

Ako (ne)opisovať výsledky štatistických analýz v diplomovej práci

Ukážky: zlá verzia vs. lepšia verzia

AUTOR: doc. RNDr. Lucia Ďuricová, PhD.

Upozornenie: Príklady a ich riešenie v tomto návode boli vygenerované pomocou AI. Toto riešenie nemusí byť štatisticky správne, chceme poukázať na spôsob prezentácie výsledkov štatistických analýz v diplomovej práci, nie na (ne)správnosť týchto výsledkov. Príklad textu v diplomovej práci nie je naformátovaný podľa aktuálnej šablóny pre písanie záverečných prác. Formulácie textov nie sú jediné možné formulácie, slúžia len ako príklad.

Príklad 1: Použitie Chí-kvadrát testu nezávislosti v diplomovej práci

Kontext príkladu

V diplomovej práci študent skúma spokojnosť zákazníkov s mobilnou aplikáciou banky. Pomocou dotazníkového prieskumu získal údaje o pohlaví respondentov a celkovej spokojnosti (veľmi spokojný, spokojný, ani spokojný ani nespokojný, nespokojný, veľmi nespokojný). Počet respondentov dotazníka je 408.

V diplomovej práci chce študent zistiť, či sa spokojnosť zákazníkov líši v závislosti od pohlavia.

✗ NEVHODNÁ VERZIA TEXTU DIPLOMOVEJ PRÁCE

Na overenie závislosti medzi pohlavím a spokojnosťou zákazníkov sme sformulovali nasledovné hypotézy:

H0: Spokojnosť zákazníkov s mobilnou aplikáciou banky nesúvisí s pohlavím, teda pohlavie a spokojnosť zákazníkov sú nezávislé premenné.

H1: Spokojnosť zákazníkov s mobilnou aplikáciou banky súvisí s pohlavím, teda pohlavie a spokojnosť zákazníkov sú závislé premenné.

Hladina významnosti $\alpha = 0,05$.

Tabuľka 1 zobrazuje empirické početnosti.

Tabuľka 1 Empirické početnosti

	Veľmi spokojný	Spokojný	Ani/ani	Nespokojný	Veľmi nespokojný	Spolu
Muž	74	94	28	8	4	208
Žena	58	94	34	9	5	200
Spolu	132	188	62	17	9	400

Z empirických početností v Tabuľke 1 sme vypočítali teoretické početnosti podľa vzorca

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n} \quad (1)$$

Tabuľka 2 uvádza vypočítané teoretické početnosti.

Tabuľka 2 Teoretické početnosti

	Veľmi spokojný	Spokojný	Ani/ani	Nespokojný	Veľmi nespokojný	Spolu
Muž	67,3	95,8	31,6	8,7	4,6	208
Žena	64,7	92,2	30,4	8,3	4,4	200
Spolu	132	186	62	17	11	400

Keďže teoretické početnosti pri kategórii veľmi nespokojných nespĺňajú podmienku $E_{ij} \geq 5$, nemôžeme pokračovať s použitím Chí-kvadrát testu nezávislosti. Zlúčili sme teda kategórie „nespokojný“ a „veľmi nespokojný“ do jednej kategórie. Empirické a teoretické početnosti po tomto zlúčení uvádza tabuľka 3.

Tabuľka 3 Teoretické a empirické početnosti po zlúčení kategórií

pohlavie / spokojnosť		Veľmi spokojný	Spokojný	Aní/ani	Nespokojný alebo veľmi nespokojný	Spolu
Muž	Skutočná početnosť	74	94	28	12	208
	Teoretická početnosť	67,3	95,8	31,6	13,3	208
Žena	Skutočná početnosť	58	94	34	14	200
	Teoretická početnosť	64,7	92,2	30,4	12,7	200
Spolu	Skutočná početnosť	132	188	62	26	408
	Teoretická početnosť	132,0	188,0	62,0	26,0	408

Po zlúčení kategórií už všetky teoretické početnosti spĺňajú podmienku $E_{ij} \geq 5$, preto môžeme Chí-kvadrát test použiť. Výsledky Chí-kvadrát testu nezávislosti uvádza Tabuľka 4.

Tabuľka 4 Chí-kvadrát test nezávislosti pohlavia a spokojnosti

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2,518 ^a	3	,472
Likelihood Ratio	2,523	3	,472
N of Valid Cases	408		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 12,75.

Záver: Keďže $p = 0,472 > 0,05$, nezamietame nulovú hypotézu H_0 na hladine významnosti $\alpha = 0,05$.

Pre overenie výsledku sme silu závislosti vypočítali pomocou Cramérovho V a testu jeho významnosti. Výsledky sú v Tabuľke 5.

Tabuľka 5 Výpočet koeficientu Cramérovo V

		Value	Approximate Significance
Nominal by Nominal	Phi	,079	,472
	Cramer's V	,079	,472
N of Valid Cases		408	

Cramérovo $V = 0,079$. Keďže hodnota Cramérovho V je $0,079$, čo je menšie ako $0,3$, ide o slabú závislosť. Pre overenie významnosti Cramérovho V testujeme hypotézy:

H_0 : Koeficient Cramérovo V je štatisticky nevýznamný.

H_1 : Koeficient Cramérovo V je štatisticky významný.

Hladina významnosti $\alpha = 0,05$.

Keďže p -hodnota testu je $p = 0,472$, nezamietame nulovú hypotézu. Koeficient Cramérovo V je štatisticky nevýznamný.

Následne sme realizovali znamienkovú schému pre zistenie podrobností o závislosti pohlavia a spokojnosti. Pre tento účel sme si vypočítali upravené štandardizované reziduály. Tieto hodnoty sú v Tabuľke 5.

Tabuľka 5 Hodnoty upravených štandardizovaných reziduálov

Adjusted Residual		Veľmi spokojný	Spokojný	Ani/ani	Nespokojný alebo veľmi nespokojný
Pohlavie	Muž	1,4	-,4	-1,0	-,5
	Žena	-1,4	,4	1,0	,5

Pomocou skriptu pre znamienkovú schému sme v SPSS dostali výsledky, uvedené v Tabuľke 6.

Tabuľka 6 Znamienková schéma

Adjusted Residual		Veľmi spokojný	Spokojný	Ani/ani	Nespokojný alebo veľmi nespokojný
Pohlavie	Muž	0	0	0	0
	Žena	0	0	0	0

Keďže sme dostali len značky **0**, medzi žiadnymi dvojicami pohlavia a stupňa spokojnosti nie je závislosť.

✗ ČO JE ZLÉ NA TEJTO VERZII?

Hlavný problém: spôsob prezentácie výsledkov štatistických analýz – zbytočný štatistický rituál

- Hypotézy H_0 a H_1 sú formulované v štatistickom jazyku, ktorý čitateľa diplomovky nezaujímajú. Píšeme diplomovú prácu, nie semestrálnu prácu zo štatistickej analýzy. Vecná otázka je „Závisí spokojnosť od pohlavia?“ – nie „Sú premenné nezávislé v zmysle Chí-kvadrát testu?“ **Hypotézy treba formulovať s dôrazom na vecný význam, radšej ako výskumné hypotézy, ktoré sa overujú pomocou testovania štatistických hypotéz.** Štatistické hypotézy nie je nutné uvádzať osobitne so zvláštnym dôrazom na ne; alternatívnu hypotézu často nie je potrebné uvádzať vôbec.
- Pri viacnásobnom použití toho istého testu (napríklad testu významnosti koeficientu) nie je nutné dookola opakovať štatistické hypotézy H_0 aj H_1 . Ak takéto hypotézy chceme vôbec uviesť, tak v metodologickej časti, nie vo výsledkoch. A úplne postačuje jedenkrát uviesť, že pre každý test významnosti, realizovaný v tejto diplomovej práci, bola nulová hypotéza formulovaná tak, že daný koeficient alebo model je štatisticky nevýznamný.
- Hladina významnosti $\alpha = 0,05$ je opakovaná pri každom jednom použití štatistického testu. Je to vstupná konvencia, nie výsledok. Ak ju vôbec uvedieme, patrí do metodologickej časti práce, nie do výsledkov. Do metodologickej časti práce treba uviesť, že pre všetky štatistické testy v tejto diplomovej práci bola využitá hladina významnosti (napríklad) 0,05. Vo výsledkovej časti práce ju už len používame.
- Záver „nezamietame H_0 “ je pre čitateľa prázdny. Čo to znamená? **Treba povedať, čo sa zistilo o vzťahu medzi pohlavím a spokojnosťou – nie čo sa stalo s nulovou hypotézou.**

Interpretačné problémy – chýba vecný obsah

- Záver „pohlavie nemá vplyv na spokojnosť“ je príliš všeobecný bez ďalšieho vysvetlenia. Cramérovho V je interpretované ako „slabá závislosť“ na základe prahovej klasifikácie bez akéhokoľvek kontextu. Čo to znamená pre tému práce? Je to očakávané? Prečo to zrejme vyšlo? Čo iné by mohlo spokojnosť vysvetľovať?
- Výsledok nie je prepojený s témou práce – text sa skončí štatistickým záverom a čitateľ sa nedozvie, aký je vecný význam tejto štatistickej analýzy a čo z toho pre ďalší výskum vyplýva.

Ďalšie problémy: zbytočná prezentácia rozsiahlych tabuliek v texte diplomovej práce

- Tabuľka teoretických početností čitateľa nezaujímajú a neprinesie mu žiadne hodnotné informácie. Služí len na overenie podmienky a na výpočet samotného Chí-kvadrát testu. Takúto tabuľku stačí uviesť v prílohe diplomovej práce.
- Pri prípadnom zlučovaní kategórií nie je potrebné znovu uvádzať oba typy početností (v jednej spoločnej tabuľke alebo dokonca v dvoch osobitných tabuľkách). Stačí v texte vysvetliť, prečo sa zlučovanie realizovalo a čo sa zlúčilo.
- Z tabuľky kontingenčných koeficientov sa reálne použije len hodnota Cramérovho V a p -hodnota testu jeho významnosti. Tieto dve hodnoty úplne postačuje uviesť v texte, nie je potrebná osobitná tabuľka.

- Tabuľku upravených štandardizovaných reziduálov nie je vôbec potrebné uvádzať (alebo len v prílohe), neprinesie čitateľovi žiadnu hodnotnú informáciu. Používa sa len na stanovenie značiek v znamienkovej schéme.

Formálne problémy:

- Prezentácia tabuliek takmer v takej podobe, ako boli vo výstupe procedúry v SPSS Statistics. Tabuľky treba upraviť: nepotrebné zlúčené bunky alebo riadky s hodnotami, ktoré nepoužívame, odstrániť.
- Texty v angličtine vo výstupoch z SPSS Statistics je nutné preložiť do slovenčiny.
- Pri desatinných číslach chýba nula pred desatinnou čiarkou – výstupy z SPSS Statistics túto nulu pred desatinnou čiarkou neobsahujú; je nutné ich doplniť.
- Uvádzanie vzorcov pre výpočet, ktoré už boli (alebo mali byť!) v metodologickej časti práce. Vo výsledkovej časti práce sa na ne treba len odvolať pomocou čísla vzorca (na to sa práve vzorce číslujú!).

✓ LEPŠIA VERZIA TEXTU DIPLOVEJ PRÁCE

Jedným z cieľov prieskumu bolo zistiť, či sa spokojnosť s mobilnou aplikáciou líši medzi mužmi a ženami. Na overenie stanovenej výskumnej hypotézy č. XY sme použili Chí-kvadrát test nezávislosti, ktorý testuje štatistickú hypotézu o nezávislosti pohlavia a spokojnosti zákazníkov. Početnosti pre jednotlivé kombinácie pohlavia a spokojnosti sú prezentované v Tabuľke 1.

Tabuľka 1 Empirické početnosti

	Veľmi spokojný	Spokojný	Ani/ani	Nespokojný	Veľmi nespokojný	Spolu
Muž	74	92	28	8	6	208
Žena	58	94	34	9	5	200
Spolu	132	186	62	17	11	400

Tabuľka teoretických početností, vypočítaných podľa vzorca (1), je uvedená v Prílohe A. Na základe týchto teoretických početností sme pred samotným testom overili podmienku minimálnych teoretických početností: kategórie *nespokojný* a *veľmi nespokojný* obsahovali len 5–9 respondentov v jednotlivých pohlaviach, čo spôsobilo pokles niektorých teoretických početností pod hranicu 5. Z tohto dôvodu sme tieto dve kategórie zlúčili do jednej (*nespokojný alebo veľmi nespokojný*) a test sme realizovali na štvorstupňovej škále. Upravená kontingenčná tabuľka je uvedená v Prílohe A.

Výsledky testovania nepreukázali štatisticky významný vzťah medzi pohlavím a spokojnosťou zákazníkov (p -hodnota testu = 0,472). Tento výsledok potvrdzuje aj hodnota Cramérovho $V = 0,079$, ktorá potenciálnu závislosť kvantifikuje ako veľmi miernu, avšak štatisticky nevýznamnú (p -hodnota testu významnosti = 0,472). Kompletná tabuľka výsledkov Chí-kvadrát testu, kontingenčných koeficientov spolu s testami ich štatistickej významnosti sa nachádza v Prílohe A. Vzhľadom na tieto výsledky ani znamienková schéma neodhalila významné vzťahy medzi jednotlivými kombináciami pohlavia a stupňa spokojnosti zákazníka.

Miera spokojnosti je teda medzi mužmi a ženami prakticky porovnateľná a pohlavie nie je premennou, ktorá by diferencovala zákaznícku skúsenosť s aplikáciou. Faktory vysvetľujúce variabilitu spokojnosti treba hľadať inde. Na základe deskriptívnej analýzy (pozri kapitolu XY) sa ako sľubné kandidáty javia frekvencia používania aplikácie a veková skupina respondentov, ktorým sa venujú nasledujúce analýzy.

✓ Čo je správne na tejto verzii?

Štatistická analýza je nástroj, nie cieľ!

- Metodologické rozhodnutie (zlúčenie kategórií) je vysvetlené jednou vetou s vecným dôvodom – nie ako byrokratický úkon, ale ako súčasť analytického uvažovania.
- Výsledky sú prezentované s veľkým dôrazom na ich vecný význam; štatistika je len nástrojom na získanie týchto výsledkov; p -hodnoty a hodnoty koeficientov sú len v zátvorke na potvrdenie formulovanej interpretácie výsledku.

- Podrobný postup realizácie výpočtov nie je uvádzaný, čitateľa nezaujíma, dôležitý je výsledok a jeho vecný význam. **Neprezentujeme, ako sme štatistické nástroje používali, ako poznáme postup výpočtu a čo nás pri tom postretlo, ale čo sme pomocou štatistických nástrojov zistili o skúmaných javoch.**
- Cramérovo V nie je len číslo s prahom, ale potvrdenie vecného záveru.
- Posledný odsek posúva text dopredu – negatívny výsledok nie je slepá ulička, ale navigácia k ďalším analýzám.

Zhrnutie odporúčaní

✘ NIE	✔ ÁNO
Uvádzať H0 a H1 v štatistickom znení	Formulovať výskumnú otázku vecne a zrozumiteľne
Zapisovať $\alpha = 0,05$ osobitne pred testom	Uviesť p-hodnotu pri konkrétnom výsledku v zátvorke
Vkladať rozsiahle tabuľku do textu	Tabuľky výpočtu do prílohy; v texte len dôležité tabuľky.
Záver: „zamietame / nezamietame H0“	Záver: čo to znamená vecne pre tému práce
Interpretácie uvádzať len ako štatistické rozhodnutia, opakovať tie isté formulácie dookola.	Interpretovať výsledky v kontexte: čo nám daný výsledok hovorí o sile vzťahu a čo z neho vyplýva?
Ignorovať praktický (vecný) význam výsledku	Komentovať, či je štatisticky významný efekt aj prakticky zaujímavý

Príklad 2 (už iba v skrátenej verzii): Lineárna regresia

Kontext príkladu

Študent skúma, čo ovplyvňuje mesačné výdavky domácností, v dátach má údaje o 150 náhodne vybraných domácnostiach. Závislá premenná: mesačné výdavky (€). Vysvetľujúce premenné: čistý mesačný príjem (€), región bydliska – tri kategórie: Západ (referenčná), Stred, Východ a počet členov domácnosti – štyri kategórie: jeden člen (referenčná), dvaja členovia, traja alebo štyria členovia, päť a viac členov.

✗ NEVHODNÁ VERZIA

Pre vytvorenie regresného modelu sme si najprv prekódovali kategoriálne premenné región a počet členov domácnosti na sústavu dummy premenných. Zvolené referenčné kategórie sú: región_západ a počet členov domácnosti_jeden.

Výsledky regresnej analýzy sú uvedené v nasledujúcich tabuľkách. Tabuľka 1 uvádza charakteristiky vytvoreného regresného modelu.

Tabuľka 1 Charakteristiky vytvoreného regresného modelu

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,842	0,709	0,697	90,74

Pre overenie štatistickej významnosti modelu ako celku sme sformulovali hypotézy:

H0: Model ako celok je štatisticky nevýznamný (všetky regresné koeficienty sú rovné nule).

H1: Model ako celok je štatisticky významný (aspoň jeden regresný koeficient je nenulový).

Hladina významnosti: $\alpha = 0,05$

Výsledky testu významnosti modelu ako celku uvádza Tabuľka 2.

Tabuľka 2 ANOVA test významnosti modelu ako celku

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	2 866 556	6	477 759	58,02	< 0,001
Residual	1 177 503	143	8 234		
Total	4 044 059	149			

Záver: Keďže $p < 0,001 < \alpha = 0,05$, zamietame nulovú hypotézu H0. Model ako celok je štatisticky významný. Tabuľka 3 uvádza koeficienty vytvoreného modelu.

Tabuľka 3 Regresné koeficienty

Premenná	B	Std. Error	Beta	t	Sig.
(Constant)	191,57	42,99		4,457	< 0,001
Príjem (€)	0,41	0,02	0,721	16,609	< 0,001
Reg_Stred	-55,53	28,09	-0,099	-1,977	0,049
Reg_Východ	-68,00	39,27	-0,091	-1,732	0,086
Členovia_2	74,31	26,20	0,127	2,836	0,005

Členovia_3až4	155,67	36,70	0,309	4,242	< 0,001
Členovia_5aviac	242,45	44,77	0,320	5,416	< 0,001

Hodnota konštanty je 180,2. To znamená, domácnosť, ktorá má teoreticky nulový príjem, je zo západného Slovenska a má jedného člena, by mala výšku mesačných výdavkov 191,57 €.

Koeficient premennej príjem je 0,41, čo znamená, že s rastom príjmu o 1 € sa mesačné výdavky zvýšia o 0,41 € za podmienky ceteris paribus (domácnosť pochádza z rovnakého regiónu a má rovnaký počet členov).

Koeficient premennej Reg_Stred je -55,53, čo znamená, že domácnosti zo Stredného Slovenska majú mesačné výdavky o 55,53 € nižšie oproti domácnostiam zo Západného Slovenska (referenčná kategória) za podmienky ceteris paribus (domácnosť má rovnaký príjem a rovnaký počet členov).

Koeficient premennej Reg_Východ je -68,00, čo znamená, že domácnosti z Východného Slovenska majú mesačné výdavky o 68,00 € nižšie oproti domácnostiam zo Západného Slovenska (referenčná kategória) za podmienky ceteris paribus (domácnosť má rovnaký príjem a rovnaký počet členov).

Koeficient premennej členovia_2 je 74,31, čo znamená, že dvojčlenné domácnosti majú mesačné výdavky o 74,31 € vyššie oproti jednočlenným domácnostiam (referenčná kategória) za podmienky ceteris paribus (domácnosť má rovnaký príjem a pochádza z rovnakého regiónu).

Koeficient premennej členovia_3až4 je 155,67, čo znamená, že domácnosti s 3 až 4 členmi majú mesačné výdavky o 155,67 € vyššie oproti jednočlenným domácnostiam (referenčná kategória) za podmienky ceteris paribus (domácnosť má rovnaký príjem a pochádza z rovnakého regiónu).

Koeficient premennej členovia_5aviac je 242,45, čo znamená, že domácnosti s 5 a viac členmi majú mesačné výdavky o 242,45 € vyššie oproti jednočlenným domácnostiam (referenčná kategória) za podmienky ceteris paribus (domácnosť má rovnaký príjem a pochádza z rovnakého regiónu).

Pre testovanie štatistickej významnosti sme si stanovili hypotézy:

H0: Koeficient je štatisticky nevýznamný.

H1: Koeficient je štatisticky významný.

Hladina významnosti $\alpha = 0,05$.

Keďže pre takmer všetky p-hodnoty v Tabuľke 3 (okrem Reg_Východ) $p < 0,05$, zamietame nulovú hypotézu. Koeficienty sú štatisticky významné.

✓ LEPŠIA VERZIA

Na identifikáciu faktorov ovplyvňujúcich mesačné výdavky domácností sme odhadli lineárny regresný model, ktorého vysvetľujúcimi premennými sú čistý mesačný príjem, región bydliska (ako referenčná kategória bol zvolený západ) a počet členov domácnosti (referenčná kategória: 1 člen domácnosti). Obe kategoriálne premenné sú do modelu zaradené prostredníctvom dummy premenných. Tabuľka charakteristík modelu a kompletná tabuľka ANOVA testu pre model ako celok sú v Prílohe B. Regresné koeficienty sú uvedené v Tabuľka 1.

Tabuľka 1 Regresné koeficienty – determinanty mesačných výdavkov domácností

Premenná	B	Štd. chyba	Beta	t	p
(Konštanta)	191,57	42,99		4,457	< 0,001
Príjem (€)	0,41	0,02	0,721	16,609	< 0,001
Reg_Stred	-55,53	28,09	-0,099	-1,977	0,049
Reg_Východ	-68,00	39,27	-0,091	-1,732	0,086
Členovia_2	74,31	26,20	0,127	2,836	0,005
Členovia_3až4	155,67	36,70	0,309	4,242	< 0,001
Členovia_5aviac	242,45	44,77	0,320	5,416	< 0,001

Model vysvetľuje 71 % variability mesačných výdavkov ($R^2 = 0,71$), čo predstavuje dobrú vysvetľovaciu silu. Celkový model je štatisticky významný ($p < 0,001$).

Príjem je najsilnejším prediktorom výdavkov: každých 100 € čistého príjmu navyše je spojených s rastom mesačných výdavkov o cca 43 € ($b = 0,43$; $p < 0,001$) za predpokladu fixných ostatných faktorov. Domácnosti teda míňajú zhruba 43 centov z každého eura príjmu navyše – čo je konzistentné s ekonomickou teóriou marginálneho sklonu k spotrebe a naznačuje, že väčšia časť príjmového rastu smeruje do úspor alebo splátok.

Región bydliska má samostatný štatisticky významný vplyv aj po kontrole príjmu a veľkosti domácnosti. Domácnosti zo Stredného Slovenska vykazujú pri rovnakom príjme výdavky priemerne o takmer 56 € nižšie ako domácnosti zo Západu ($p = 0,049$), je však nutné poznamenať, že tento výsledok je na hranici štatistickej významnosti a bolo by ho potrebné potvrdiť pomocou rozsiahlejšieho súboru dát. Domácnosti z Východu vykazujú pri rovnakom príjme a počte členov domácnosti výdavky priemerne o 68 € nižšie. Na zvolenej hladine významnosti 0,05 je však tento rozdiel štatisticky nevýznamný ($p = 0,086$) a opäť by ho bolo treba overiť na rozsiahlejšej vzorke domácností na Slovensku. Tieto regionálne rozdiely pravdepodobne odrážajú nižšiu cenovú hladinu a odlišné spotrebné vzorce mimo západného regiónu; presnejšie určenie príčin by si vyžadovalo zaradenie ďalších kontrolných premenných.

Počet členov domácnosti sa ukázal ako štatisticky aj vecne významný determinant výdavkov. Oproti jednočlenným domácnostiam míňajú dvojčlenné domácnosti pri rovnakom príjme a regióne v priemere o 74 € mesačne viac ($p = 0,005$), domácnosti s 3 až 4 členmi o takmer 156 € viac ($p < 0,001$) a domácnosti s 5 a viac členmi až o 242 € viac ($p < 0,001$). Výdavky teda s veľkosťou domácnosti rastú, no nie proporcionálne – kým prechod z jedného na dvoch členov pridá 74 €, každý ďalší člen v kategórii 3–4 pridá v priemere zhruba 40–50 € a v najväčších domácnostiach sa prírastok na jedného člena ďalej znižuje. Tento vzorec je konzistentný s existenciou úspor z rozsahu v spotrebiteľskom správaní:

spoločné náklady na bývanie, energie či domácnosť sa rozkladajú na viac členov, čo znižuje priemerný výdavok na osobu napriek tomu, že celkové výdavky domácnosti rastú.

Zhrnutie pravidiel – lineárna regresia

✗ NIE	✓ ÁNO
Vkladať všetky tabuľku výstupov regresnej analýzy do textu	R^2 uviesť jednou vetou v texte s interpretáciou; tabuľka do prílohy
Uvádzať testy významnosti modelu a koeficientov premenných s H_0/H_1 rituálom	p-hodnotu stačí uviesť v zátvorke; tabuľka ANOVA do prílohy, výsledky opísať len v texte
Záver: "model je štatisticky významný"	Záver: čo model vysvetľuje, ako dobre, čo z toho plynie
Interpretovať každý koeficient rovnakým vzorom	Interpretovať vecne s kontextom a hodnotením potenciálneho významu danej premennej
Ignorovať výsledky, ktoré sú na hranici štatistickej významnosti	Vyjadriť sa aj k výsledkom, ktoré sa nepotvrdili štatistickou analýzou

Štatistika je nástroj, nie cieľ.