

# Experimentálna ekonómia

## Prednáška 7: Štatistická inferencia

Matej Lorko

[matej.lorko@euba.sk](mailto:matej.lorko@euba.sk)

Materiály: [www.lorko.sk/lectures](http://www.lorko.sk/lectures)

Referencie:

- Weimann, J., & Brosig-Koch, J. (2019). *Methods in experimental economics*. Springer International Publishing. Chicago
- Jacquemet, N., & l'Haridon, O. (2018). *Experimental economics*. Cambridge University Press.
- Rajčáni, J., Kačmár, P., Bavolar, J., Cavojova, V., Adamkovič, M., Vargová, L., Kočišová, L., Martončík, M. (2024). *Štatistika pre reprodukovateľný výskum v spoločenských vedách*.

# Štatistické testy

- V každodennom živote sa až príliš často stretávame s tým, že vyvodzujeme úplne nevedecké a neplatné závery, ako napríklad „Môj priateľ bol raz okradnutý v meste A, takže je to zločinecké mesto“ alebo „Bezpečnostný pás nie je potrebný. Koniec koncov, nikdy som nemal nehodu.“
- Aj bez formálnej analýzy si môžeme byť celkom istí, že tieto závery príliš zovšeobecňujú, keďže sú založené len na jednom pozorovaní. Ako však možno urobiť konkrétne tvrdenia o kvalite záveru? Ako si môže byť experimentátor istý, že pozorovaný efekt nie je úplne náhodný?
- V takýchto situáciách nám pomáhajú nástroje z inferenčnej štatistiky. Dôraz je kladený na testovanie štatistických hypotéz. Takýmito testami možno určiť, nakoľko je všeobecné tvrdenie o charakteristikách populácie v súlade s pozorovanými laboratórnymi údajmi alebo so vzorkou.

# Formulácia testovateľných hypotéz

- Východiskovým bodom štatistického testovania je výskumná hypotéza. Zvyčajne postuluje obsah výskumnej otázky. V podstate to, čo považujeme za pravdivé, je formulované ako nulová hypotéza  $H_0$  a jej opak alebo doplnok ako alternatívna hypotéza  $H_1$ .
- Tento princíp štatistického testovania je porovnateľný s prezumpciou nevinu v súdnom spore. Počiatočná alebo nulová hypotéza je: "Obžalovaný je nevinný." Namiesto priameho preukázania, že obžalovaný je vinný, prokurátor predkladá viac či menej silné dôkazy, ktoré nie sú v súlade s nevinou obžalovaného. Ak sú tieto dôkazy dostatočne silné, prezumpcia nevinu už neplatí a obžalovaný je uznaný vinným. Ak však nie je možné predložiť dostatočne silné dôkazy proti prezumpcii nevinu, obžalovaný nebude uznaný vinným, pretože jeho predtým predpokladanú nevinu nebolo možné spochybniť bez akýchkoľvek pochybností.
- Predpokladá sa, že nulová hypotéza je pravdivá, až pokiaľ zhromaždené dáta nebudú dostatočne silné proti nej, a vtedy musí byť zamietnutá. Akonáhle sa tak stane, alternatívna hypotéza je nepriamo prijatá. Ak však údaje nedokážu vyvrátiť nulovú hypotézu, treba stále predpokladať, že je pravdivá a výskumná hypotéza nie je akceptovaná. Keďže v teste hypotéz sa testuje iba nulová hypotéza a hľadajú sa dôkazy proti nej, nulová hypotéza môže byť iba zamietnutá alebo nezamietnutá, ale nikdy nie akceptovaná.

# Inferenčná štatistika

- Ak je výskumná otázka formulovaná vo forme štatistickej hypotézy, môžeme ju testovať. Vždy však existuje určitá pravdepodobnosť chýb.
- Vždy treba mať na pamäti, že žiadny štatistický test nedokáže určiť, či je hypotéza skutočne pravdivá alebo nepravdivá. Aj keď je testovacia štatistika vzorky v kritickej oblasti a dospejeme k záveru, že nulovú hypotézu treba zamietnuť, stále môže byť pravdivá.
- Čím väčšiu kritickú oblasť alebo hladinu významnosti zvolíme, tým pravdepodobnejšia je táto takzvaná chyba I. typu. Avšak, ak je nulová hypotéza v skutočnosti nepravdivá, môžeme urobiť chybu v prípade, že ju neodmietneme (chyba typu II).

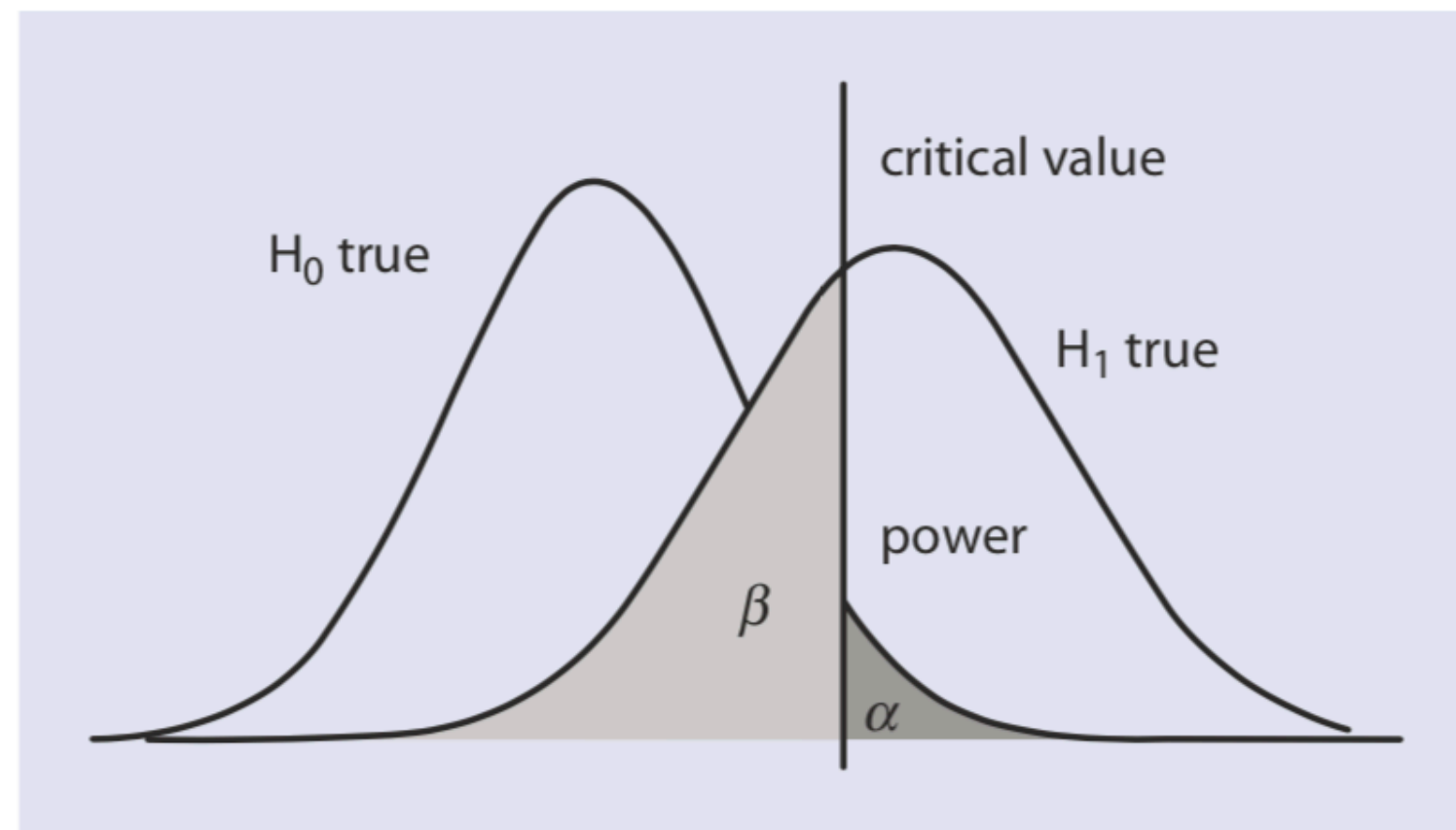
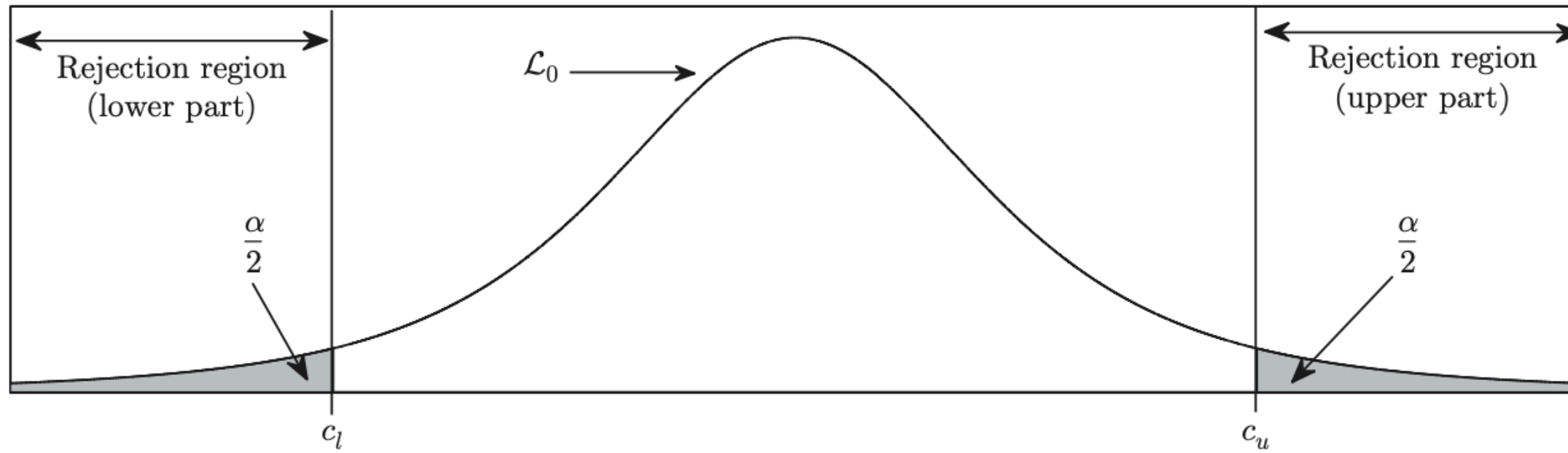


Table 4.1 Summary of error probabilities

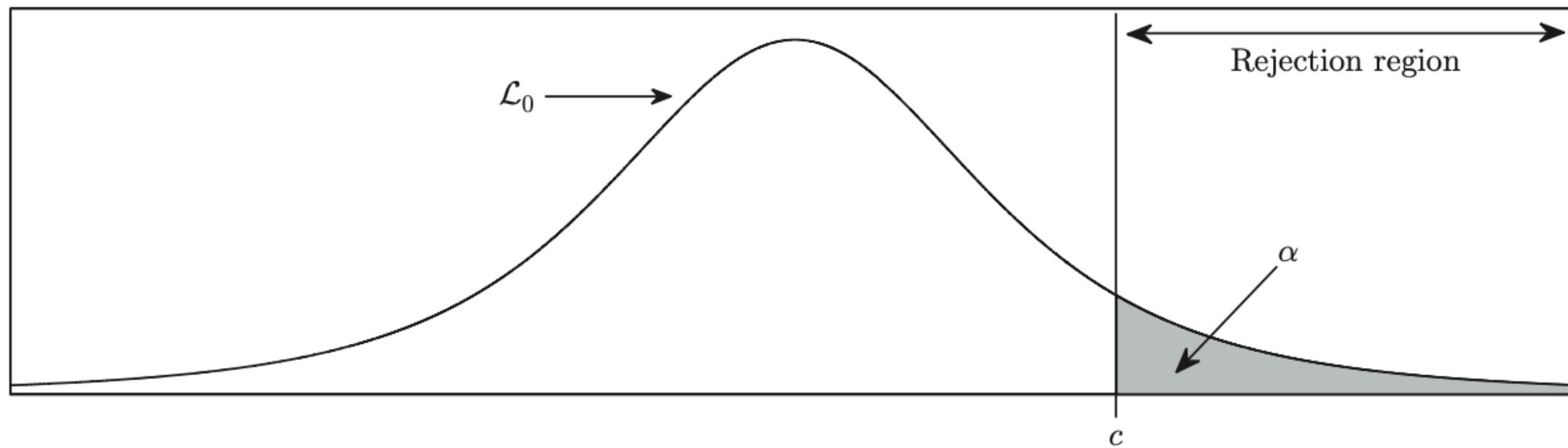
	Truth	
	H <sub>0</sub> true	H <sub>0</sub> not true
Rejection H <sub>0</sub>	Type I error (Prob. $\alpha$ )	correct
Non-rejection H <sub>0</sub>	correct (Prob. $1 - \alpha$ )	Type II error (Prob. $1 - \beta$ )

# Inferenční statistika

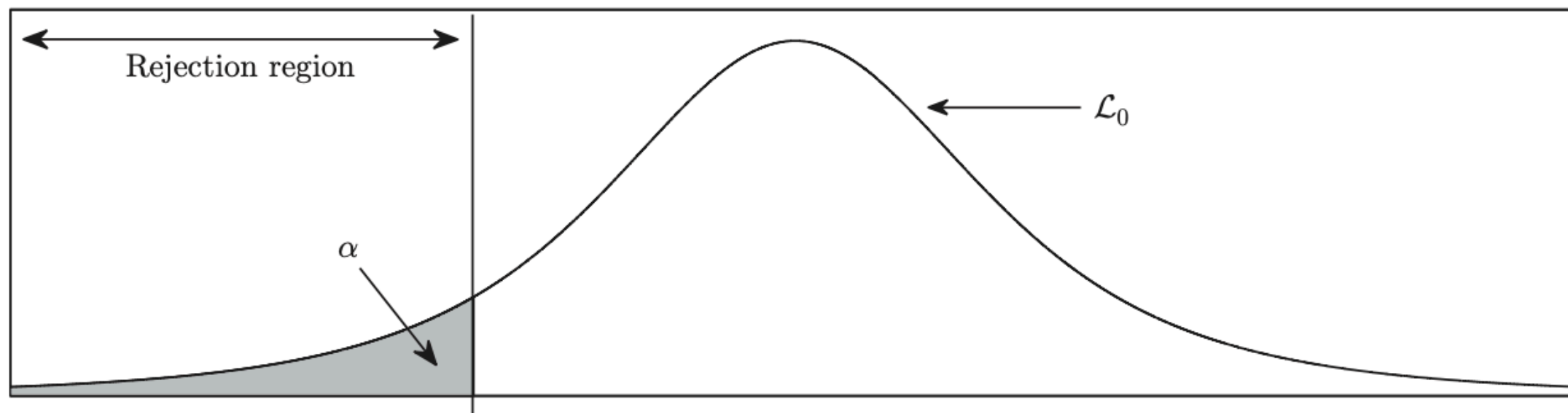
A bilateral rejection region



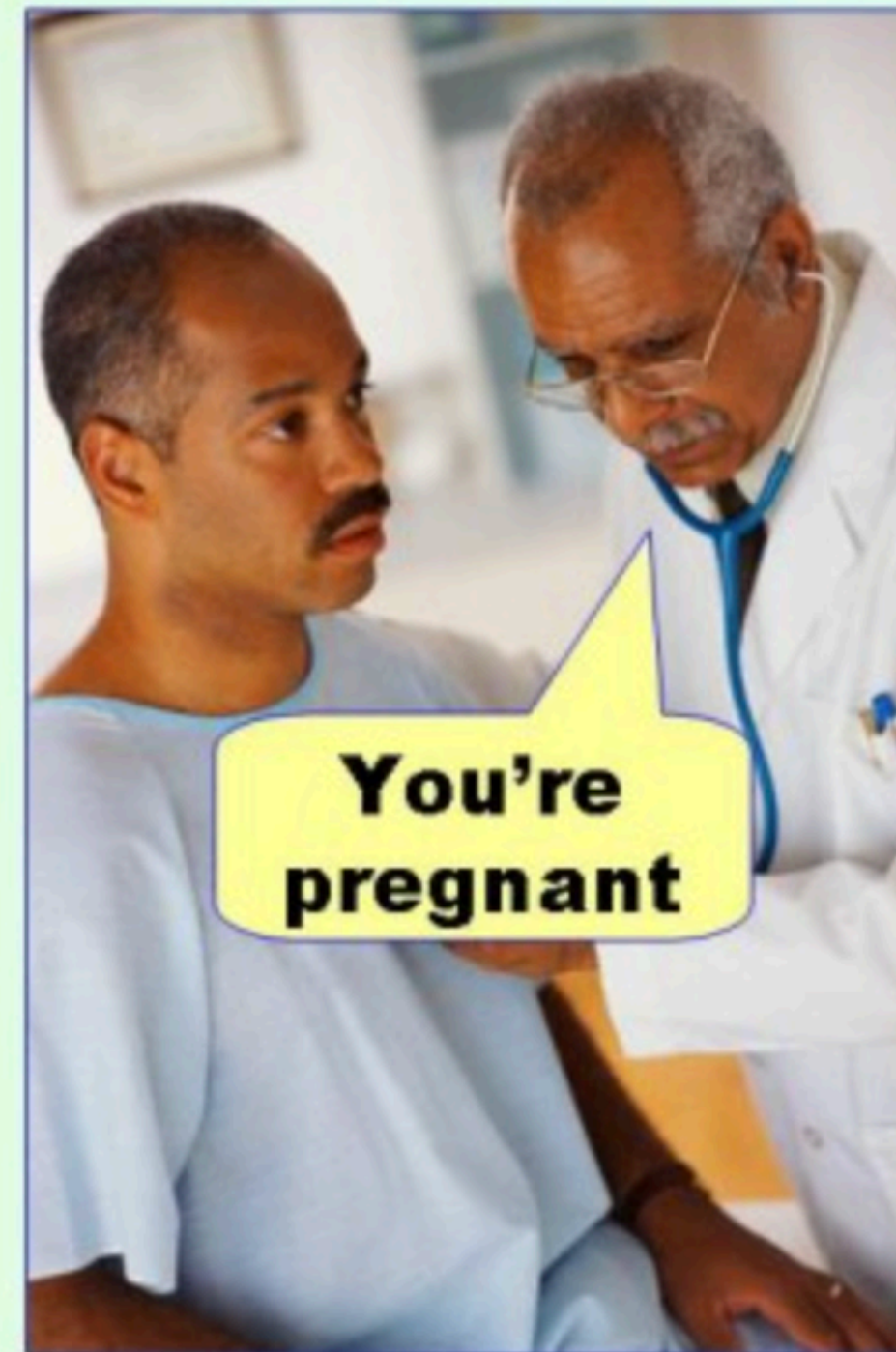
An upper rejection region



A lower rejection region



**Type I error**  
(false positive)



**Type II error**  
(false negative)



# Výber štatistického testu

- „Správna“ voľba metódy na analýzu experimentálnych dát vždy leží medzi dvoma extrémami. Jedným extrémom je úplne svojvoľné rozhodnutie použiť konkrétnu metódu analýzy, druhým extrémom je predpoklad, že existuje len jedna metóda analýzy, ktorá je dokonale vhodná pre daný experiment. Oba prístupy sú, samozrejme, rovnako nesprávne. Na jednej strane je určite možné a potrebné obmedziť počet metód, ktoré možno použiť. Všetky experimentálne dáta majú určité charakteristiky, ktoré vylučujú určité štatistické analýzy, pričom umožňujú vykonávať iné. Na druhej strane, experiment nie je nikdy taký špecifický, aby sa dala použiť iba jedna optimálna metóda analýzy.
- Základným prístupom k výberu vhodných metód štatistickej analýzy je predovšetkým zosúladenie formálnych požiadaviek na aplikáciu metódy s danými charakteristikami dát. Všetky metódy inferenčnej štatistiky, korelačnej analýzy a regresnej analýzy sú založené na určitých predpokladoch.
- Aby sme sa vyhli pri výbere metódy štatistickej analýzy závažným chybám, je dôležité o tom uvažovať už pri tvorbe experimentálneho dizajnu. Keď sú údaje k dispozícii a až potom sa zistí, že neexistuje vhodný postup na ich analýzu, zvyčajne je na opravy neskoro.
- Štatistická analýza údajov by mala byť v zásade vždy založená na odborných znalostiach a štatistická metóda by sa mala použiť len vtedy, ak výsledky môžu poskytnúť skutočný pohľad na experimentálne skúmanú výskumnú otázku. Malo by sa vyhnúť ad-hoc aplikácii metódy „iba kvôli metóde“, pretože štatistická analýza potom často míňa zmysel pôvodnej otázky.

# Klasifikácia testovacích metód

- Testy štatistických hypotéz možno kategorizovať pomocou niekoľkých kritérií. Jedným z najzákladnejších rozlišovacích znakov je počet skupín alebo vzoriek, ktoré test porovnáva. Ak sa skúma iba jedna skupina, je možné napríklad otestovať, či jej priemer zodpovedá určitému parametru populácie, o ktorom sa predpokladá, že je pravdivý. Týmto spôsobom sa vykoná porovnanie medzi špecifickou vzorkou a predpokladanou skutočnou hodnotou populácie pomocou jednovýberových testov.
- Keď sa pozrieme na výskum v experimentálnej ekonómii, psychológii a iných spoločenských vedách, porovnania skupín patria k najčastejším typom výskumných dizajnov. Ide o postup bežne používaný pri experimentoch, kde sledujeme rozdiely v experimentálnej skupine, ktorá podstúpila určitý zásah alebo intervenciu, a kontrolnej skupine. Ak sa majú porovnávať dve skupiny, napríklad pri klasickom porovnaní kontrolnej a treatment skupiny, predpokladá sa, že vzorky boli odobraté z dvoch samostatných populácií. V tomto prípade sa musia použiť iné testy.
- So štatistickým porovnávaním dvoch skupín sa však stretneme aj v mnohých iných prípadoch. Porovnávanie skupín sú bežné aj pri vnútrosubjektových výskumoch, napr. porovnania meraných vlastností pred určitým zásahom (napr. absolvovaním krátkodobej psychoterapie) a potom u tých istých ľudí, tak aj v prierezovom výskume, kde porovnáваме určité existujúce skupiny (napríklad na základe demografických vlastností).
- Na porovnanie medzi viac ako dvoma skupinami boli vyvinuté ďalšie testy. Ak sa má porovnávať niekoľko skupín, výber vhodného testu závisí od toho, či sú skupiny od seba štatisticky nezávislé (nesúvisiace alebo nespárované) alebo nie (príbuzné alebo spárované). Na túto otázku do značnej miery odpovedá použitý experimentálny dizajn. Testovanie jednotlivých subjektov v množstve experimentálnych podmienok alebo skupín nevyhnutne vedie k príbuzným vzorkám.
- Dve po sebe nasledujúce rozhodnutia tej istej osoby nemôžu byť zo svojej podstaty navzájom nezávislé. Nezáleží na tom, či sa osoba rozhoduje dvoch rôznych podmienach (cross-over design) alebo v jednej a tej istej podmienke (pozdĺžny design). Na druhej strane, ak sa každý subjekt rozhoduje iba raz, možno za podmienok úplnej anonymity a bez spätnej väzby predpokladať, že rozhodnutie jednej osoby neovplyvňuje rozhodnutie inej osoby.

# Klasifikácia testovacích metód

- Tretím kritériom ovplyvňujúcim výber štatistických metód je otázka, aké predpoklady o rozdelení pravdepodobnosti premenných platia. V závislosti od odpovede sú k dispozícii dve široké triedy metód, parametrické a neparametrické. Parametrické metódy poskytujú zmysluplné výsledky len vtedy, ak platia špecifické predpoklady o forme (napr. normálne rozdelenie) a parametroch (napr. priemer, rozptyl, stupne voľnosti) rozdelenia.
- Pokiaľ je vzorka veľmi veľká, nehrá otázka rozdelenia veľkú rolu. Aj keď skutočné rozdelenie nie je normálne a parametrický test vyžaduje normálne rozdelenie, tento test bude stále poskytovať pre veľké vzorky spoľahlivé výsledky. Z tohto dôvodu sa hovorí, že parametrické testy sú robustné (voči odchýlkam v rozdelení) pre veľké vzorky. Pri malých vzorkách sa však veľmi odporúča uistiť sa, že predpoklady týkajúce sa rozdelenia sú správne. Aj malé odchýlky od predpokladaného rozdelenia môžu spôsobiť, že výsledok testu bude úplne nepoužiteľný.
- Alternatívou k tomu sú neparametrické metódy. Nezávisia od formy a parametrov rozloženia populácie, z ktorej bola vzorka odobratá. To však, samozrejme, neznamená, že neparametrické postupy nevyžadujú žiadne predpoklady. Predpoklady sú len menej obmedzujúce ako v parametrickom prípade.
- Pokiaľ ide o veľké vzorky, nie je potrebné sa príliš obávať, ktorá trieda je lepšia voľba. Parametrický test má len o niečo vyššiu silu ako jeho neparametrický náprotivok, ale druhý môže byť o niečo jednoduchší.

# Ako si vybrať správny test?

- Pri výbere testu štatistickej hypotézy je potrebné zvážiť aspoň tieto kritériá:
  - Jedna alebo viac skupín?
  - Súvisiace alebo nesúvisiace skupiny?
  - Parametrické alebo neparametrické údaje alebo mierky merania údajov?

**Table 7.5** Frequently used statistical tests

	Level of measurement and parametric assumptions		
	Interval/ratio and normal	Ordinal or interval/ratio and not normal	Categorical
<b>One sample</b>	<i>t</i> -test <i>z</i> -test Chi-square test for the variance	Wilcoxon test Sign test Kolmogorov–Smirnov test	Binomial test Chi-squared test
<b>Independent samples</b>			
2-sample	<i>t</i> -test <i>z</i> -test  Welch’s test <i>F</i> -test One-way ANOVA	Mann–Whitney test Kolmogorov–Smirnov test  Siegel–Tukey test Kruskal–Wallis test	Fisher exact test Chi-squared test  Chi-squared test
K-sample	Barlett’s test	Levene’s test	
<b>Dependent samples</b>			
2-sample	Paired <i>t</i> -test	Matched-pairs Wilcoxon test	McNemar test
K-sample	Repeated-measure ANOVA	Friedman test	Cochran’s Q test

# z-Test a t-Test pre jednu vzorku

- z-test pre jednu vzorku skúma, či priemer  $x$  náhodnej vzorky je dostatočne konzistentný s priemerom danej populácie  $\mu_0$ , ktorý sa považuje za pravdivý. Ak je rozdiel medzi  $x$  a  $\mu_0$  významný, údaje nepodporujú hypotézu, že vzorka bola odobratá z populácie s priemerom  $\mu = \mu_0$ . V súlade s tým je nulová hypotéza  $H_0: \mu = \mu_0$  a alternatívne hypotézy sú  $H_1: \mu \neq \mu_0$  alebo  $H_1: \mu < \mu_0$  alebo  $H_1: \mu > \mu_0$ .
- Keďže ide o parametrickú metódu, dôležitým predpokladom je, že vzorka bola odobratá z normálne rozloženej populácie so známym rozptylom  $\sigma^2$ . Veľkosť vzorky z-testu by mala zahŕňať aspoň 30 pozorovaní.
- Na rozdiel od nulovej distribúcie z-testu je nulová distribúcia t-testu rozdielna pre rôzne veľkosti vzorky, pretože závisí od stupňov voľnosti  $n-1$ .
- Príklad
  - Výsledky celoštátneho matematického testu sú zvyčajne rozdelené s priemerom  $\mu = 78$  bodov a štandardnou odchýlkou  $\sigma = 12$  bodov. Učiteľ konkrétnej školy chce otestovať, či jeho novozavedený spôsob vyučovania matematiky má pozitívny významný vplyv na bodové skóre žiakov. Jeho výskumná hypotéza je teda  $H_1: \mu > 78$ .
  - 36 študentov v jeho kurze získalo priemerné skóre  $x = 82$  z hodnôt 94, 68, 81, 82, 78, 94, 91, 89, 97, 92, 76, 74, 74, 92, 98, 70, 55, 56, 83, 65, 83, 91, 76, 79, 79, 86, 82, 93, 86, 82, 62, 93, 95, 100, 67, 89.
  - Ide v našom prípade o náhodné, nesystematické rozdiely vysvetliteľné výberovou chybou, alebo ide o skutočný rozdiel v populácii? Inými slovami, môžeme tvrdiť, že študentky a študenti učiteľa dosiahli naozaj lepšie skóre?
  - Štatistika testu potom  $z = (82-78) / (12/\sqrt{36}) = 2$ .
  - P-hodnota je  $2,5\% < 5\%$  a zamietame  $H_0: \mu = 78$  na hladine významnosti  $5\%$ .

# t-Test pre dve nezávislé vzorky (Between-Subject)

- Aby sme mohli porovnať dve vzorky, musíme upraviť jednovzorkový t-test. Pritom predpokladáme, že nikto nie je zastúpený v oboch vzorkách súčasne a že výsledky jednej vzorky nie sú žiadnym spôsobom ovplyvnené výsledkami druhej vzorky.
- Test určí, či sa priemery  $x_1$  a  $x_2$  týchto dvoch nezávislých vzoriek líšia natolko, že možno dospieť k záveru, že medzi nimi existuje významný rozdiel.
- Ak je rozdiel medzi  $x_1$  a  $x_2$  významný, údaje nepodporujú hypotézu, že vzorky boli odobraté z populácií s rovnakým priemerom,  $\mu_1 = \mu_2$ . Preto je nulová hypotéza  $H_0: \mu_1 - \mu_2 = \mu_0$ , teda vo všeobecnosti sa testuje „bez rozdielu“, t. j.  $\mu_0 = 0$ . Alternatívne hypotézy sú  $H_1: \mu_1 - \mu_2 \neq \mu_0$  alebo  $H_1: \mu_1 - \mu_2 < \mu_0$  alebo  $H_1: \mu_1 - \mu_2 > \mu_0$ .
- Keďže sme stále v oblasti parametrických metód, je potrebné predpokladať, že obe vzorky boli náhodne vybrané z normálne rozloženej populácie. Tieto dve populácie majú rovnaký, aj keď neznámy rozptyl  $\sigma^2$ , ale nie je potrebné, aby vzorky mali rovnakú veľkosť. Je mimoriadne dôležité, aby boli subjekty priradené k rôznym treatmentom náhodne. Iba úspešná randomizácia môže zabezpečiť, že sa vyhneme efektom selekcie.

# t-Test pre dve závislé vzorky (Within-Subject)

- Ďalšia úprava t-testu je potrebná, ak realizácia jednej vzorky nie je nezávislá na realizácii druhej vzorky. Toto je vždy prípad within-subject dizajnu, pretože jeden subjekt sa rozhoduje v dvoch rôznych treatmentoch a je teda v oboch vzorkách.
- Preto v dátach existujú dvojice rozhodnutí toho istého subjektu. Nulová hypotéza je  $H_0: \mu_1 - \mu_2 = \mu_0$ , pričom sa zvyčajne testuje  $\mu_0 = 0$  a alternatívne hypotézy sú  $H_1: \mu_1 - \mu_2 \neq \mu_0$  alebo  $H_1: \mu_1 - \mu_2 < \mu_0$  alebo  $H_1: \mu_1 - \mu_2 > \mu_0$ .
- Opäť sa vzorky náhodne odoberú z normálne distribuovanej populácie neznámeho, ale rovnakého rozptylu  $\sigma^2$ . Štandardná chyba sa vypočíta z váženého priemeru rozptylov vzoriek, korigovaných stupňom korelácie medzi týmito dvoma vzorkami.

# Kolmogorov Test

- Kolmogorovov test je jedným z testov, ktoré sa nazývajú testy zhody. Tieto testy skúmajú, či rozdelenie hodnôt vzorky je také, aké by sa dalo očakávať na základe špecifického, vopred definovaného rozdelenia. To znamená, že tento test poskytuje štatistický dôkaz o tom, či je alebo nie je splnený predpoklad konkrétneho rozdelenia.
- Na tento účel sa porovnáva empirická distribučná funkcia  $F_x$  vzorky, teda podiel pozorovaných hodnôt  $x$ , ktoré sú menšie alebo rovné špecifickej hodnote  $x$  (pre všetky reálne hodnoty  $x$ ), s vopred definovanou, resp. predpokladanou distribučnou funkciou  $F_0$ . Štatistika testu  $D$  meria stupeň zhody a je maximálnou vzdialenosťou medzi  $F_x$  a  $F_0$ .
- Nulová hypotéza postuluje zhodu medzi teoretickým a empirickým rozdelením a alternatívna hypotéza tvrdí, že vzorka nepochádza z teoretického rozdelenia. Z tohto dôvodu sa v praxi zvyčajne používa dvojstranná hypotéza, ktorá pripúšťa odchýlku v oboch smeroch.
- Na rozdiel od väčšiny ostatných testov pri Kolmogorovovom teste nechceme zamietnuť nulovú hypotézu, pretože zvyčajne očakávame potvrdenie predpokladu týkajúceho sa konkrétneho rozdelenia (napr. normálneho rozdelenia). Čím viac sa údaje líšia od referenčného rozdelenia, tým vyššia je pravdepodobnosť, že nulová hypotéza bude zamietnutá.

# Wilcoxon Rank-Sum test a Mann-Whitney U Test

- Wilcoxonov test poradia súčtu je populárnou alternatívou k t-testu, keď sa nezdá realistické predpokladať normálne rozdelenie a/alebo údaje nie sú metricky škálované.
- Podobne ako t-test porovnáva rovnosť „stredových bodov“ dvoch nezávislých vzoriek. Pri tomto neparametrickom teste však už nepoužívame aritmetické priemery a pri porovnávaní skupín namiesto toho hovoríme o „centrálnych tendenciách“.
- Alternatívnou metódou, ktorá vždy vedie k rovnakému výsledku testu ako Wilcoxonov rank-sum test, je Mann-Whitney U test.
- Wilcoxonov rank test (dve závislé vzorky)
  - Rovnako ako Wilcoxonov rank-sum test možno považovať za neparametrický náprotivok k t-testu s dvoma nezávislými vzorkami, Wilcoxonov rank test možno použiť ako neparametrickú alternatívu k t-testu s dvoma závislými vzorkami. Je to jeden zo štandardných testov pre dáta vo within-subject dizajne.
  - Hypotézy sú rovnaké ako hypotézy vo Wilcoxonovom rank-sum teste. Ak je nulová hypotéza platná, predpokladá sa, že rozdiely pochádzajú z populácie, ktorá je symetricky rozložená okolo mediánu s hodnotou 0.

# Porovnávanie viacerých skupín

- t-testy a Mannov-Whitney U-test sú postačujúce vtedy, ak nás zaujímajú len dve skupiny, ako napr. pri porovnaní jednej experimentálnej a kontrolnej skupiny.
- Čo však v prípade, že je potreba porovnávať viac ako dve skupiny? Ako prvé riešenie tejto potreby nám môže napadnúť, že jednoducho opakovane použijeme metódy, ktoré slúžia na porovnanie dvoch skupín. Ale aj keď sa to môže javiť ako veľmi jednoduché a použiteľné riešenie, v skutočnosti má základné nevýhody, kvôli ktorým uprednostíme iný postup, analýzu rozptylu (ANOVA).
- Pozrime sa teda na dva problémy, ktoré by mohli nastať pri porovnaní dvojitých skupín prostredníctvom t-testov. Prvým (a menej významným) problémom je stúpajúci počet porovnaní. Stačí, ak máme len tri skupiny, už počet dvojíc (a teda aj počet testov) stúpne na 3 (ak označíme skupiny písmenami A, B a C, sú to dvojice A-B, A-C a B-C. To pri používaní štatistického softvéru nie je ešte žiaden problém, aj keď to môže chvíľu trvať, počet porovnaní však bude výrazne stúpať pri zvyšujúcom sa počte skupín. Pri štyroch skupinách to bude už 6 porovnaní, pri piatich skupinách 10 porovnaní.
- Omnoho závažnejší je však druhý dôvod, ktorý do úvahy berie už nielen stúpajúci počet porovnaní, ale aj s ním stúpajúcu pravdepodobnosť chyby. Vieme, že pri vyhodnocovaní hypotéz môžu nastať dva druhy chýb, pričom v klasickom (frekventistickom) prístupe sa snažíme kontrolovať najmä chybu  $\alpha$  (chyba prvého typu) – nesprávne zamietnutie nulovej hypotézy. Klasicky používanou hranicou je 0,05, teda maximálne 5 % percentná pravdepodobnosť, že ak nulovú hypotézu zamietneme, nedopustíme sa chyby. To, samozrejme, znamená, že v piatich percentách prípadov (to je jeden z dvadsiatich) sa môže stať, že sa rozhodneme nesprávne – nulovú hypotézu zamietneme, aj keď by sme nemali.

# Porovnávanie viacerých skupín

- Prečo to však spomíname práve teraz? Ved', samozrejme, platí, že aj keď by sme použili testy určené pre porovnanie dvoch skupín na porovnanie viacerých skupín vyčlenených na základe jednej premennej (napr. veku), stále budeme tú známu 5-percentnú hranicu rešpektovať. Nuž, odpoveď je dvojznačná – áno aj nie. Áno znamená, že je pravda, že ju rešpektujeme pri každom jednom teste – každom porovnaní vybranej dvojice skupín. Nie znamená, že chyby sa navzájom kombinujú. Ilustrovať to môžeme na zjednodušenom príklade dvoch ľudí, o ktorých vieme, že aj jeden, aj druhý hovoria pravdu s 95 % pravdepodobnosťou. Ak mi niečo tvrdí ten prvý, tak viem, že sa na jeho tvrdenie viem s touto mierou pravdepodobnosti spoľahnúť, alebo, inak povedané, že ak mi bude niečo tvrdiť 20-krát, v 19 prípadoch to bude pravda. Čo ak však jedno tvrdenie pochádza od prvého a druhé (úplne iné, týkajúce sa niečoho iného) od druhého človeka? Vtedy síce pri oboch tvrdeniach môžeme povedať, že sú (jedno aj druhé) s pravdepodobnosťou 95 % pravdivé, ale nemôžeme povedať, že je aj pravdepodobnosť 95 %, že je pravdivé aj prvé, aj druhé tvrdenie.
- Tu sa dostávame do oblasti kombinácie pravdepodobností, kde pri vzájomne nezávislých udalostiach (pravdepodobnosť výskytu prvej udalosti nezávisí na výskyte druhej) platí, že pravdepodobnosť ich súčasného výskytu vypočítame ako súčin pravdepodobností výskytu izolovaných udalostí.
- A presne to je aj prípad viacnásobného použitia porovnania dvojíc. Aj keď pri každom izolovanom t-teste môžeme mať pravdepodobnosť správneho rozhodnutia o nulovej hypotéze 95-percentnú, pri realizácii viacerých takýchto testov je potrebné túto hodnotu vynásobiť. Môžeme sa odvolať na predchádzajúci výpočet počtu porovnaní, keď pri troch skupinách to boli tri porovnania, pri štyroch skupinách šesť a pri piatich desať porovnaní. Aká bude dôveryhodnosť výsledkov a im zodpovedajúce riziko a chyby? Pri troch skupinách (a troch porovnaníach) to je  $0,95 \cdot 0,95 \cdot 0,95$ , teda 0,86, pravdepodobnosť a chyby je potom  $1 - 0,86 = 0,14$ , pri štyroch skupinách (šesť porovnaní) je to  $1 - 0,74 = 0,26$ . V priemere jeden zo štyroch prípadov (26 %) porovnávaní štyroch skupín po dvojiciach tak zistí významné rozdiely medzi nimi, aj keď v skutočnosti neexistujú.
- Na základe uvedeného by teda malo byť zrejmé, že nejaká metóda, ktorá bude túto kumuláciu a chyby kontrolovať a udržiavať na akceptovateľnej úrovni, je na porovnanie viacerých skupín veľmi vhodná.
- Touto metódou je analýza rozptylu (analysis of variance, ANOVA), alebo neparametrická alternatíva - Kruskal-Wallis test.

# ANOVA, Kruskal-Wallis

- Základná ANOVA a základný Kruskal-Wallis test overujú, či existujú významné rozdiely v aspoň jednej z porovnávaných dvojíc. Ak zamietneme nulovú hypotézu (nájdú sa významné rozdiely aspon v jednej dvojici), ako zistíme, o ktorú dvojicu/ktoré dvojice ide?
- Najčastejšie používanou metódou po realizácii ANOVA, presnejšie po zamietnutí nulovej hypotézy v tomto teste, sú tzv. post hoc testy. Sú to metódy, ktoré sa používajú na porovnanie (najčastejšie všetkých možných) dvojíc, aby bolo možné povedať, pri ktorej dvojici môžeme hovoriť o štatisticky významných rozdieloch v priemerných hodnotách. Už názov hovorí, že ich použijeme po niečom, v tomto prípade po zamietnutí nulovej hypotézy.
- Medzi najčastejšie používané patria Scheffého a Tukeyho test, pri nezhodných rozptyloch sa odporúča Gamesov-Howellov test (Sauder & DeMars, 2019).
- Pri problémoch s rozdelením závislej premennej je odporúčanou alternatívou ANOVY neparametrický Kruskalov-Wallisov test. V prípade post hoc testov je možnosť použiť už známy Mannov-Whitneyho U-test alebo Dunnov test.

# Binomiálny test

- Mnohé premenné v experimentoch majú len dva možné výsledky, ako napríklad „prijatť ponuku/odmietnuť ponuku“ v ultimátnej hre, „spolupracovať/zradiť“ v hre s väzňovou dilemou alebo „vybrať párne číslo/vybrať nepárne číslo“ v matching pennies. Takouto dichotomickou premennou môže byť reprezentované aj hádzanie mincou s výsledkami „hlava“ alebo „oroľ“. Jednorazové uskutočnenie takéhoto experimentu nazývame Bernoulliho pokus a dva možné výsledky ako „úspech“ a „neúspech“.
- Ak hádzeme mincou, pravdepodobnosť oboch možností je 50 percent. V laboratórnom experimente zahrňajúcom rozhodovanie ľudí však nie je vo všeobecnosti vopred známa pravdepodobnosť, s ktorou sa subjekty rozhodnú pre jednu alebo druhú alternatívnu akciu. Ak však teória špecifikuje konkrétnu hodnotu pravdepodobnosti, v laboratóriu môžeme overiť, či dáta štatisticky podporujú danú teoretickú hodnotu alebo nie.
- Vo vyššie spomínanej matching pennies napríklad teória hier predpovedá rovnováhu, v ktorej obaja hráči hrajú obe alternatívy s rovnakou pravdepodobnosťou, teda s  $p = P(\text{výber párneho čísla}) = 1 - p = P(\text{výber nepárneho čísla}) = 0,5$ . Ak sa táto hra hrá v laboratóriu dostatočne často, relatívna frekvencia pre „párne číslo“ („úspech“) a „nepárne číslo“ („neúspech“) sa získa jednoduchým spočítaním príslušných realizácií. Táto frekvencia sa tiež označuje ako empirická pravdepodobnosť úspechu  $p_U$ .
- Binomický test skúma, či pozorovaná hodnota  $p_U$  je taká, aká by sa dala očakávať. Ak je rozdiel medzi nameranou a očakávanou hodnotou ostatočne veľký, potom je nulová hypotéza zamietnutá, t. j. pri zohľadnení danej pravdepodobnosti chyby nie je špecifikovaná hodnota konzistentná s pozorovanou vzorkou. Ak však nulovú hypotézu nemožno zamietnuť, experimentálne dáta podporujú teoretickú predpoveď.
- Uvažovaná premenná je buď dichotomická, t. j. môže mať podľa definície iba dve hodnoty, ako napríklad výsledok hodu mincou, alebo je kategoricky škálovaná s 2 kategóriami, napr. „vysoké“ vs. „nízke“ príspevky v hre o verejné statky (public goods game).

# Multinomiálny test (1 × k)

- Multinomiálny test je zovšeobecnením binomiálneho testu na kategorické premenné s  $k > 2$  kategóriami. Napríklad by mohlo byť žiaduce klasifikovať sumy uvedené v hre diktátora nielen ako „vysoké“ a „nízke“, ale skôr presnejšie ako „vysoké“, „stredné“ a „nízke“, zodpovedalo by to kategorickej premennej s tromi kategóriami.
- Inak je princíp testu multinomiálneho testu úplne analogický s princípom binomického testu. Test skúma, či empirické frekvencie  $\pi_1, \dots, \pi_k$  kategórií  $k$  sú také, ktoré by sa dali očakávať za predpokladu, že v skutočnosti pravdepodobnosti úspechu kategórií nadobudnú určité dané hodnoty  $p_1, \dots, p_k$  (nulová hypotéza).

•

# Fisherov exaktný test ( $2 \times 2$ )

- Multinomiálny test porovnáva frekvencie jednej vzorky v  $k$  kategóriách s očakávanými hodnotami referenčného rozdelenia (napr. rovnomerné rozloženie vo všetkých  $k$  kategóriách). Ak teraz chceme navzájom porovnať dve nezávislé, kategoricky škálované vzorky (alebo skupiny alebo treatmenty), potom ponúka dobré riešenie Fisherov exaktný test.
- Pozorované frekvencie sa najskôr zhrnú do kontingenčnej tabuľky. Riadky a stĺpce tejto tabuľky obsahujú príslušné hodnoty dvoch kategorických premenných.
- Fisherov exaktný test kontroluje, či sú frekvencie dostatočne odlišné, aby naznačovali významný rozdiel medzi skupinami. Nulová hypotéza predpokladá, že frekvencie populácie sú rovnaké resp. že tieto dve vzorky pochádzajú z rovnakej populácie.

Table 4.13 Contingency table 1 for Fisher's exact test in the example

		Measured categorical variable		
		High school diploma	No high school diploma	
Gender	Male	$x_{11} = 2$	$x_{12} = 6$	$n_1 = 8$
	Female	$x_{21} = 9$	$x_{22} = 3$	$n_2 = 12$
		$N_1 = 11$	$N_2 = 9$	$N = 20$

- Fisherov exaktný test sa však stane nepraktickým, keď sa zvýši počet tried kategorickej premennej alebo počet pozorovaní. Pre tieto prípady ponúka zjednodušenú aproximáciu test  $\chi^2$ .

# Lineárna regresia

- Počas minulej prednášky sme zistili, že je možné pomocou jedného čísla – korelačného koeficientu – vyjadriť silu a smer vzťahu medzi dvomi premennými. Tak sa môžeme dozvedieť napr. to, ako veľmi je spojená inteligencia so školským prospechom, miera svedomitosti s úspechom v práci či úzkostná vzťahová väzba s problémami v partnerskom vzťahu.
- Pri uvedených vzťahoch nám však môže hneď napadnúť ďalšia otázka – ak sú v týchto dvojiciach identifikované nezanedbatelné vzťahy, nebolo by možné na základe jednej premennej predpovedať druhú? Pri poznaní inteligencie predpovedať školský prospech, pri poznaní svedomitosti úspešnosť pri vykonávaní práce a pri informácii o úzkostnej vzťahovej väzbe problémy v partnerskom vzťahu?
- To je práve otázka, na ktorú sa bude snažiť odpovedať metóda, ktorú nazývame lineárna regresia. Ak poznáme hodnotu jednej premennej, je možné predpovedať hodnotu inej premennej, ktorá s ňou súvisí? Lineárna regresia bude, samozrejme, vedieť poskytnúť spoľahlivú predpoveď len v tom prípade, ak tieto dve premenné navzájom súvisia, miera spoľahlivosti predpovede tak môže kolísať od veľmi dobrej po bezcennú. To je ďalším cieľom lineárnej regresie – posúdiť mieru spoľahlivosti predpovede.
- Princíp lineárnej regresie
- Základný princíp lineárnej regresie je možné odvodiť už z prvého slova jej názvu. Lineárny znamená majúci tvar priamky, v tomto prípade sa teda budeme snažiť najprv znázorniť vzťah dvoch premenných pomocou priamky a potom túto priamku použiť na predpovedanie jednej premennej z druhej. Vizuálnou pomôckou je bodový graf (scatter plot). V danom grafe sa na osi x nachádza jedna, na osi y druhá premenná, body v grafe zobrazujú individuálne prípady, ktorých umiestnenie v grafe je dané konkrétnymi hodnotami týchto premenných.

# Štatistické modely

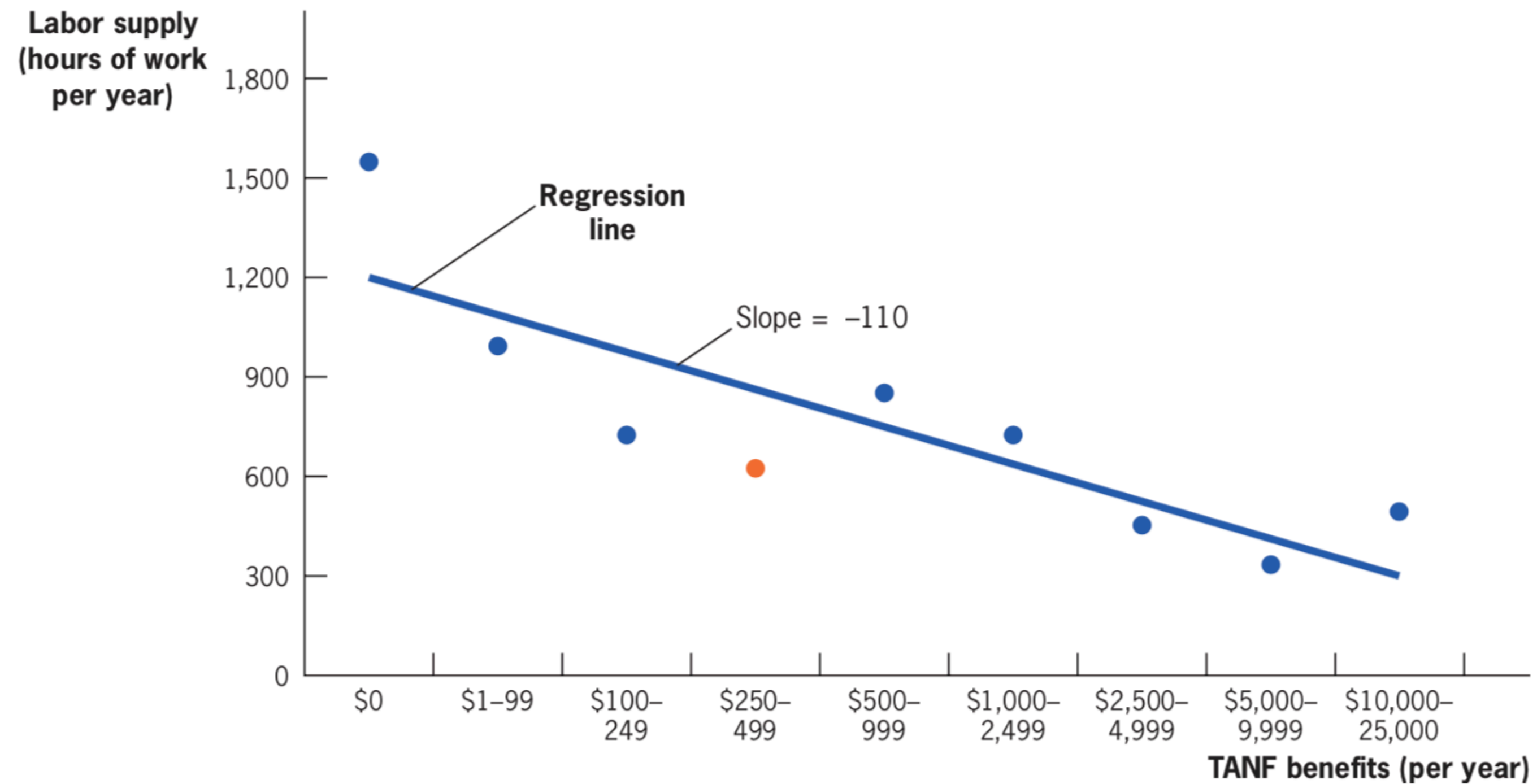
- Testovanie štatistickej významnosti treatment efektov nie je jediným dôvodom mnohých experimentálnych štúdií. Ďalším je odhadnúť vzťahy medzi premennými. Na tento účel je zvyčajne vyvinutý štatistický model s cieľom čo najlepšie získané údaje vysvetliť. Takýto model možno použiť na zodpovedanie ďalších otázok, ako napríklad:
- Ako možno vysvetliť atribúty premennej pomocou iných premenných a ako dobre sa to dá urobiť? Akú hodnotu by pravdepodobne mala premenná, ak by bola ovplyvnená atribútom inej premennej, ktorá nebola vyvolaná v experimente?
- Východiskovým bodom pre štatistický model je snaha vysvetliť zmeny v experimentálne pozorovanej premennej  $y$ . Premenná  $y$  sa preto nazýva aj endogénna premenná. Informácie, ktoré používame na vysvetlenie endogénnej premennej, pochádzajú z jednej alebo viacerých vysvetľujúcich premenných (exogénnych premenných). Základným predpokladom každého štatistického modelu je, že medzi týmito dvoma premennými existuje skutočný vzťah, ktorý však nie je známy. V obzvlášť jednoduchých prípadoch môže byť vhodné predpokladať lineárny vzťah medzi endogénnou premennou  $y$  a presne jednou exogénnou premennou  $x$ . To by potom malo tvar  $y = a + bx$ .
- Takýto model vždy podlieha určitým predpokladom. Najdôležitejšie z nich sú:
  - V ekonometrickom modeli nechýbajú žiadne relevantné exogénne premenné a použité exogénne premenné nie sú irelevantné.
  - Skutočný vzťah medzi exogénnou premennou a endogénnou premennou je lineárny.
  - Parametre priesečníka a sklonu sú konštantné pre všetky pozorovania.
  - Chyba je normálne rozdelená s  $u_i \sim N(0, \sigma^2)$  pre všetky pozorovania  $i$  a chyby všetkých pozorovaní  $i$  sú navzájom štatisticky nezávislé.
  - Hodnoty nezávislej premennej  $x$  sú štatisticky nezávislé od chybovej premennej  $u$ .

# Prierezová regresná analýza

- Ďalším postupom pre identifikovanie kauzálnych efektov je prierezová regresná analýza, ktorá vyhodnocuje vzťah dvoch premenných, pričom drží ostatné faktory konštantné. “Prierezová” znamená porovnanie množstva jednotlivcov v jednom časovom bode (teda nie vo viacerých obdobiach ako to robia časové rady)
- Regresná analýza popisuje (a kvantifikuje) vzťah medzi premennou, ktorej vývoj sa snažíme objasniť (závislá premenná) a súborom premenných, o ktorých sa domnievame, že by ju vysvetliť mohli (nezávislé premenné)
- Odhad takéhoto vzťahu zobrazuje regresná priamka. Keďže priamka, ktorá by prešla všetkým bodmi v dátovom súbore neexistuje (museli by sme nakresliť krivku), lineárna regresia hľadá priamku, ktorá je k bodom čo najbližšie
- Technicky priamka identifikuje súčet najmenších štvorcov, teda kvadrátov vzdialenosti jednotlivých bodov od priamky. Táto technika je tým pádom citlivá na body, ktoré sú od nej príliš ďaleko (outliers), keďže im dáva veľmi vysokú váhu a teda vplyv. Niekedy je preto nutné zvážiť aj postupy, ktoré budú k vzdialeným pozorovaniam menej citlivé.

# Prierezová regresná analýza

■ FIGURE 3-4



**TANF Benefit Income and Labor Supply of Single Mothers, Using CPS Data** • Using data from the CPS, we group single mothers by the amount of TANF income they have. Those who are receiving the lowest level of TANF income are the ones providing the highest number of work hours.

Source: Calculations based on data from Current Population Survey's annual March supplements.

# Prierezová regresná analýza

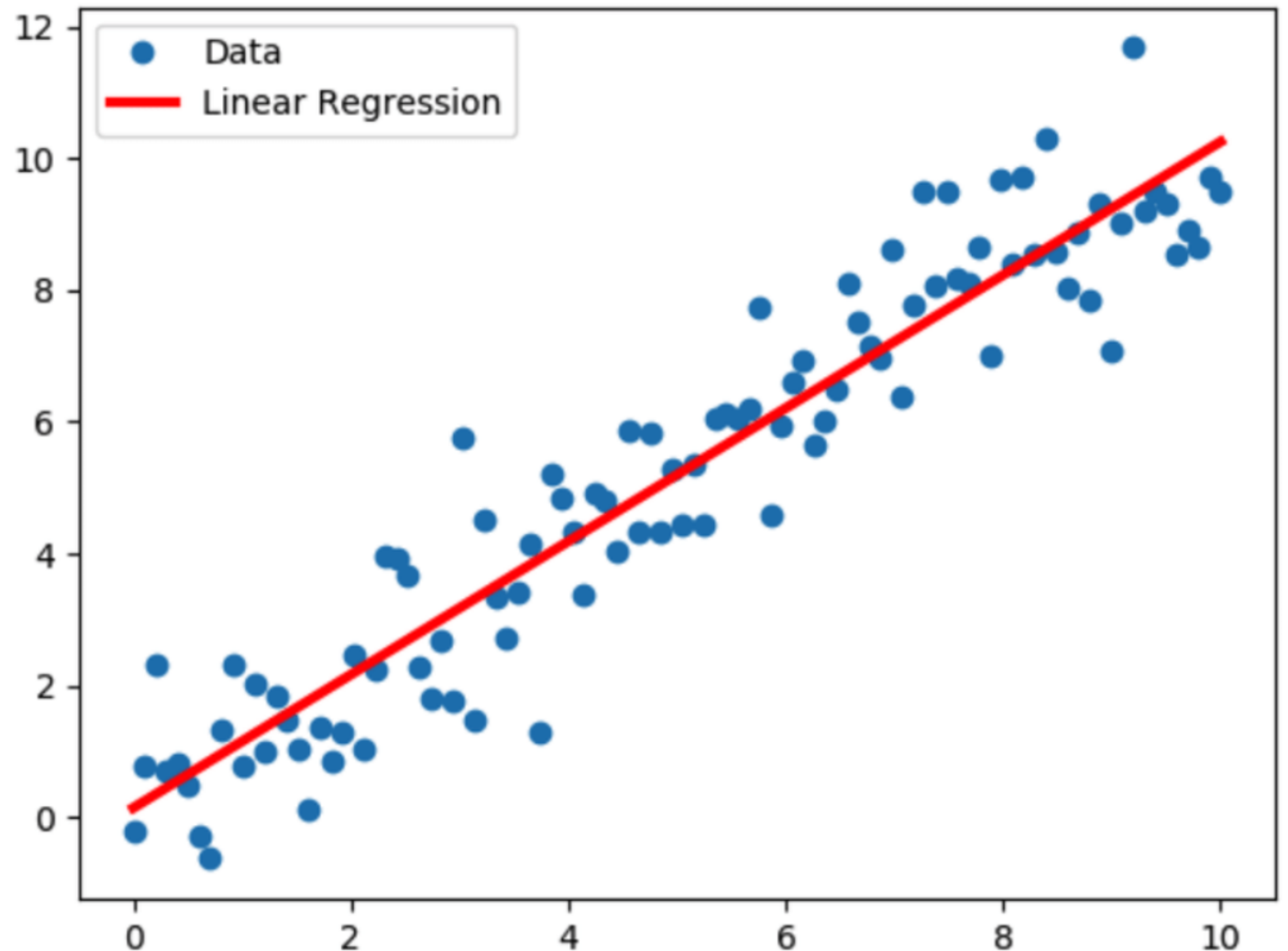
- Vzťah medzi dvoma premennými, ktorý aproximujeme regresnou priamku nemusí byť kauzálny. Preto výsledky regresie neinterpretujeme slovami “x-percentný pokles/nárast premennej A spôsobuje y-percentný pokles/nárast premennej B”, ale “x-percentný pokles/nárast premennej A je spojený s y-percentný pokles/nárast premennej B”
- Regresná analýza má oproti korelácii výhodu v tom, že dokáže pracovať so skreslením, vďaka možnosti zahrnúť do nej kontrolné premenné. Kontrolné premenné vnesú do analýzy iné ako intervenčné rozdiely medzi jedincami vo vzorke, a tým pádom zostávajúcu koreláciu medzi A a B je možné o trochu spoľahlivejšie označiť za kauzálnosť.
- Použitie kontrolných premenných však nikdy nevyrieši problém skreslenia úplne, keďže väčšinou nemáme k dispozícii všetky charakteristiky, ktoré by mohli závislé premenné ovplyvňovať.

# Prierezová regresná analýza

- $y = \alpha + \beta x + e$
- kde
  - $\alpha$  = konštanta (hodnota pre  $x = 0$ )
  - $\beta$  = koeficient sklonu, vyjadrujúci zmenu  $y$  pri jednotkovej zmene  $x$
  - $e$  = odchýlka, ktorá vyjadruje rozdiel medzi skutočnou hodnotou premennej a jej hodnotou odhadovanou modelom

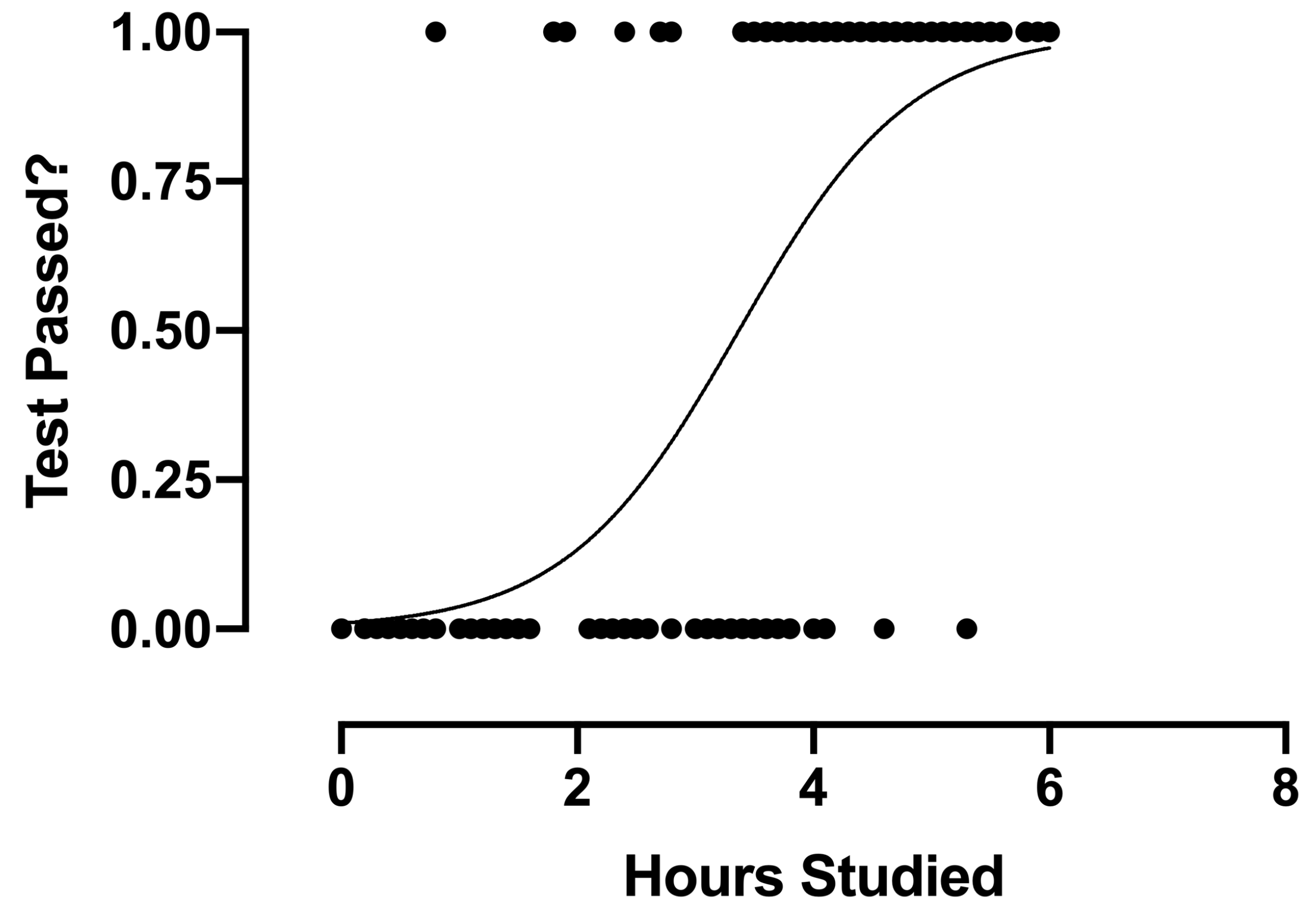
# Regressné modely

- Lineárna regresia
- Najčastejšie používanou technikou modelovania je lineárna regresia, ktorá predpokladá lineárny vzťah medzi závislou premennou (Y) a nezávislou premennou (X). Vyjadruje sa formou regresnej priamky, známej aj ako priamka najlepšej zhody. Lineárne spojenie je definované ako  $Y = \alpha + \beta \cdot X + e$ , kde " $\alpha$ " označuje priesečník, " $\beta$ " označuje sklon priamky a " $e$ " je chybový člen.
- Lineárny regresný model môže byť jednoduchý (len s jednou závislou a jednou nezávislou premennou) alebo zložitý (s mnohými závislými a nezávislými premennými) (s jednou závislou premennou a viac ako jednou nezávislou premennou).



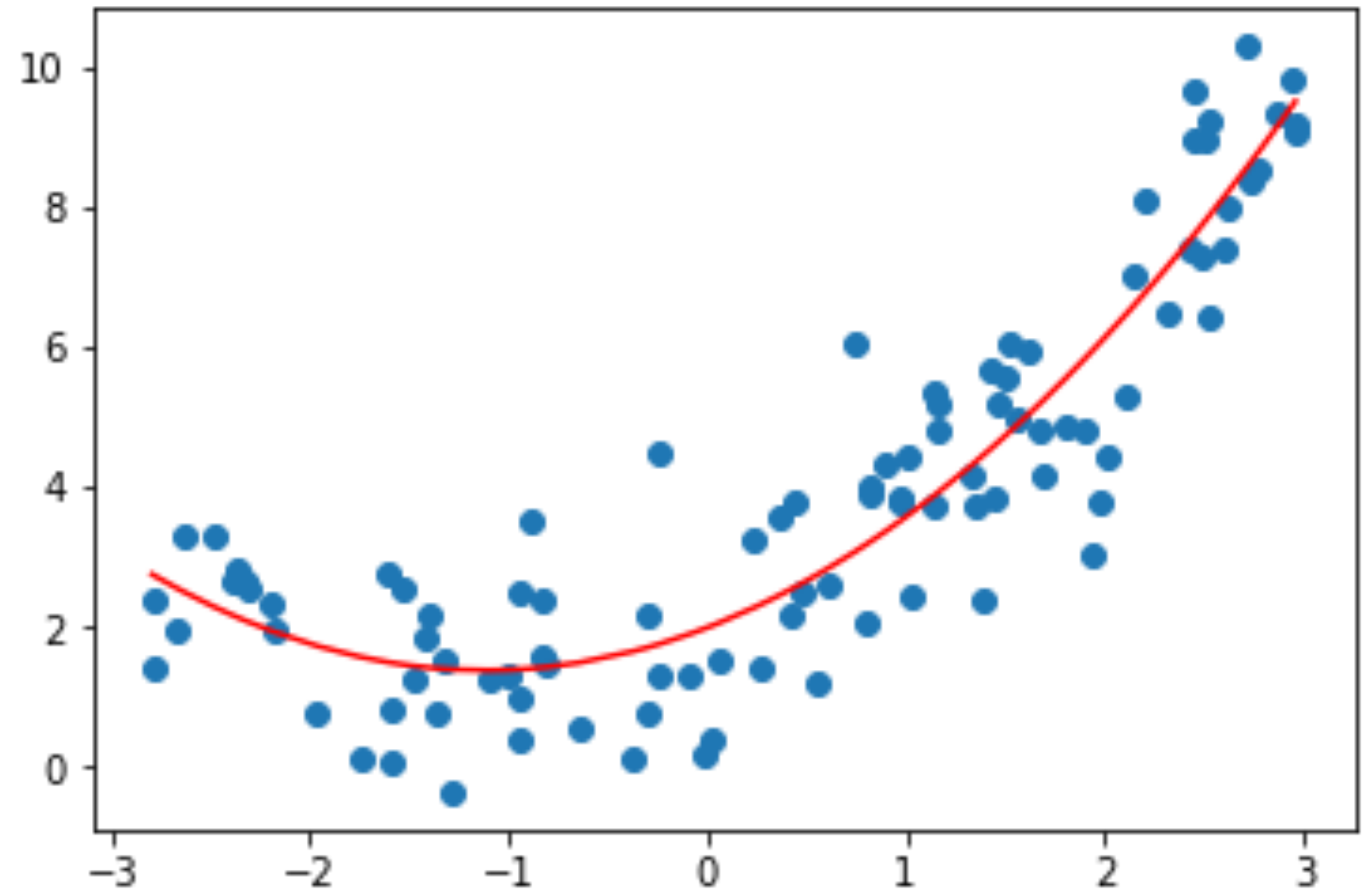
# Regresné modely

- Logistická regresia
- Ak je závislá premenná diskretná, používa sa technika logistickej regresie. Inými slovami, táto technika sa používa na výpočet pravdepodobnosti vzájomne sa vylučujúcich výskytov, ako je napríklad vyhovel/nevyhovel, pravda/nepravda, 0/1 atď. Cieľová premenná teda môže nadobúdať len jednu z dvoch hodnôt, sigmoidná krivka predstavuje jej vzťah s nezávislou premennou a pravdepodobnosť má hodnotu medzi 0 a 1.



# Regresné modely

- Polynomiálna regresia
- Technika polynomickej regresnej analýzy sa používa na vyjadrenie nelineárneho vzťahu medzi závislými a nezávislými premennými. Ide o variant viacnásobného lineárneho regresného modelu s tým rozdielom, že najlepšie vyhovujúca priamka je skôr zakrivená ako rovná.



# Prezentácia regresnej analýzy

Laboratórium			
	(1)	(2)	(3)
	Pravdivé	Nepravdivé	Dezinfo
<b>Prvé meranie</b>	0.65***	0.49***	
	(0.06)	(0.08)	
<b>Prebunk</b>	-3.36	-0.32	1.61
	(1.81)	(1.86)	(2.97)
<b>Debunk</b>	-6.73**	-0.08	-11.89**
	(2.13)	(1.90)	(3.84)
<b>Druhá odpoveď</b>			4.58
			(3.20)
<b>Kontrola na presvedčenia</b>	YES	YES	YES
<b>Kontrola na demografiu</b>	YES	YES	YES
<b>Konštanta</b>	40.48***	36.18***	68.50***
	(11.54)	(8.59)	(11.75)
<b>N</b>	220	220	220
<b>R2</b>	0.51	0.30	0.19

Reprezentatívna vzorka			
	(1)	(2)	(3)
	Pravdivé	Nepravdivé	Dezinfo
<b>Prvé meranie</b>	0.87***	0.78***	
	(0.02)	(0.03)	
<b>Prebunk</b>	-1.12	-2.79**	-4.18*
	(0.87)	(0.91)	(1.72)
<b>Debunk</b>	-2.15*	-1.87*	-5.33**
	(0.94)	(0.89)	(1.73)
<b>Druhá odpoveď</b>			1.27
			(1.73)
<b>Kontrola na presvedčenia</b>	YES	YES	YES
<b>Kontrola na demografiu</b>	YES	YES	YES
<b>Konštanta</b>	4.90	9.40***	35.82***
	(2.65)	(2.67)	(3.68)
<b>N</b>	925	925	925
<b>R2</b>	0.66	0.59	0.30

**Table 2: OLS regressions**

	(1) COVID- 19 FND	(2) Russo- Ukrainian War FND	(3) COVID- 19 FND	(4) Russo- Ukrainian War FND	(5) COVID- 19 FND	(6) Russo- Ukrainian War FND
<b>Media literacy tips</b>	0.22** (0.07)	0.15* (0.07)	0.21** (0.07)	0.14* (0.06)	0.16** (0.06)	0.10 (0.06)
<b>Inoculation</b>	0.06 (0.06)	-0.03 (0.06)	0.05 (0.06)	-0.03 (0.06)	0.04 (0.06)	-0.06 (0.06)
<b>Both interventions</b>	-0.01 (0.06)	0.04 (0.07)	-0.01 (0.06)	0.05 (0.07)	0.03 (0.06)	0.05 (0.06)
<b>Female</b>	0.01 (0.05)	-0.01 (0.05)	0.05 (0.05)	0.03 (0.05)	0.05 (0.04)	0.03 (0.04)
<b>Age</b>	-0.00* (0.00)	-0.00** (0.00)	-0.00 (0.00)	-0.00* (0.00)	-0.00 (0.00)	-0.00 (0.00)
<b>Analytic thinking</b>			0.01 (0.02)	0.02 (0.02)	-0.01 (0.01)	-0.00 (0.01)
<b>Scientific reasoning</b>			0.06*** (0.02)	0.06*** (0.02)	0.02 (0.01)	0.03 (0.01)
<b>Intellectual humility</b>			0.03*** (0.01)	0.02*** (0.01)	0.02*** (0.01)	0.01** (0.01)
<b>Epistemic curiosity - Interest</b>			0.01 (0.01)	0.03*** (0.01)	-0.00 (0.01)	0.02** (0.01)
<b>Epistemic curiosity - Deprivation</b>			0.00 (0.01)	-0.02 (0.01)	0.01 (0.01)	-0.01 (0.01)
<b>Bullshit receptivity</b>			-0.01* (0.00)	-0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
<b>Conspiracy mentality</b>					-0.03 (0.02)	-0.08*** (0.02)
<b>COVID-19 unfounded beliefs</b>					-0.02*** (0.00)	
<b>Pro-Kremlin conspiracy beliefs</b>						-0.21*** (0.03)
<b>Constant</b>	0.51*** (0.09)	0.54*** (0.09)	-0.39* (0.17)	-0.34* (0.17)	0.92*** (0.18)	0.60*** (0.17)
<b>N</b>	1420	1420	1420	1420	1420	1420
<b>R<sup>2</sup></b>	0.01	0.01	0.07	0.06	0.24	0.21

Notes: \*, \*\*, and \*\*\* indicate significance at the 5%, 1%, and 0.1%-level, respectively.