Obsah

Vědecké a odborné články

Zdeněk Půlpán
Posouzení homogenity dotazníku 3

Petr Volf
Zpráva o Statistických dnech konaných 10. – 11. září v Olomouci 16

Zprávy a informace

Jaromír Antoch
ROBUST 2026 19

Ondřej Vencálek
Robust 2026 – nejen robustní metody, ale také sníh, těžiště, toulky a zpěv 21

Redakce
Pozvánka na Energy Days 2026 24

Pavel Stříž
Pozvánka na překonání rekordu 25

Redakce
Pozvánka na OSSConf 2027 26

Informační bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214. Časopis je sázen v programu TEX, ve formátu LuaHBTEX s písmy balíku CSfonts.

The Information Bulletin of the Czech Statistical Society is published quarterly.
The contributions in the journal are published in English, Czech and Slovak languages.

Předsedkyně společnosti: Ing. Martina Litschmannová, Ph.D., Katedra aplikované matematiky, Fakulta elektrotechniky a informatiky, Vysoká škola báňská – Technická univerzita Ostrava, 17. listopadu 2172/15, 708 33 Ostrava-Poruba, Martina.Litschmannova@vsb.cz.

Redakce: prof. RNDr. Gejza Dohnal, CSc. (šéfredaktor), prof. RNDr. Jaromír Antoch, CSc., RNDr. Marek Malý, CSc., doc. RNDr. Jiří Michálek, CSc., prof. Ing. Jiří Militký, CSc., doc. Ing. Iveta Stankovičová, PhD., doc. Mgr. Ondřej Vencálek, Ph.D.

Redaktor časopisu: doc. Mgr. Ondřej Vencálek, Ph.D., ondrej.vencalek@upol.cz.
Informace pro autory jsou na stránkách společnosti, http://www.statspol.cz/.

DOI: 10.5300/IB, http://dx.doi.org/10.5300/IB
ISSN 1804–8617 (Online)

INFORMAČNÍ BULLETIN

České statistické společnosti

Ročník 37, číslo 3, září 2026

POZVÁNKA NA OSSCONF 2027
INVITATION TO OSSCONF 2027

Redakce

Open-source software ve vzdělávání, výzkumu & vývoji a v IT systémech

Fakulta řízení a informatiky Žilinské univerzity
ve spolupráci se **Společností pro otevřené informační technologie** (SOIT), sdruženími **CSTUG**, **Česká statistická společnost**, **Geokomunita**

pořádají již 15. ročník konference

OSSConf 2027

Konference se uskuteční v prostorách Fakulty řízení a informatiky v Žilině na Slovensku ve dnech

30. června – 2. července 2027

Cílem naší konference je poskytnout prostor pro informování o novinkách ve vývoji otevřeného softwaru a otevřených technologií, o možnostech využití těchto nástrojů ve vědě a vzdělávání a také poskytnout prostor pro neformální přátelské setkání uživatelů a příznivců otevřených informačních technologií.

Při příležitosti konference bude vydán sborník příspěvků. Příspěvky autorů přijímáme buď jako nerecenzované příspěvky, z nichž uveřejňujeme pouze abstrakt, nebo jako recenzované příspěvky, které budou posuzovány dvěma nezávislými recenzenty. Předepsaná šablona pro psaní příspěvků v typografickém systému TEX/LATEX je k dispozici na webové stránce konference: `https://ossconf.fri.uniza.sk/pre-autorov/`. Příspěvky přijímáme v českém, slovenském nebo anglickém jazyce.

Důležité termíny

Recenzované příspěvky: do 10. června 2027.
Abstrakty nerecenzovaných vystoupení: do 20. června 2027.

Sekce

Open AI | **LATEX, R a jejich přátelé** | **Vývoj OSS**
OSS ve vzdělávání | **Open GIS & Open Data** | **Open Hardware**
Kyberbezpečnost

Další informace o konferenci budou postupně zveřejňovány na `https://ossconf.fri.uniza.sk/`.

POSOUZENÍ HOMOGENITY DOTAZNÍKU
ASSESSMENT OF QUESTIONNAIRE HOMOGENEITY

Zdeněk Půlpán

Adresa: Na Brně 1952, 500 09 Hradec Králové 9

E-mail: `zdenek.pulpan@post.cz`

Abstrakt: Dotazník (případně test) se považuje za homogenní, když pravděpodobnost správné odpovědi na každou z položek je ve zkoumané populaci stejná. Jsou diskutovány některé modely pro posouzení homogenity dotazníkových položek vzhledem k určité populaci.

Klíčová slova: homogenita dotazníkových položek, odhad míry homogenity dotazníkových položek.

Abstract: A questionnaire (or test) is considered homogeneous when the probability of a correct answer to each item is the same in the studied population. Some models are discussed for assessing the homogeneity of questionnaire items with respect to a certain population.

Keywords: homogeneity of questionnaire items, estimation of the degree of homogeneity of questionnaire items.

1. Úvod

V celém článku budeme označovat slovem dotazník (případně i slovem test, bude-li dotazník sofistikovaně konstruován) výhradně výkonový dotazník, který se také označuje jako CR test (criterion referenced test). Ukážeme některé možnosti odhadu míry homogenity jeho položek. V situaci, kdy je velká snaha po automatizovaném zpracování výsledků dotazníkových šetření (zvláště u vědomostních testů), je nutné předem rozhodnout, jak zajistit objektivitu skórování. Dotazníky mohou mít (krom jiného) rozdílnou vzájemnou úroveň svých položek. Úroveň položek je odhadována hlavně empiricky vzhledem k testované populaci pomocí relativních četností správných odpovědí ve výběrovém souboru. Tyto relativní četnosti se považují za nestranné odhady pravděpodobností správných odpovědí na jednotlivé položky. Na druhou stranu se pravděpodobnosti správných odpovědí považují (interpretují) také jako pravděpodobnosti správných odpovědí „průměrného respondenta“. Položka je tím obtížnější, čím menší je pravděpodobnost správné reakce „průměrného respondenta“ na položku, případně je tím snadnější, čím je tato

pravděpodobnost větší. Dotazníkové položky (nebo krátce dotazník) se *považují za homogenní*, když mají stejnou obtížnost, tj. pravděpodobnosti správných odpovědí se od sebe „moc“ neliší. [2] Rozdíly mezi relativními četnostmi je třeba posoudit statistickou cestou, protože základem pro odhad míry homogenity položek jsou pravděpodobnosti správných odpovědí. Kdy relativní četnosti správných (nebo nesprávných) odpovědí mohou svědčit o postačující homogenitě položek a kdy již tento předpoklad musíme opustit? K tomuto cíli vede řada i sofistikovaných statistických metod jako je např. analýza rozptylu, metody hlavních komponent, … Každá z uvedených metod má trochu jinou základní filozofii. Problém pak vzniká v pohledu na interpretaci rozdílných výsledků rozdílných metod. Podle mého názoru, jednodušší metody (pokud aspoň trochu vyhovují požadavku postačujícího řešení problému) dávají snadnější návod pro interpretaci.

Ještě se zmiňme o tom, proč je důležité mít jistotu, zda dotazník je možné považovat za homogenní. Je-li dotazník homogenní, lze za výsledný dotazníkový skór považovat tzv. hrubý skór, což je celkový počet jeho správných odpovědí (rozhodnutí o tom, že na položku se správně odpovědělo, je velmi často formální, automatizované). Není-li dotazník homogenní, pak je z důvodu vhodné interpretace nutné za výsledný skór respondenta považovat tzv. vážený skór, kdy každé správně zodpovězené položce přiřazujeme váhu, která je úměrná například obtížnosti položky.

Ukážeme jednoduché statistické techniky jak ověřit, zda mohou být položky považovány za homogenní či je třeba předpoklad homogenity zamítnout.

Obsah tohoto příspěvku navazuje na článek [1].

2. První model

Předpokládejme, že máme dotazník s n položkami, které skórujeme pouze dichotomicky: 0 – 1, podle předem daných kritérií (0 například ve významu: neuspěl, nevyřešil, …, 1 ve významu opačném: uspěl, vyřešil, …). Součástí metodiky skórování celého dotazníku je zjistit, zda dotazníkové položky jsou vzhledem ke zkoumané populaci přibližně stejné úrovně.

Jak jsme se již zmínili, úroveň i-té položky může být definována pomocí pravděpodobnosti p_i odpovědi se skórem 1 (s uvedenou definicí úrovně položky budeme dále pracovat). Tuto pravděpodobnost odhadneme z relativní četnosti $\widehat{p}_i$ reakcí na i-tou položku ($i = 1, 2, \ldots, n$) odpovídající skóru 1 v reprezentativním výběrovém souboru rozsahu r. Pokud se budou odhady pravděpodobností $\widehat{p}_i$ pro různá i „významně“ lišit, nebudou položky považovány z hlediska skórování za rovnocenné. Z tohoto hlediska budeme také

POZVÁNKA NA PŘEKONÁNÍ REKORDU
INVITATION TO A RECORD-BREAKING ATTEMPT

Pavel Stříž

E-mail: pavel@striz.cz

POZVÁNKA NA ENERGY DAYS 2026
INVITATION TO ENERGY DAYS 2026

Redakce

5. až 7. listopadu 2026, Praha

Zveme vás na workshop Energy Days 2026, který propojí odborníky na matematické a statistické modelování s pracovníky energetiky. Letošním ústředním tématem je modelování sítí: účinné využití jejich kapacity, řízení rizik, sdílení energie a optimalizace provozu. Setkání nabídne prostor pro výměnu zkušeností mezi výzkumem a praxí i pro diskusi o otázkách, které přináší proměna energetických systémů.

Elektrizační soustavy se potýkají s novými nároky. Přibývají obnovitelné zdroje, decentralizovaná výroba a nové způsoby spotřeby; roste význam bezpečného a odolného provozu. Při plánování i každodenním řízení sítí je proto nutné rozhodovat s ohledem na omezenou kapacitu a nejistotu. Právě zde mohou pomoci matematické modely, statistické metody, výpočetní postupy a využití dat.

Program se zaměří na modelování složitých energetických sítí, efektivní využití přenosových a distribučních soustav, sdílení kapacity, řízení přetížení a zapojení flexibility. Dalšími tématy jsou hodnocení rizik, kvantifikace nejistoty, odolnost propojených soustav a optimalizace plánování i provozu. Vítány jsou rovněž přístupy založené na datech a kombinace datových a fyzikálních modelů.

Součástí setkání budou zvané i přihlášené přednášky. Diskuse u kulatého stolu spojí představitele české energetiky s pozvanými evropskými odborníky; jednání proběhne v češtině a angličtině se zajištěným tlumočením. Workshop je příležitostí představit vlastní výsledky, porovnat různé metodické přístupy a navázat spolupráci na úlohách, které mají bezprostřední význam pro energetickou praxi.

Úvodní den 5. listopadu se uskuteční na Matematicko-fyzikální fakultě Univerzity Karlovy na Malostranském náměstí. Dne 6. listopadu bude jednání pokračovat v Českém statistickém úřadě v Praze 10. Závěrečný den 7. listopadu proběhne na Fakultě strojní ČVUT na Karlově náměstí. Podrobný program, místa konání a aktuální organizační informace najdete na webu workshopu.

Web: `https://energy-days.cz/`

hovořit o tom, že položky jsou vzhledem k uvažované populaci nehomogenní. Protože nejjednodušší případ skórování se realizuje za předpokladu homogenity položek, měly by být metody skórování založeny na prvotním odhadu míry homogenity položek.

Znakem X_j^i označíme dichotomickou náhodnou veličinu reprezentující skór j-tého respondenta pro i-tou položku, kde $i = 1, 2, \ldots, n$ a $j = 1, 2, \ldots, r$. Předpokládejme, že

$$P\left(X_j^i = 1\right) = p_i, \quad P\left(X_j^i = 0\right) = 1 - p_i = q_i$$

pro všechna $j = 1, 2, \ldots, r$. Součet $Y_i = \sum_{j=1}^{r} X_j^i$ je pak pro každé $i = 1, 2, \ldots, n$ binomickou náhodnou veličinou $Bi(r, p_i)$, pro kterou je střední hodnota $EY_i = r \cdot p_i$ a rozptyl $DY_i = D\sum_{j=1}^{r} X_j^i = E(Y_i - rp_i)^2 = r \cdot p_i q_i$, protože zřejmě respondenti reagují na položky vzájemně nezávisle (lze předpokládat, že i odpovědi různých respondentů na i-tou položku budou statisticky nezávislé). Rozptyl DY_i charakterizuje rozdílnosti reakce r vybraných respondentů na i-tou položku (ovšem za předpokladu známé pravděpodobnosti p_i). [4]

Pro průměr $p = \frac{\sum_1^n p_i}{n} = \frac{\sum_1^n rp_i}{r \cdot n} = \frac{\sum_1^n EY_i}{r \cdot n} = 1 - q$ lze pak psát $\sum_1^n EY_i = r \cdot n \cdot p$.

Počítejme nyní střední kvadratickou odchylku veličiny Y_i od průměru střední hodnoty $r \cdot p$:

$$\begin{aligned} E(Y_i - rp)^2 &= E(Y_i - rp_i + rp_i - rp)^2 = \\ &= E\left[(Y_i - rp_i)^2 + 2 \cdot (Y_i - rp_i)\,(rp_i - rp) + r^2(p_i - p)^2\right] = \\ &= E(Y_i - rp_i)^2 + 2r \cdot (p_i - p) \cdot E\,(Y_i - rp_i) + r^2(p_i - p)^2 = \\ &= rp_i q_i + r^2(p_i - p)^2. \end{aligned}$$

Celkový rozptyl n položek $E(Y_i - rp)^2$ chápaný jako součet výše uvedených kvadratických odchylek je roven

$$\sum_1^n E(Y_i - rp)^2 = r\sum_1^n p_i q_i + r^2 \sum_1^n (p_i - p)^2. \tag{1}$$

Když je dotazník složen z položek stejné obtížnosti, tj. vykazuje „absolutní“ homogenitu, pak

$$p = p_1 = p_2 = \cdots = p_n \tag{2}$$

a druhý člen zprava z výrazu (1) je roven nule.

Poznámka 1: Pro $p = \frac{\sum_1^n p_i}{n}$ má výraz $\sum_1^n (p_i - p)^2$ minimální hodnotu z možných různých hodnot p při daných pravděpodobnostech p_i. Nulovou hodnotu má však, jen když platí (2).

Blízkost p_i určité konstantní hodnotě (kterou neznáme) můžeme posuzovat jednoduše pomocí variačního koeficientu $\frac{S_n}{p} \geq 0$, kde $S_n = \sqrt{S_n^2}$ a $S_n^2 = \frac{1}{n}\sum_1^n (p_i - p)^2$. Přitom p může být hodnotou, blízkou k předpokládané, tedy buď aritmetický průměr z p_i nebo jejich medián. Čím pak bude $\frac{S_n}{p}$ bližší 0, tím spíše můžeme předpokládat homogenitu dotazníku. □

Míru homogenity položek lze posuzovat z poměru H_1/r, kde H_1 je poměr dvou rozptylů v (1)

$$H_1 = r \cdot \frac{\sum_1^n (p_i - p)^2}{\sum_1^n p_i q_i}, \quad H_1 \geq 0. \tag{3}$$

Čím je ukazatel H_1/r bližší 0, tím spíše můžeme spoléhat na to, že dotazník se blíží dotazníku homogennímu. Z (2) plyne $H_1 = 0$.

Vzhledem k tomu, že teoretické pravděpodobnosti p_i neznáme, můžeme je odhadnout takto:

$$p_i \sim \widehat{p}_i = 1 - \widehat{q}_i = \frac{\sum_j X_j^i}{r}, \quad p \sim \widehat{p} = \frac{\sum_i \widehat{p}_i}{n}.$$

Pak pro odhad veličiny H_1 použijeme číslo $\widehat{H}_1$ určené vztahem (4a):

$$\widehat{H}_1 = r \cdot \frac{\sum_1^n (\widehat{p}_i - \widehat{p})^2}{\sum_1^n \widehat{p}_i \widehat{q}_i} \tag{4a}$$

Odhad $\widehat{p}$ můžeme určit místo z aritmetického průměru $\frac{\sum_i \widehat{p}_i}{n}$ i řešením kvadratické rovnice $\frac{1}{n}\sum_1^n \widehat{p}_i\widehat{q}_i = \frac{A}{n} = \widehat{p}\cdot(1-\widehat{p})$ za předpokladu, že $\frac{A}{n} \leq \frac{1}{4}$ ve tvaru $\widehat{p}_{01,02} = \frac{1 \pm \sqrt{1 - 4.\frac{A}{n}}}{2}$, kde $\widehat{p}_{01} = 1 - \widehat{p}_{02}$. Tato rovnice na pravé straně obsahuje odhad rozptylu Y_i pro $i = 1, \ldots, n$, když předpokládáme homogenitu, tj. platnost (2). Tento odhad je učiněn jako průměr odhadů rozptylů jednotlivých Y_i (viz levá strana). Máme dvě řešení, a musíme rozhodnout (což pro odhad kritéria není důležité), které je odhadem p a které odhadem q. Použijeme pak předchozí aproximativní vztah (4a), ale s odhadem $\widehat{p} = \widehat{p}_{01}$, resp. $\widehat{p} = \widehat{p}_{02}$ (tedy s jiným než je ve (4a)) ve vzorci

$$\widehat{H}_1 = r \cdot \frac{\sum_1^n (\widehat{p}_i - \widehat{p})^2}{n \cdot \widehat{p} \cdot (1 - \widehat{p})} \tag{4b}$$

u vzniku její bakalářské práce, ze které příspěvek vycházel, jsem jako školitel byl přítomen. Vlastně to byl pro mě docela nový pocit – vůbec poprvé se stalo, že by byla oceněna práce, u jejíhož vzniku jsem byl v roli vedoucího. A musím říct, že to je ještě daleko lepší pocit, než když byl před léty v Králíkách oceněn můj vlastní poster. Může se mi pak někdo divit, že mi bylo do zpěvu? A tak, když vzal Martin Romaňák do ruky kytaru, přidal jsem se ke skupince, která zpívala ještě dlouho po půlnoci toho nezapomenutelného lednového čtvrtečního večera.

však nutno dodat, že druhý středobod se nachází doslova „hned za dveřmi“ – je jím jídelna, kde se kromě konzumace vybraných pokrmů odehrávají četná neformální (přátelská) setkání. Právě tato setkání jsou druhým, neméně významným, těžištěm Robustů.

Organizátoři při sestavování programu dbají na to, aby oba aspekty konference – odborný a společenský – byly v rovnováze. Program využívá ranní svěžesti účastníků, kterým hned po snídaňovém menu nabízí první vydatný chod přednášek, následovaný vydatným obědem, a … teď přijde to kouzlo … dlouhou poobědní pauzou. Tu je možno využít k všemožným sportovním a společenským aktivitám. Kouzlo spočívá v tom, že pak člověk opět s chutí usedá k podvečerní dávce přednášek, následované večeří a volným večerním programem.

Šumavské toulky

Měl jsem to štěstí, že jsem s přáteli využil všechny poobědové pauzy k toulkám po okolích šumavských hvozdech – v pondělí jsme objevovali Hauswaldskou kapli, ve středu se prošli z Modravy kolem řeky Vydry až zpět do Srní a ve čtvrtek došli k zamrzlé kráse na Čeňkově pile. Každý z těch výletů byl jedinečný. A pak tu byl životní zážitek z úterního odpoledne, kdy mě mé milé kolegyně postavily na běžky a nějakým zázrakem jsem se (zcela začátečnickou technikou a jen a pouze díky jejich vydatnému povzbuzování) vydrápal z Modravy na Březník a díky dobře namazaným zapůjčeným běžkám jsem se pak i v pořádku dostal zpět.

Dokud se zpívá…

Program Robustu, chápeme-li tuto konferenci komplexně jako událost se dvěma těžišti, vydrží jen robustní statistici. I přes celotýdenní náročný program mají ale někteří ještě ve čtvrtek večer dost sil, aby zvládli i vrchol společenské části konference. Čtvrteční „závěrečný“ večer bývá slavnostní. My zkušení účastníci Robustu se těšíme, až Zdeněk Fabián s kytarou v ruce zahraje „Járu“. Napjatě očekáváme zejména novou sloku věnovanou právě se končícímu ročníku. A Zdeněk nezklamal ani tentokrát a umně připomněl, že hlavní organizátor Jára[1] Antoch je opravdový šaman, tedy kouzelník.

Byl to pak právě hlavní organizátor, který účastníky Robustu seznámil s výsledky hodnocení odborných příspěvků studentů a doktorandů. Mně osobně udělalo největší radost ocenění práce Elizabeth Princové, mj. protože

[1] Prof. RNDr. Jaromír Antoch, CSc. (v textu použitá domácí podoba křestního jména odkazuje na text Fabiánovy písně)

Dostaneme tak dva možné odhady $\widehat{H}_1(\widehat{p}_{01})$, resp. $\widehat{H}_1(\widehat{p}_{02})$. Výslednou kriteriální hodnotou je normovaná hodnota $\min\limits_{i=1,2}\left\{\widehat{H}_1(\widehat{p}_{0i})/r\right\}$.

3. Druhý model

Předchozí postup můžeme ještě upravit tak, že ve vztahu (1) součin p_iq_i přepíšeme takto:

$$p_iq_i = [p+(p_i-p)]\,.\,[1-(p+(p_i-p))] = [p+(p_i-p)]\,.\,[q-(p_i-p)] = \\ = pq-(p_i-p)\,(p-q)-(p_i-p)^2.$$

Pro součet $\sum_1^n p_iq_i$ pak bude platit $\sum_1^n p_iq_i = npq-(p-q)\sum_1^n(p_i-p)-\sum_1^n(p_i-p)^2 = npq-\sum_1^n(p_i-p)^2$, protože $\sum_1^n(p_i-p)=0$.

Dosazením z předchozího vztahu do (1) dostaneme

$$\sum_1^n E(Y_i-rp)^2 = r\cdot npq - r\cdot\sum_1^n(p_i-p)^2 + r^2\cdot\sum_1^n(p_i-p)^2 \\ = rnpq + r(r-1)\sum_1^n(p_i-p)^2. \qquad (5)$$

První člen na pravé straně předchozího výrazu představuje rozptyl veličiny $\sum_{i=1}^n Y_i$ za předpokladu, že všechny veličiny X_j^i jsou alternativní a nezávislé se stejným parametrem p. [4]

Míru homogenity dotazníkových položek pak můžeme určovat z podobného důvodu jako v prvním modelu i pomocí ukazatele, definovaného jako poměr $H_2/(r-1)$, kde

$$H_2 = (r-1)\,\frac{\sum_1^n(p_i-p)^2}{npq}, \quad r>1, \quad H_2\geq 0. \qquad (6)$$

Opět zřejmě je v případě homogenity, tj. za platnosti (2), $H_2=0$.

Odhadem ukazatele (6) je ukazatel $\widehat{H}_2$

$$\widehat{H}_2 = (r-1)\,\frac{\sum_1^n(\widehat{p}_i-\widehat{p})^2}{n\cdot\widehat{p}\,\widehat{q}}. \qquad (7)$$

Vzor záznamu experimentálních výsledků pro výpočet ukazatelů (4) a (7) je schematicky naznačen v Tab. 1.

V Tab. 1 označují x_j^i hodnoty 0 nebo 1 pro i-tou položku a j-tého jedince, součet $\sum_{i=1}^n x_j^i = v_j$ je výsledek dotazníku j-tého respondenta (tzv. hrubý skór), hodnoty y_i jsou počty správných odpovědí pro i-tou položku od r respondentů, $\widehat{p}$ se určí z podílu $\sum_{i=1}^n \widehat{p}_i/n$.

i,j	1	2	…	r	$y_i = \sum_{j=1}^{r} x_j^i$	$z_i = r - y_i$	$\widehat{p}_i = \frac{y_i}{r}$	$\widehat{q}_i = 1 - \widehat{p}_i$	$\widehat{p}_i\widehat{q}_i$
1	x_1^1	x_2^1	…	x_r^1	y_1	z_1	$\widehat{p}_1$	$\widehat{q}_1$	$\widehat{p}_1\widehat{q}_1$
2	x_1^2	x_2^2	…	x_r^2	y_2	z_2	$\widehat{p}_2$	$\widehat{q}_2$	$\widehat{p}_2\widehat{q}_2$
…	…	…	…	…	…	…			
…	…	…	…	…	…	…			
n	x_1^n	x_2^n	…	x_r^n	y_n	z_n	$\widehat{p}_n$	$\widehat{q}_n$	$\widehat{p}_n\widehat{q}_n$
$v_j = \sum_{i=1}^{n} x_j^i$	v_1	v_2	…	v_r	$\sum_1^n y_i$		$\sum_1^n \widehat{p}_i$	$\sum_1^n \widehat{q}_i$	$\sum_1^n \widehat{p}_i\widehat{q}_i$

Tabulka 1

i,j	1	2	3	4	5	y_i	$\widehat{p}_i$	$\widehat{q}_i$	$\widehat{p}_i \cdot \widehat{q}_i$
1	0	0	1	1	1	3	$0{,}6 = 3/5$	$0{,}4 = 1 - 0{,}6$	0,24
2	0	1	0	0	0	1	0,2	0,8	0,16
3	1	1	1	1	1	5	1,0	0,0	0,00
4	0	1	0	0	1	2	0,4	0,6	0,24
$v_j = \sum_{i=1}^{4} x_j^i$	1	3	2	2	3	$\sum_{i=1}^{4} y_i = 11$	$\sum_i \widehat{p}_i = 2{,}2$	$\sum_i \widehat{q}_i = 1{,}8$	$\sum_i \widehat{p}_i\widehat{q}_i = 0{,}64$

Tabulka 2

ROBUST 2026 – NEJEN ROBUSTNÍ METODY, ALE TAKÉ SNÍH, TĚŽIŠTĚ, TOULKY A ZPĚV

ROBUST 2026 – NOT JUST ROBUST METHODS, BUT ALSO SNOW, THE CENTER OF GRAVITY, WANDERING, AND SINGING

Ondřej Vencálek

E-mail: ondrej.vencalek@upol.cz

Tento krátký text jsem sepsal, abych doplnil „oficiální“ zprávu o Robustu 2026 o zcela subjektivní, daty málo podložené (naštěstí) postřehy z tradičního setkání českých a slovenských statistiků.

Zasněžená imprese

Zasněžená pláň se začínala třpytit v odlescích slunce, které se pozvolna šplhalo nad obzor. Nebe bylo žluté, bílé a světle modré. Stromy oděny ve slavnostní bílou, posvátné ticho, a mráz … oj, mráz, co štípe na konečcích prstů a špičce nosu. Odvrátil jsem pohled od vycházejícího slunce a nechal jej klouzat po horizontu směrem doleva po okolních kopcích vystupujících z ranního mlžného oparu, až spočinul u nezaměnitelné siluety hradu Kašperk. Ten pohled mi připomněl můj první Robust. Před 20 lety, v lednu 2006 (nechce se tomu věřit)! Tenkrát jsme spolu s Danem Hlubinkou a spolužáky z matfyzu Šárkou, Matúšem, Michalem, Honzou Kalužou a Andreou Kvitkovičovou došli do Kašperských Hor ze Lhoty nad Rohanovem, kde se tehdy Robust konal. Jo, na první Robust se nezapomíná. Jen nás předem nikdo nevaroval, že je to návykové – od té doby jsem žádný Robust nevynechal… malinko jsem se zasnil a slunce bylo mezitím už nad obzorem. Znovu jsem se rozhlédl a bylo mi nádherně.

Toto všechno se odehrálo jednoho lednového rána roku 2026, v poslední den letošního Robustu. Stál jsem na Klostermanově vyhlídce kousek od hotelu Vydra na okraji šumavského střediska Srní, kde se zimní škola JČMF konala. Bylo 8 hodin ráno, cvakl jsem pár fotek a vydal se zpět do hotelu na snídani. Nechtěl jsem přijít pozdě na dopolední blok přednášek.

Dva středobody a dvě těžiště Robustu

Přednášky jsou totiž těžištěm Robustu a místnost, kde se přednášky konají, se na pět dnů stává pro účastníky konference středobodem jejich pohybu. Je

Konference ROBUST mají dlouhou tradici účasti doktorandů a studentů magisterského studia v samostatné studentské sekci. Na konferenci ROBUST 2026 studenti představili 23 posterů. Ceny byly uděleny ve třech kategoriích: studenti bakalářského a magisterského studia, kteří dosud neukončili magisterské studium, dokončené diplomové práce a doktorandi. Jako obvykle ceny sponzorovaly Nadace RSJ a konference ROBUST. Konferenční poplatky mnoha účastníků, převážně studentů magisterského studia a doktorandů prvního ročníku, stejně jako na předchozích konferencích hradily RSJ a Česká statistická společnost.

Šéfredaktorka časopisu *Statistika: Statistics and Economy Journal* (`https://csu.gov.cz/2-statistika`) Stanislava Hronová laskavě vyzvala účastníky, aby do časopisu zasílali články na odpovídající témata.

4. Třetí model

K posouzení homogenity dotazníkových položek je možné také použít výběrovou charakteristiku χ^2

$$\chi^2 = \sum_1^n \frac{(n_i - n_i^t)^2}{n_i^t}, \tag{8}$$

kde n_i znamená pozorovanou četnost v i-té kategorii, n_i^t příslušnou teoretickou četnost, určenou na základě jisté hypotézy. Pro nás to bude hypotéza o homogenitě položek. Charakteristika χ^2 je ukazatelem rozdílu mezi skutečností a hypotézou. Ze vztahu (8) je vidět, že když pozorovaná četnost je v úplném souladu s hypotézou, je $\chi^2 = 0$. Čím větší je rozdíl pozorovaných četností od četností teoretických, tím větší je hodnota χ^2. Protože charakteristika χ^2 je určena na základě druhých mocnin rozdílů četností, nevyjadřuje směr odchylek. K charakteristice χ^2 přísluší určitý počet stupňů volnosti. Veličina (8) má $n-1$ stupňů volnosti pokud teoretické četnosti nejsou vázány dalšími podmínkami, v opačném případě se číslo $n-1$ musí snížit o počet vazebních podmínek. Jak to prakticky realizovat, ukážeme na řešení našeho problému.

Nechť je r respondentů podrobeno testu o n položkách. Výsledky pro test, kde každá položka je hodnocena pouze dichotomicky, je možné zaznamenat do tabulky typu Tab. 1. Náhodnou veličinou X_j^i opět označíme výsledek j-tého respondenta v i-té položce (může tedy jen být $X_j^i = 0$ nebo $X_j^i = 1$). Podobně použijeme odhady pravděpodobnosti správné, resp. nesprávné odpovědi na i-tou položku $\widehat{p}_i$, resp. $\widehat{q}_i$ a jejich průměrné hodnoty p, q. Ještě označíme $Z_i = r - Y_i$.

Podobně jako v předchozích modelech, zde máme $\widehat{p} = \sum_1^n Y_i/(rn)$, $\widehat{q} = 1 - \widehat{p}$.

Veličinu χ^2 pak můžeme napsat na základě hypotézy o homogenitě položek ve tvaru

$$\chi^2 = \sum_{i=1}^n \left(\frac{(Y_i - pr)^2}{pr} + \frac{(Z_i - qr)^2}{qr} \right)$$

Snadno vidíme, že $(Y_i - pr)^2 = (-Z_i + r - pr)^2 = (Z_i - qr)^2$ a tedy dostaneme, že

$$\chi^2 = \frac{1}{pqr} \cdot \sum_1^n (Y_i - pr)^2$$

nebo

$$\chi^2 = \frac{1}{pq} \cdot \left[\sum_1^n \frac{(Y_i)^2}{r} - p^2 rn \right] \tag{9}$$

Počet stupňů volnosti zde je $n-1$. Je-li hodnota χ^2, vypočtená podle (9) větší než v tabulkách nalezená hodnota kritická na 5% či 10% hladině významnosti, hypotézu o homogenitě dotazníkových položek zamítáme na příslušné hladině významnosti (jinak ji nezamítáme). Zde je třeba připomenout, že charakteristika χ^2 je charakteristikou pouze aproximativní a má pro malá n „slabou schopnost“ zamítnout nulovou hypotézu. A to bývá často problém, protože například v pedagogické praxi jde o hodnoty n většinou menší než 20. [3]

Ve vztahu (9) používáme již dříve zavedené odhady, takže dostaneme odhad pro χ^2 ve tvaru

$$\chi^2 \sim \frac{1}{\widehat{p}\,\widehat{q}} \cdot \left[\sum_1^n \frac{(y_i)^2}{r} - \widehat{p}^2 rn \right]. \tag{10}$$

5. Příklad

Mějme dotazník o $n = 4$ položkách. Reakce respondentů budou hodnoceny pouze dichotomicky (1 – uspěl, 0 – neuspěl). Dotazník byl aplikován u $r = 5$ náhodně vybraných respondentů, jejichž výsledky jsou uvedeny v tabulce Tab. 2.

$$\widehat{p} = \frac{\sum_i \widehat{p}_i}{4} = \frac{2{,}2}{4} = 0{,}55 = 1 - \widehat{q}, \quad \widehat{p} \cdot \widehat{q} = 0{,}2475, \quad \sum_i \widehat{p}_i \widehat{q}_i = 0{,}64,$$

$$\sum_i (\widehat{p}_i - \widehat{p})^2 = (0{,}6-0{,}55)^2 + (0{,}2-0{,}55)^2 + (1{,}0-0{,}55)^2 + (0{,}4-0{,}55)^2 = 0{,}35.$$

Variační koeficient je $\sqrt{\frac{0{,}35}{4}} \Big/ 0{,}55 \sim 0{,}54$, což je velké číslo, které svědčí o relativně velkých rozdílech mezi jednotlivými $\widehat{p}_i$. Podle (4a) dostaneme $\widehat{H}_1 = 5 \cdot \frac{0{,}35}{0{,}64} \sim 2{,}7$ a $\frac{\widehat{H}_1}{5} \sim 0{,}54$, resp. podle (4b) je $\widehat{H}_1 = 5 \cdot \frac{0{,}60}{0{,}64} \sim 4{,}7$ když $\widehat{p} = 0{,}8$ a $\frac{\widehat{H}_1}{5} \sim 0{,}94$; a $\widehat{H}_1 \sim 6{,}6$ když $\widehat{p} = 0{,}2$ a $\frac{\widehat{H}_1}{5} \sim 1{,}32$; kriteriální hodnota je pak $\min\{4{,}7; 6{,}6\}/5 = 0{,}94$; opět dostáváme dosti velké hodnoty.

ROBUST 2026
24. STATISTICKÁ KONFERENCE ROBUST

Jaromír Antoch

E-mail: antoch@karlin.mff.cuni.cz

Ve dnech 18. až 23. ledna 2026 se v Srní u Sušice konal 24. ročník tradiční statistické konference ROBUST 2026, pořádané jednou za dva roky. Společnými pořadateli byly Odborná skupina pro výpočetní statistiku České matematické společnosti (sekce Jednoty českých matematiků a fyziků), Katedra pravděpodobnosti a matematické statistiky Matematicko-fyzikální fakulty Univerzity Karlovy v Praze a Česká statistická společnost. Konference se uskutečnila v hotelu Vydra.

Celkem 93 účastníků ze tří zemí předneslo a diskutovalo příspěvky pokrývající široké spektrum témat od teoretické statistiky, pravděpodobnosti a stochastické analýzy přes strojové učení a informatiku až po aplikovanou statistiku v řadě oborů, včetně pojistné a finanční matematiky, veterinárního lékařství, medicíny, zdravotnictví a epidemiologie či metrologie. Zahraniční účastníci přijeli ze Slovenska (15) a Rakouska (1).

Základní myšlenkou konferencí ROBUST bylo vždy sdružovat statistiky všech generací a ze všech oborů, především z českých a slovenských institucí, ale také české a slovenské statistiky působící v zahraničí. Cílem je umožnit výměnu myšlenek a nabídnout účastníkům mezioborový pohled na současný statistický výzkum. Jednacími jazyky jsou tradičně čeština a slovenština, příspěvky v angličtině zaznívají jen příležitostně.

Zaznělo šest zvaných přednášek. M. Maciak z MFF UK v Praze a R. Matúš z MSD CR v Praze podrobně vysvětlili, proč a jak zveřejnit vlastní balíčky pro R v repozitáři CRAN. Na jejich přednášku navázal J. Lacko představením balíčku RCzechia pro R, které přirozeně doplnilo předchozí výklad. J. Baruník z FSV UK v Praze přednesl příspěvek „Přizpůsobení volby portfolia pomocí strategií cílených na kvantily“, zatímco K. Hron z UPOL v Olomouci se zaměřil na „Vícerozměrné Bayesovy prostory“. Přední odborník na výzkum mozku R. Rosipal z ÚM SAV v Bratislavě přednesl velmi zajímavou přednášku nazvanou „Od tenzorů k terapii: Statistické modely pro kolaborativní rehabilitaci s využitím rozhraní mozek–počítač a virtuální reality“. Konferenci uzavřela přednáška D. Hlubinky o testech využívajících charakteristické funkce.

blok nazvaný mnohoznačně Statistické metody. Dá se říci, že společným jmenovatelem těchto (i dalších příspěvků v závěrečném bloku nazvaném Aplikace) byla detekce nějaké odchylky od standardního průběhu sledovaného procesu či od standardní struktury dat. Gejza Dohnal předvedl metody detekce změny v sledovaném procesu prostřednictvím detekce změny struktury vhodného modelu. Blanka Šedivá se potom zabývala přímo metodami rozpoznání outlierů v mnohorozměrných datech. Musím předběhnout a zmínit už zde příspěvek Zuzany Hübnerové, která detekovala atypické průběhy funkcí v analýze stavu motorů sportovních letadel, s využitím konceptu hloubky v množině funkcí.

Následovala zasloužená (a dobrá) večeře v restauraci Fontána (u fontány v parku, kde se tradičně koná Flora Olomouc). Pak už jsme se přesunuli do příjemného prostředí penzionu Töde.

Dopolední blok druhého dne se věnoval výuce, a samozřejmě, nedaly se obejít výzvy související s rychlým rozvojem používání AI. Jiří Dvořák ukázal, jaký vliv může mít AI a používání internetu vůbec na skladbu matematické přednášky, které etapy je možné upozadit a nechat na studentech (předávání základních informací, výklad pojmů) a na které by naopak měl být zvýšen důraz (interpretace, použití, ale i samotná idea určitého tvrzení i důkazu, tj. spíš „proč“ místo „jak“). Podobnými problémy se zabýval i Ondřej Vozár, tentokrát v oblasti výuky metod výběrových šetření. Zdůraznil také, že inovace by se měla týkat i věcného obsahu, současných technik, včetně těch, které preferují budoucí zaměstnavatelé studentů. Příspěvek Zuzany Pacákové popisoval různé přístupy studentů k přípravě během kurzu: počáteční snaha, než dojdou síly, průběžné učení (když sil je dost) nebo pomalý start a silný finiš. Ukázalo se, že poslední strategie je optimální co do poměru výsledek/úsilí.

Na oběd jsme znovu zašli do oblíbené již restaurace Fontána, a pak následoval závěrečný blok příspěvků, nazvaný Aplikace. O přednášce Zuzany Hübnerové již byla řeč. Pak nás Tomáš Hlavsa seznámil s jedním dílčím problémem v stochastické hraniční analýze (SFA). Tím je výběr jednostranného rozdělení popisujícího tzv. technickou neefektivnost, neboť volba této distribuce podstatně ovlivňuje predikované hraniční hodnoty. Závěr patřil Marii a Ladislavu Budíkovým. Představili jedno z vhodných a flexibilních rozdělení k popisu pravděpodobnosti překročení. Ač inspirace přišla z oblasti hydrologie (sledování průtoků), podobný problém je řešen i v analýze životnosti, nemluvě o analýze finančních dat a modelování extrémních hodnot. I tady se ukázalo, jak problémy z různých oblastí souvisejí. A setkání jako Statistické dny snad vedou i k tomu, abychom si tyto souvislosti uvědomovali, vzájemně si rozuměli a obohacovali se. Takže ještě jednou velký dík organizátorům i účastníkům a někdy brzo opět nashledanou.

Podle (7) máme $\widehat{H}_2/(r-1) = \frac{0{,}35}{4 \cdot 0{,}55 \cdot 0{,}45} \sim 0{,}56$. Tato hodnota nesvědčí o velké nehomogenitě.

Interpretace hodnot těchto ukazatelů je vázána na zkušenost. Ta v podobných případech, když jejich hodnoty jsou větší než 1, předpoklad o homogennosti položek zamítá.

Podle vzorce (10) v našem případě dostáváme

$$\chi^2 = \frac{1}{0{,}55 \cdot 0{,}45} \cdot \left[\frac{1}{5} \cdot (9 + 1 + 25 + 4) - 0{,}55^2 \cdot 20 \right] \sim 7{,}1.$$

Kritická hodnota χ^2 při 5% hladině významnosti a $(n-1)(2-1) = 3$ stupních volnosti je rovna 7,81, tedy námi vypočtená hodnota χ^2 tuto hodnotu nepřevyšuje. Na základě našeho experimentu (a volené hladině významnosti) nemůžeme hypotézu o homogenitě položek na 5% hladině významnosti zamítnout. Všimněme si ještě, že na hladině významnosti 10% je kritická hodnota rovna 6,25 a při této volbě hladiny významnosti hypotézu zamítáme. Náš případ je pro statistické rozhodování pomocí limitního kritéria χ^2 hraniční. To není nic mimořádného, χ^2 má, jak jsme se již zmínili, „slabou schopnost“ hypotézu při malém počtu stupňů volnosti zamítnout.

V případě, že můžeme předpokládat homogenitu dotazníkových položek, je výsledný skór j-tého jedince roven

$$v_j = \sum_1^4 x_j^i, \quad v_j \in \{0, 1, 2, \ldots, 4\}.$$

U nehomogenního dotazníku musíme v zájmu objektivity výsledný skór určovat vzorcem, který více hodnotí například vyřešení položek obtížných

$$v_j = \frac{1}{\sum_1^n \widehat{q}_i} \sum_{i=1}^4 x_j^i \cdot \widehat{q}_i, \quad v_j \in \langle 0; 1 \rangle.$$

Pak bude $v_2 = \frac{1}{1{,}8} \cdot (0{,}8 + 0{,}0 + 0{,}6) \sim 0{,}78$ a $v_5 = \frac{1}{1{,}8} \cdot (0{,}4 + 0{,}0 + 0{,}6) \sim 0{,}56$, přitom v případě homogenity mají oba jedinci skóry stejné. Jak se oba výsledky od sebe liší, poznáme, když je budeme transformovat do rozsahu původní škály: $0{,}78 \cdot 4 = 3{,}12$ a $0{,}56 \cdot 4 = 2{,}24$. Použijeme-li při určování vah místo obtížnosti $\widehat{q}_i$, hodnoty $\widehat{p}_i$, preferujeme položky snadné. Krom toho vážený skór můžeme určit na jiném základě, např. v závislosti na důležitosti položky. Podle toho se může měnit význam výsledného váženého dotazníkového skóru.

6. Čtvrtý model

Mějme k dispozici reprezentativní výběr r respondentů a uvažujme náš předchozí příklad (zde je $r = 5$). Za předpokladu, že dotazník je homogenní a pravděpodobnost správné odpovědi na každou položku je $p = 0{,}5$ a odpovědi na položky jsou statisticky nezávislé, mají skóry v_j jednotlivých respondentů $v_j = \sum_{i=1}^{4} x_j^i$, $j = 1, 2, \ldots, r$, binomická rozdělení se stejnými parametry $n = 4$, $p = 0{,}5$. Mezi odpověďmi jednotlivých respondentů můžeme předpokládat také statistickou nezávislost. Pak i součet $\sum_1^r v_j = V$ má binomické rozdělení s parametry $rn = 5 \cdot 4 = 20$, $p = 0{,}5$. Jsou-li splněny předchozí předpoklady, pak veličina V by měla s pravděpodobností $1 - \alpha$ nabývat hodnoty mezi V_α a $V_{1-\alpha}$ za platnosti nerovností $\sum_{k=0}^{V_\alpha} \binom{rn}{k} \cdot 0{,}5^{rn} \leq \alpha/2$; $\sum_{V_{1-\alpha}}^{20} \binom{rn}{k} \cdot 0{,}5^{rn} \leq \alpha/2$.

Pro náš případ, když je $\frac{\alpha}{2} \sim 0{,}02$, je $V_\alpha = 5$, $V_{1-\alpha} = 15$. Podobně dostaneme pro $\frac{\alpha}{2} \sim 0{,}06$ hodnoty $V_\alpha = 6$, $V_{1-\alpha} = 14$.

Protože $\sum_1^r v_j = V = \sum_{i=1}^4 y_i = 11$, nemůžeme na základě předchozí úvahy zamítnout předpoklad $p = 0{,}5$ jak na hladině významnosti $\alpha = 0{,}04$, tak i $\alpha = 0{,}12$.

Poznámka 2: Podobně můžeme postupovat při ověřování homogenity dotazníku, když bychom volili hodnotu pravděpodobnosti p jinou (odlišně zdůvodněnou, například když bychom chtěli ověřit, zda položky nejsou extrémně snadné, resp. obtížné), například můžeme jako nulovou hypotézu volit $p = 0{,}25$ nebo $p = 0{,}75$. □

Poznámka 3: Důležitým parametrem dotazníku, který ovlivňuje jeho obtížnost, je povolený čas pro odpovědi. Cítíme-li, že ho potřebujeme v daném případu uvažovat, můžeme se rozhodnout pro určitý model stochastického procesu, kterým je pak proces konstrukce odpovědí modelován. Dovolujeme si v této souvislosti upozornit na podrobnější vysvětlení v [7] nebo náš příspěvek k této problematice v [6]. □

7. Pátý model

V předchozích úvahách o podmínkách homogenity dotazníku jsme uvažovali pouze globální kritéria (dotazník jsme chápali jako určitý celek, protože položky by měly být spolu významově vázány a uplatňuje se jen celkový hrubý skór). Přesto si zde ukážeme jednoduché hodnocení variability odpovědí mezi respondenty u každé položky zvlášť. Vycházíme při tom z jednoduché úvahy, že položka s vysokou variabilitou odpovědí nemůže být snadná ani obtížná. Požaduje se, aby ukazatel variability i-té položky δ_i byl pro vhodnou položku

příspěvek Ondry Vencálka, který nabídl exploratorní analýzu časového vývoje úmrtnosti v souvislosti s proočkovaností populace. Kromě této analýzy měl tento příspěvek i nezanedbatelný filozofický přesah: Co vlastně vidíme v „zrcadle“ našich analýz? Asi ne to, co si přejeme (to umí jen zrcadlo z Erisedu), ale i zrcadla ukazující nekompromisně pravdu se vyskytují spíš v pohádkách (viz riskantní dotaz „Kdo je na světě nejkrásnější?“ ve Sněhurce). Tu se nabízí až otázka, zda jediná pravda existuje (idealisté asi řeknou, že ano, ale nikdo ji neviděl, stejně jako nikdo neviděl Boha a možná bychom byli rádi, kdyby existoval). Námi pozorované zrcadlo spíš připomíná vodní hladinu zčeřenou větrem, kde z reality vidíme jen spoustu víceméně deformovaných fragmentů. Jako statistici o tom něco víme. Na toto téma navázal příspěvek Tomáše Fürsta, dotýkající se občasné samoúčelnosti používání populárních statistických metod bez ohledu na cíl zamýšlené analýzy. A vlastně implicitně varoval i před tím, aby se pozorovaný fragment považoval za celou realitu.

I další sekce byla věnována biostatistice, z čehož se trochu vymykala přednáška Tomáše Chupáně o efektivní práci s velkými datovými soubory a o možnostech, které nabízí R-ko. Pak následoval mezinárodní blok: V příspěvku Dominiky Mikšové jsme se dozvěděli, jak pomocí sledování rakouských (i jiných) zdravotnických registrů je možné monitorovat a analyzovat kvalitu zdravotnické péče a její trend. Další příspěvek dorazil z Bratislavy. Přednáška Ivety Waczulíkové se týkala modelování změn průtokových vlastností krve, podstatou sdělení byl ale obecný problém identifikace rolí různých veličin, které je třeba do modelu zahrnout (konfoundery, mediátory, supresory a jiné modifikátory výsledného efektu). Po kávě a občerstvení jsme šli na

ZPRÁVA O STATISTICKÝCH DNECH KONANÝCH 10. – 11. ZÁŘÍ V OLOMOUCI

REPORT ON THE STATISTICAL DAYS HELD ON 10–11 SEPTEMBER IN OLOMOUC

Petr Volf

E-mail: volf@utia.cas.cz

Česká statistická společnost vyslyšela přání členů i příznivců a připravila ve dnech 10. a 11. září 2026 Statistické dny v Olomouci. Tedy, sama „Společnost“ by je těžko připravila, nebýt organizátorů v čele s Ondrou Vencálkem a jeho olomouckým týmem. Podařilo se jim pro setkání zajistit pavilon uprostřed zeleně v Botanické zahradě PřF UP Olomouc a v něm krásný sál Josefa Otruby. Takže unavený účastník, kterému už bylo málo i podávané občerstvení, se mohl ve venkovní expozici potěšit pohledem na mnoho druhů rostlin z celého světa (žel, některé vypadaly, že jsou poznamenány letošním suchým počasím).

Setkání bylo tentokrát dvoudenní, s noclehem v penzionu Töde v historickém centru města. Pečlivě vybraném nejen pro nocleh, ale i pro večerní neformální povídání na témata pracovní i jiná na zahradě penzionu (počasí nám první den přálo). Povídání se protáhlo, ale druhý den jsme se disciplinovaně sešli a pokračovali v rokování. Navíc začlo pršet, což také možná přispělo k zaplnění přednáškového sálu. Dík tomu, že se setkání konalo přímo ve městě, kde sídlí spolupořádající pracoviště (PřF Univerzity Palackého), programu se mohli zúčastnit i studenti a další místní příznivci statistiky. Takže do prezenční listiny se nakonec zapsalo 50 účastníků, což je rekordní počet na Statistických dnech.

A nyní již k programu Statistických dnů: Program, abstrakty i další informace jsou dostupné na stránce https://www.statspol.cz/STATDNY2026/. Nejdříve nás přivítala a rokování slavnostně zahájila naše milá předsedkyně Martina Litschmannová (viz obrázek). Příspěvky byly rozděleny do několika tematických okruhů. Tak hned první, Biostatistika I, měl společného jmenovatele v návratu k časům Covidu se snahou stále se dobrat nějakých nesporných závěrů statistických analýz.

Jaroslav Janošek nás na začátek seznámil se soutěží v analýze dat DATACON, která se v první etapě zaměřila na souvislost mezi spotřebou kortikoidů (používaných na potlačení autoimunitních onemocnění) a vakcinací proti Covidu-19, která také mohla způsobovat autoimunitní problémy. Následoval

co nejblíže své maximální hodnotě $\delta_i^{\max}$. Podle [7], str. 142, se uvažovaná variabilita odpovědí dá charakterizovat při dichotomickém hodnocení odpovědí pro i-tou položku ukazatelem

$$\delta_i = \binom{r}{2}^{-1} \cdot n_0 \cdot n_1, \quad \delta_i \geq 0, \quad i = 1, 2, \ldots, n, \tag{11}$$

kde n_0, resp. n_1, je počet hodnocení 0, resp. 1, mezi r respondenty, proto $n_0 + n_1 = r$.

Maximální hodnota je (v případě dichotomického skórování)

$$\delta_i^{\max} = \frac{1}{2} \cdot \frac{r}{r-1}. \tag{12}$$

V našem příkladu je $r = 5$, takže $\delta_i^{\max} = \frac{1}{2} \cdot \frac{5}{4} \sim 0{,}625$.

Podíváme-li se na řádky Tab. 2, které odpovídají odpovědím na určitou položku, dostaneme výsledky, uvedené v Tab. 3 pro $i = 1, 2, \ldots, 4$.

i	δ_i	Hodnocení δ_i
1	$10^{-1} \cdot 2 \cdot 3 = 0{,}6$	C
2	0,4	B
3	0,0	A
4	0,6	C

Tabulka 3

Exaktní testování hodnot veličiny δ_i je obtížně dostupné, ale pro běžnou praxi nám stačí „rozumné“ slovní hodnocení „blízkosti“ hodnoty δ_i hodnotě maximální $\delta_i^{\max} = 0{,}625$.

Jako ukázku možného slovního hodnocení jsme interpretovali hodnoty δ_i tak, jak je uvedeno v Tab. 4. Jak s danými informacemi naloží sestavovatel a hodnotitel určitého dotazníku, záleží ještě na dalších (nestatistických) podmínkách konstrukce a jejich souvislostech.

Interval hodnot δ_i	Interpretace	Označení
$\langle 0; 0{,}30)$	Položka je nevhodná	A
$\langle 0{,}30; 0{,}50)$	O vhodnosti položky lze uvažovat	B
$\langle 0{,}50; 0{,}625\rangle$	Položka je vhodná	C

Tabulka 4

Poznámka 4: V celém příspěvku jsme předpokládali, že pravděpodobnost výskytu úrovně 1 je specifická pro každou položku a nezávisí na tom, o kterého respondenta se jedná. Je-li $n(r-1) > 30$, můžeme tento předpoklad otestovat Cochranovým Q-testem, který má kritérium Q dané vztahem

$$Q = \frac{(r-1)\left[r \cdot \sum_1^r v_j^2 - (\sum_1^r v_j)^2\right]}{r\sum_1^n y_i - \sum_1^n y_i^2}$$

s přibližně χ^2_{r-1} rozdělením. Velké hodnoty χ^2_{r-1} svědčí o neplatnosti předpokladu. V našem případě je $Q = 3{,}5$, ale $n\,(r-1) = 16 < 30$. Proto tuto hodnotu nemůžeme použít pro srovnání s rozdělením χ^2, viz např. [7]. □

8. Závěr

Některé důležité dotazníky, zkoumající výkon (např. tzv. vědomostní testy), z důvodu utajení obsahu svých položek (i obsah má být náhodně vybraný) postrádají apriorní exaktní odhad pravděpodobnosti správné reakce „průměrného respondenta“ na položky, a proto se všechny položky hodnotí stejně (0 nebo 1). Nerespektování možné rozdílné obtížnosti položek je hrubou chybou pro interpretaci skórování. Z hlediska správné interpretace nejčastěji užívané automatizované metody skórování by se měla krom jiného zajistit (při 0 – 1 skórování) stejná obtížnost všech položek, dotazník by měl být homogenní.

Uvažujme homogenní výkonový dotazník (CR test) s položkami buď velmi snadnými, resp. obtížnými (p velké, resp. p malé). Takový dotazník velmi špatně odlišuje jedince navzájem podle hrubého skóru (protože rozptyl npq je malý). Je proto výhodnější, když homogenní dotazník má položky s obtížnostmi $p \sim 0{,}5$, pak je rozptyl npq maximální.

U nehomogenních dotazníků předpokládáme předběžný odhad obtížností položek a pro zkoumanou populaci konstrukci vážených skórů. [1, 5]

Ukázali jsme použití statistiky i tam, kde se to nepovažuje až za tak důležité. Přesto statistika nabízí vnímavému hodnotiteli řadu postupů, jak korigovat úvahy, které se někdy o statistiku neopírají (při metodě skórování, ale i při volbě témat jednotlivých položek), ač se velmi často používají. Ukazuje se, že použití statistiky i zde vyžaduje navíc úvahy zdravého rozumu.

Poděkování

Autor děkuje Pavlu Střížovi za technické dopracování příspěvku.

Literatura

[1] Půlpán, Z.: Dávají nám dotazníky dostatek informací?, *Obzory matematiky, fyziky a informatiky* 2/2025 (54), s. 1–15. *cit. 4, 14*

[2] Lindquist, E. F.: *Statistická analýza v pedagogickém výzkumu*, SPN, Praha 1967, 264 s. *cit. 4*

[3] Hátle, J., Likeš, J.: *Základy počtu pravděpodobnosti a matematické statistiky*, SNTL, ALFA, Praha 1972, 463 s. *cit. 10*

[4] Janko, J.: *Jak vytváří statistika obrazy světa a života*, 1. díl, Cesta k vědění, JČMF 1942. *cit. 5, 7*

[5] Půlpán, Z.: *Odhad informace z dat vágní povahy*, Academia, Praha 2012, 199 s. *cit. 14*

[6] Půlpán, Z.: *K problematice hledání podstatného v humanitních vědách*, Studie 1, Academia, Praha 2001, 135 s. *cit. 12*

[7] Komenda, S., Klementa, J.: *Analýza náhodného v pedagogickém experimentu a praxi*, SPN, Praha 1981, 316 s. *cit. 12, 13 a 14*