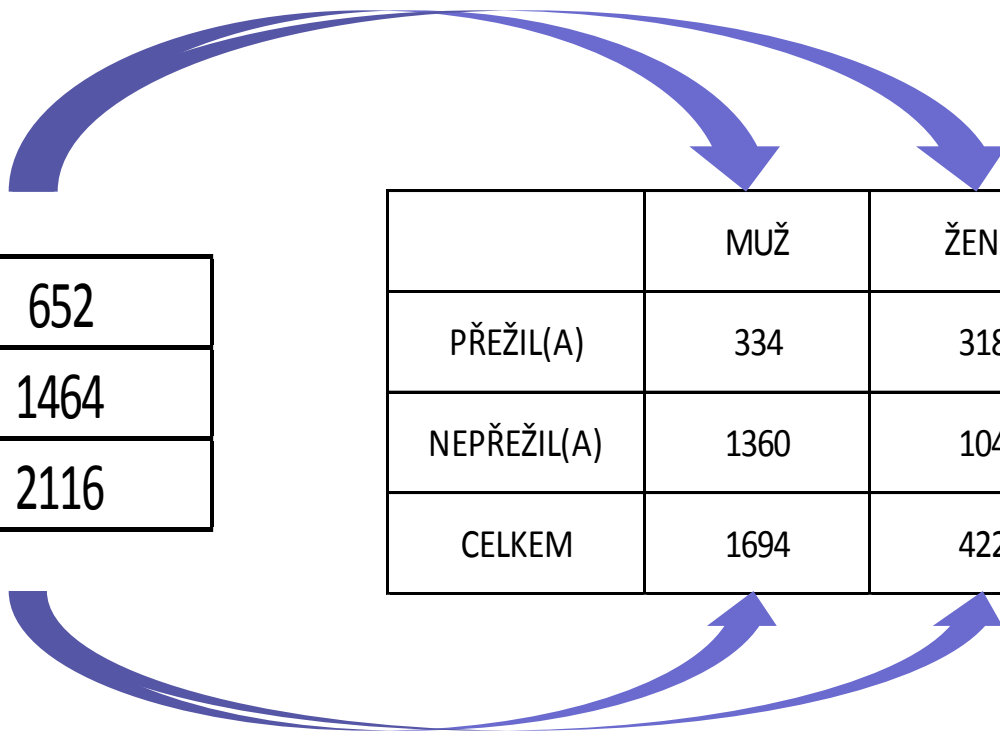
Correlation Vs Regression



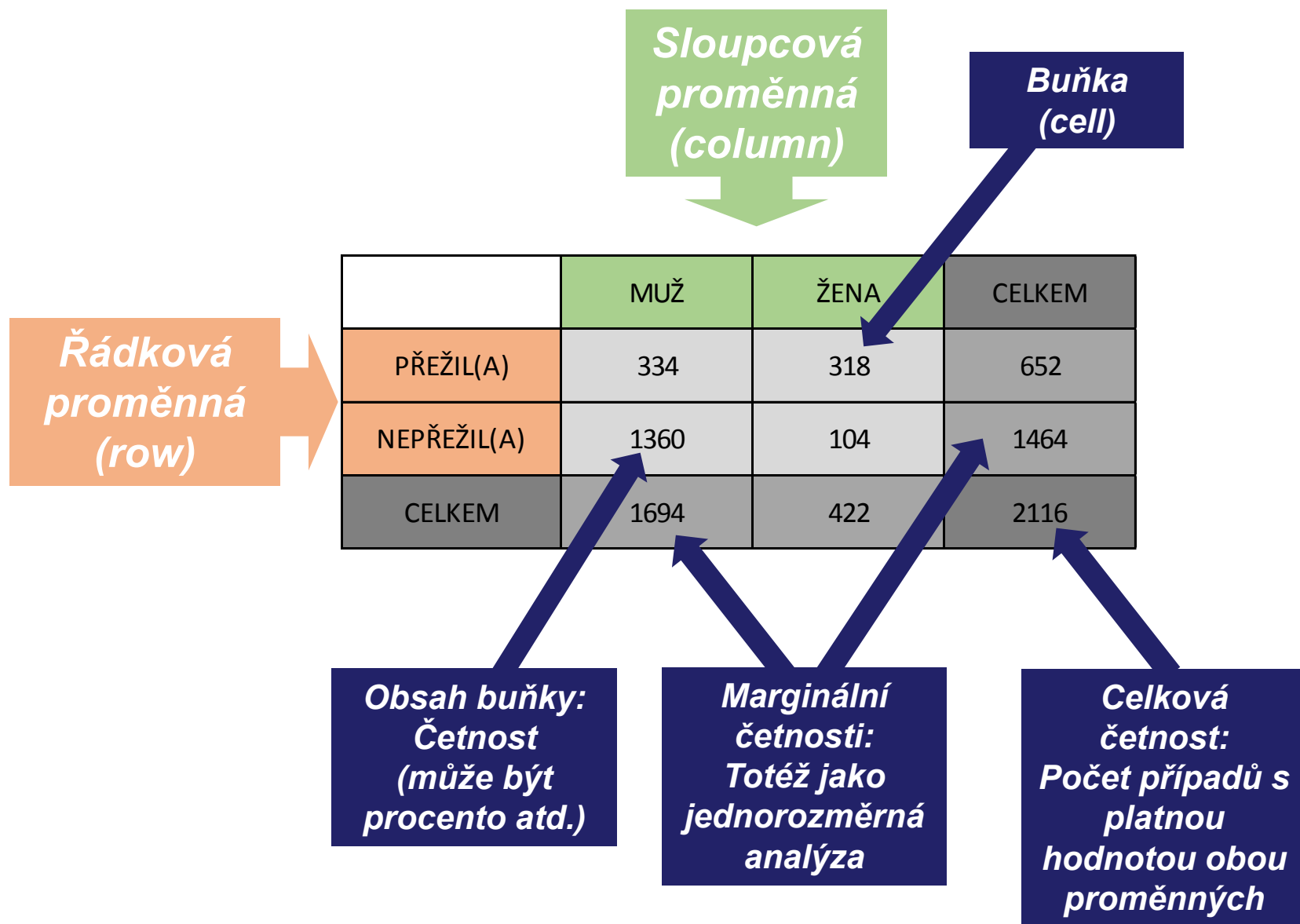
**KONTINGENČNÍ TABULKA:
ROZDĚLENÍ ČETNOSTÍ JEDNOTLIVÝCH KATEGORIÍ PODLE DALŠÍ PROMĚNNÉ**

| | |
|-------------|------|
| PŘEŽIL(A) | 652 |
| NEPŘEŽIL(A) | 1464 |
| CELKEM | 2116 |

| | MUŽ | ŽENA | CELKEM |
|-------------|------|------|--------|
| PŘEŽIL(A) | 334 | 318 | 652 |
| NEPŘEŽIL(A) | 1360 | 104 | 1464 |
| CELKEM | 1694 | 422 | 2116 |



Terminologie:



TYPY RELATIVNÍCH ČETNOSTÍ (PROCENTA)

SLOUPCOVÁ PROCENTA

*Jaký podíl z celku mužů přežil?
Jaký podíl z celku žen přežil?*

...

| | MUŽ | ŽENA | CELKEM |
|-------------|--------|--------|--------|
| PŘEŽIL(A) | 19,7% | 75,4% | 30,8% |
| NEPŘEŽIL(A) | 80,3% | 24,6% | 69,2% |
| CELKEM | 100,0% | 100,0% | 100,0% |

ŘÁDKOVÁ PROCENTA

*Jaký podíl z přeživších byli muži?
Jaký podíl z přeživších byly ženy?*

...

| | MUŽ | ŽENA | CELKEM |
|-------------|-------|-------|--------|
| PŘEŽIL(A) | 51,2% | 48,8% | 100,0% |
| NEPŘEŽIL(A) | 92,9% | 7,1% | 100,0% |
| CELKEM | 80,1% | 19,9% | 100,0% |

CELKOVÁ PROCENTA

*Jaký podíl z celku tvoří muži, kteří přežili?
Jaký podíl z celku tvoří ženy, které přežily?*

...

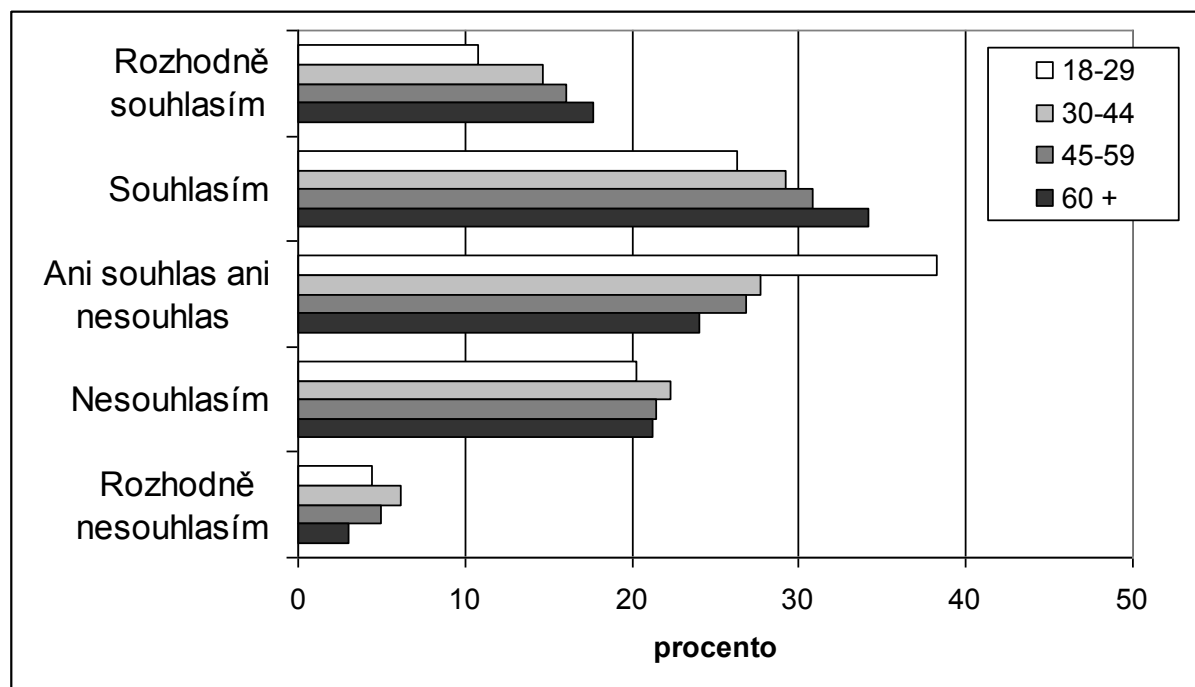
| | MUŽ | ŽENA | CELKEM |
|-------------|-------|-------|--------|
| PŘEŽIL(A) | 15,8% | 15,0% | 30,8% |
| NEPŘEŽIL(A) | 64,3% | 4,9% | 69,2% |
| CELKEM | 80,1% | 19,9% | 100,0% |

V České republice je dnes příliš mnoho přistěhovalců.

Věková kategorie

| | 18-29 | 30-44 | 45-59 | 60 + | Celkem |
|---------------------------|-------|-------|-------|-------|--------|
| Rozhodně souhlasím | 10,7 | 14,7 | 16,1 | 17,6 | 14,8 |
| Souhlasím | 26,3 | 29,2 | 30,8 | 34,1 | 30,1 |
| Ani souhlas ani nesouhlas | 38,2 | 27,7 | 26,8 | 24,0 | 29,2 |
| Nesouhlasím | 20,3 | 22,3 | 21,4 | 21,2 | 21,3 |
| Rozhodně nesouhlasím | 4,4 | 6,1 | 4,9 | 3,1 | 4,6 |
| Celkem | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 |

KONTINGENČNÍ TABULKA A GRAF



DVOUROZMĚRNÁ ANALÝZA

**KATEGORIE
VS
KONTINUUM**

**KATEGORIE
VS
KATEGORIE**

**KONTINUUM
VS
KONTINUUM**

**SROVNÁVÁNÍ
PRŮMĚRŮ**

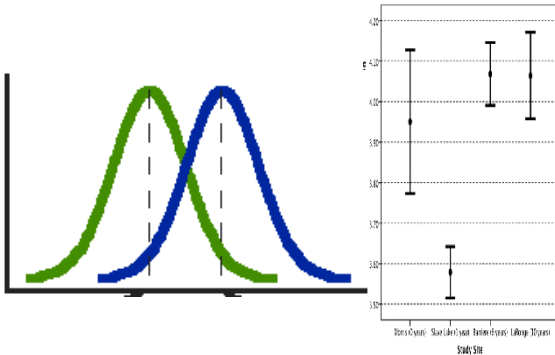
(např. příjem podle
pohlaví)

**KONTINGENČNÍ
TABULKA**

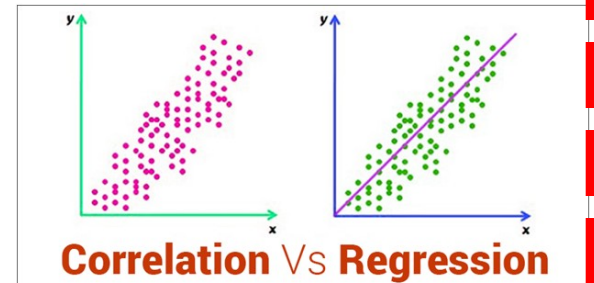
(např. vzdělanostní
kategorie podle
pohlaví)

**KORELACE,
REGRESE**

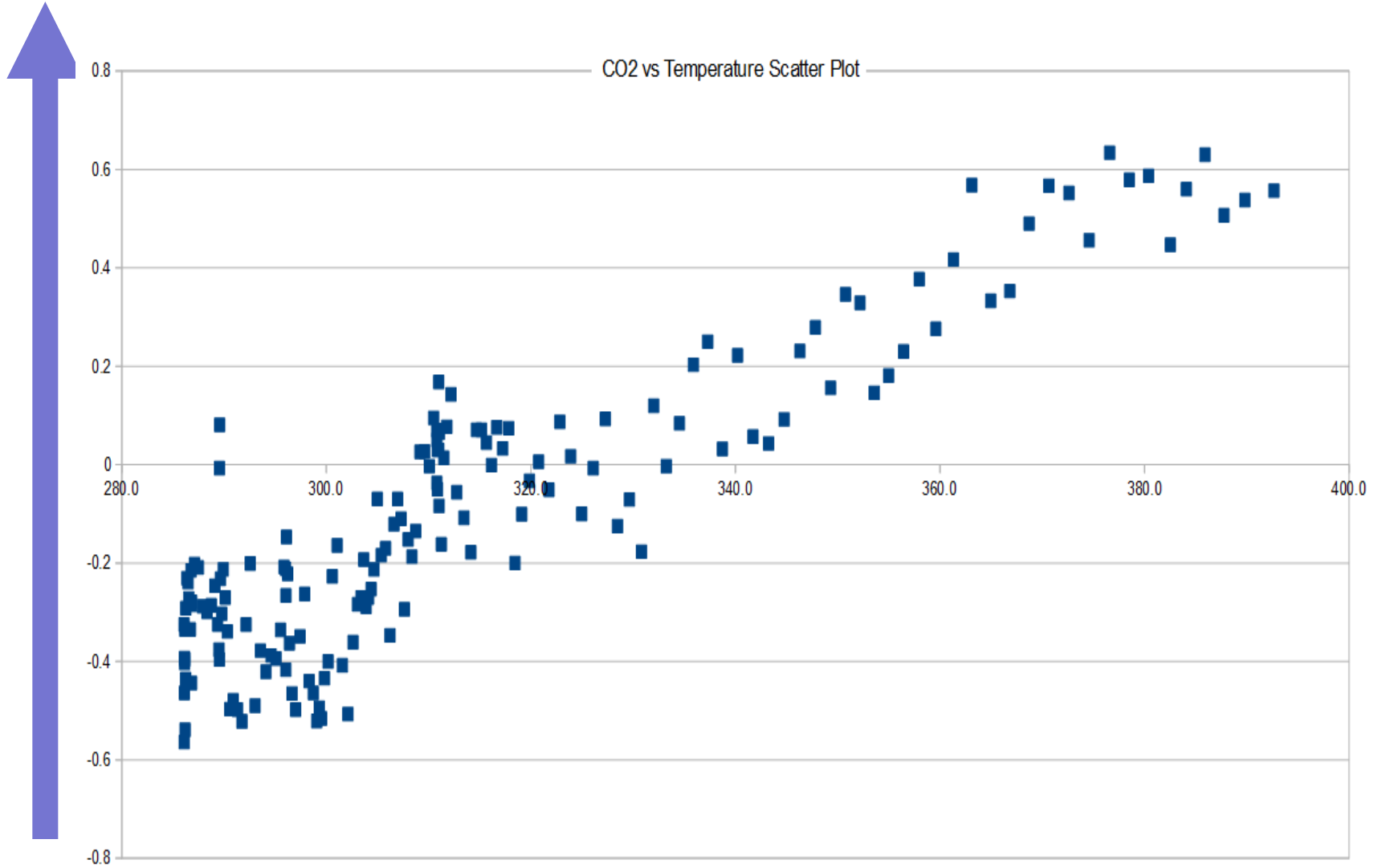
(např. příjem podle
věku)



| | Dog | Cat | Total |
|--------|-----|-----|-------|
| Male | 42 | 10 | 52 |
| Female | 9 | 39 | 48 |
| Total | 51 | 49 | 100 |

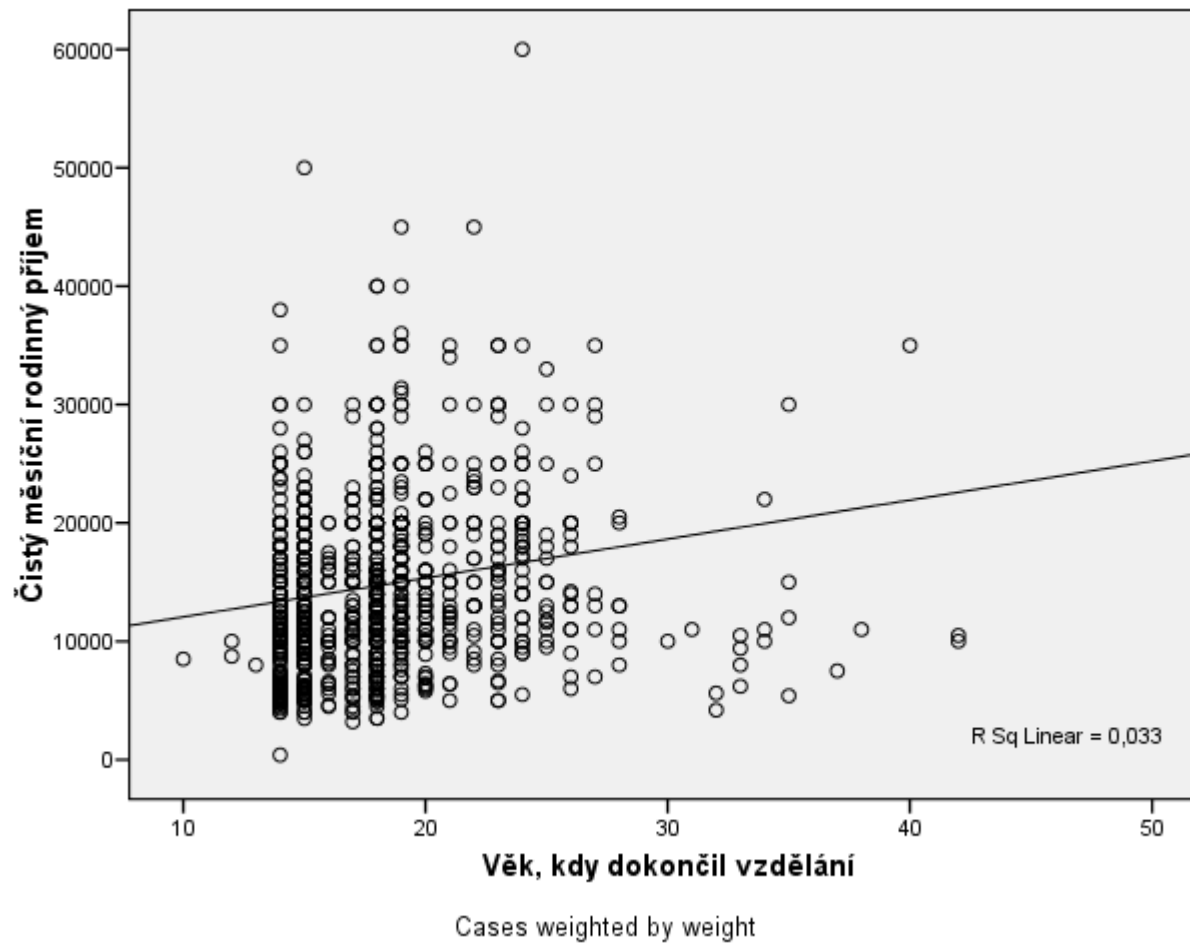


Osa Y

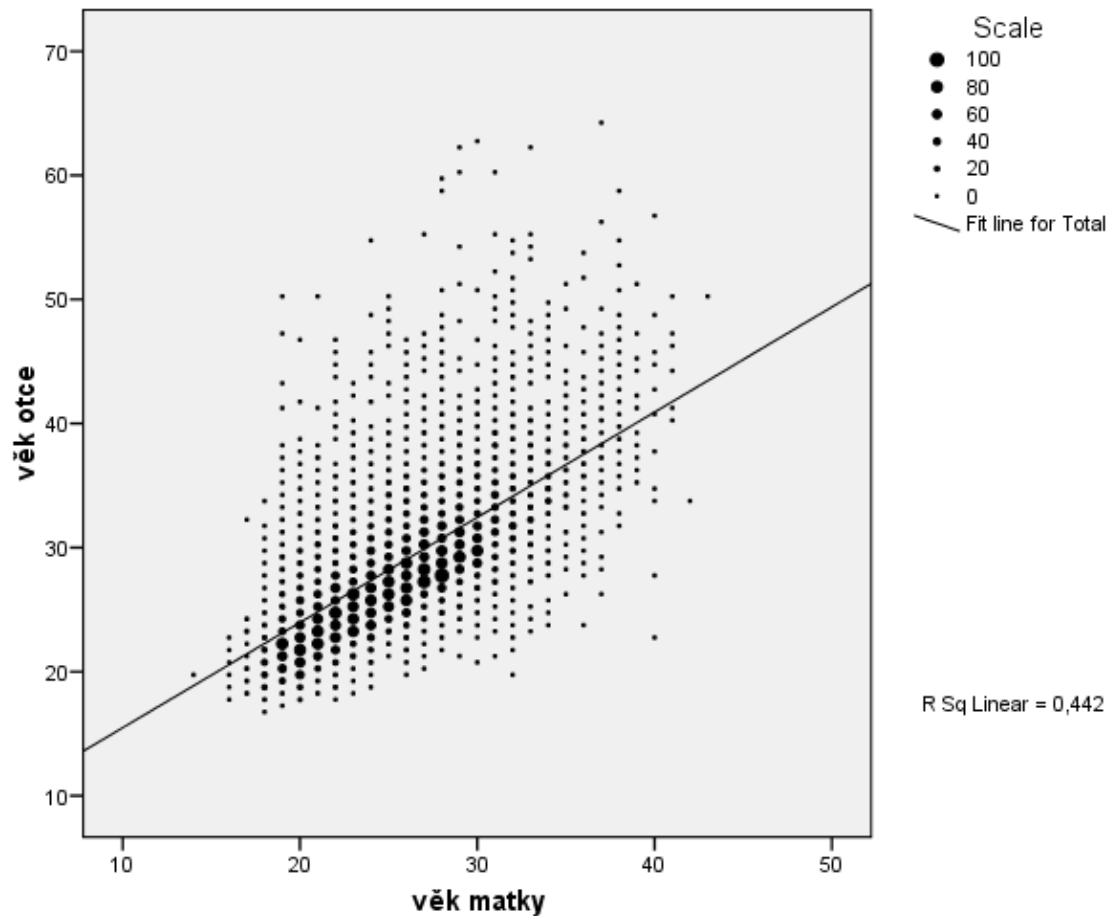


Osa X

KORELACE, REGRESE – SLABÁ SOUVISLOST



KORELACE, REGRESE – SILNÁ SOUVISLOST



KORELAČNÍ KOEFICIENT UKAZUJE, JAK SE DATA PŘIMYKAJÍ K PŘÍMCE (TĚSNOST VZTAHU)
REGRESE TUTO PŘÍMKU POPUSUJE (VELIKOST EFEKTU)

ELABORACE

(prostředky běžných analýz – bez vícerozměrných modelů)

KATEGORIE

VS

KATEGORIE

KATEGORIE

VS

KONTINUUM

KONTINUUM

VS

KONTINUUM

*KONTINGENČNÍ
TABULKA*

*SROVNÁVÁNÍ
PRŮMĚRŮ*

KORELACE

*PŘIDAT DALŠÍ
VRSTVY TABULKY
(např. vzdělanostní
kategorie podle
pohlaví a věkové kat.)*

*PŘIDAT DALŠÍ
VRSTVU TŘÍDĚNÍ
PRŮMĚRŮ
(např. příjem podle
pohlaví a vzdělání)*

*PARCIÁLNÍ
KORELACE
(např. příjem podle
věku a délky praxe v
dané pozici)*

Testování modelů

Z matematického hlediska je model rovnice, která vystihuje určitý princip v realitě a umožňuje ho popsat a predikovat výsledky

Např. vztah mezi léty vzdělání a dosaženým příjmem

Obecná rovnice regresního modelu

$$Y = a + bX$$

a = konstanta

b = směrnice – regresní koeficient

X, Y = nezávisle a závisle proměnná

Regresní rovnice může vypadat takto:

$$PLAT = 8772 + 329 \cdot \text{věk, kdy dokončil vzdělání}$$