



Manuálny Výpočet Deskriptívnej Štatistiky, Intervalov Spôľahlivosti, Korelačnej Analýzy, T-Testov, F-Testu, Multivariačnej Analýzy a Analýzy prežívania V Ére Softvéru Typu "Black Box" Pre Hodnotenie Zmeny Zdravotného stavu Pacienta

¹Daniel FURKA, ¹Saul DUEÑAS CASAS, ¹Andrej TÓTH, ¹Jerguš VENGRÍNI, ²Ondrej LAURINEC, ³Samuel HORVÁTH, ⁴Anton GÚTH, ⁵Karol HORNÁČEK, ^{1,6}Ivan FURKA, ^{1,7*}Samuel FURKA

¹ Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Katedra fyzikálnej a teoretickej chémie, Mlynská dolina, Ilkovičova 6, 841 04 Bratislava, Slovensko.

² Klinika fyziatrie, balneológie a liečebnej rehabilitácie, Univerzitná nemocnica Bratislava, Nemocnica akad. L. Déreza, Limbová 5, 833 05 Bratislava, Slovensko.

³ Landesklinikum Mödling, Landesgesundheitsagentur Niederösterreich, Restituta-Gasse 12, 2340 Mödling, Rakúsko.

⁴ Rehabilitačný ústav Rehabilitácia - ambulancia profesora Gútha, Na Barine 16, 841 03 Bratislava, Slovensko.

⁵ Slovenská zdravotnícka univerzita v Bratislave, Lekárska fakulta, Klinika fyziatrie, balneológie a liečebnej rehabilitácie, Bratislava, Slovensko.

⁶ BEKO Engineering & Informatik GmbH, Stubenbastei 2, 1010 Viedeň, Rakúsko.

⁷ Oddelenie molekulovej a biochemickej genetiky, Ústav lekárskej biológie, genetiky a klinickej genetiky, Lekárska fakulta Univerzity Komenského v Bratislave, Univerzitná nemocnica, Mickiewiczova 13, 813 69 Bratislava, Slovensko.

*Korešpondujúci autor: samuel.furka@uniba.sk

SÚHRN/ABSTRAKT

Úvod: Moderné štatistické modely sú nevyhnutné na metodologicky správne hodnotenie zmien zdravotného stavu pacientov a presnú prognostiku. V období rokov 2003 až 2025 bol vo vedeckej komunite zaznamenaný nárast nesprávnej aplikácie štatistických metód. Spoliehanie sa na softvér typu „Black Box“, umelú inteligenciu alebo automatizáciu bez dôkladného pochopenia kľúčových parametrov často vedie k nepresnostiam. Pokročilý štatistický softvér zefektívňuje výpočty, nenahrádza však porozumenie modelu.

Súbor: Hodnotili sme zmeny zdravotného stavu pred a po podstúpení liečby a následnej rehabilitácie v dvoch odlišných populáciách, u pacientov so zlomeninou zápästia a u pacientov s intramedulárnym osteosarkómom. Svalová obnova po zlomenine si vyžaduje presnú diagnostiku, najmä štandardizovaný Jandov test na spoľahlivé monitorovanie priebehu. Naopak, pri rehabilitácii pacientov s osteosarkómom má prioritu vertikalizačný tréning zameraný na optimalizáciu kontroly ťažiska a aktiváciu posturálnych svalov. Tento prístup zachováva maximálnu heterogenitu vzorky a umožňuje sledovať zmeny pri živote neohrozujúcich aj život ohrozujúcich ochoreniach.

Metódy: Navrhujeme metodologický prístup zahŕňajúci hodnotovú maticu založenú na fuzzy logike pri rôznych hladinách významnosti na hodnotenie subjektívnych, kvázi subjektívnych a objektívnych zdravotných parametrov. Následne uvádzame podrobné manuálne výpočty deskriptívnej štatistiky, intervalov spoľahlivosti, korelačnej analýzy, t-testov, F-testu, viacrozmernej analýzy, Kaplanových-Meierových kriviek a log-rank testu s cieľom zabezpečiť metodologickú zrozumiteľnosť.

Výsledky: V našej štúdii sme zistili, že 65 zo 600 vedeckých prác použilo nepresné alebo neoptimalizované štatistické modely.

Záver: Cieľom tejto práce je poskytnúť metodologický návod na správnu aplikáciu pokročilých štatistických modelov vo vedeckom výskume, najmä pre lekárov na začiatku odbornej kariéry.

KEYWORDS

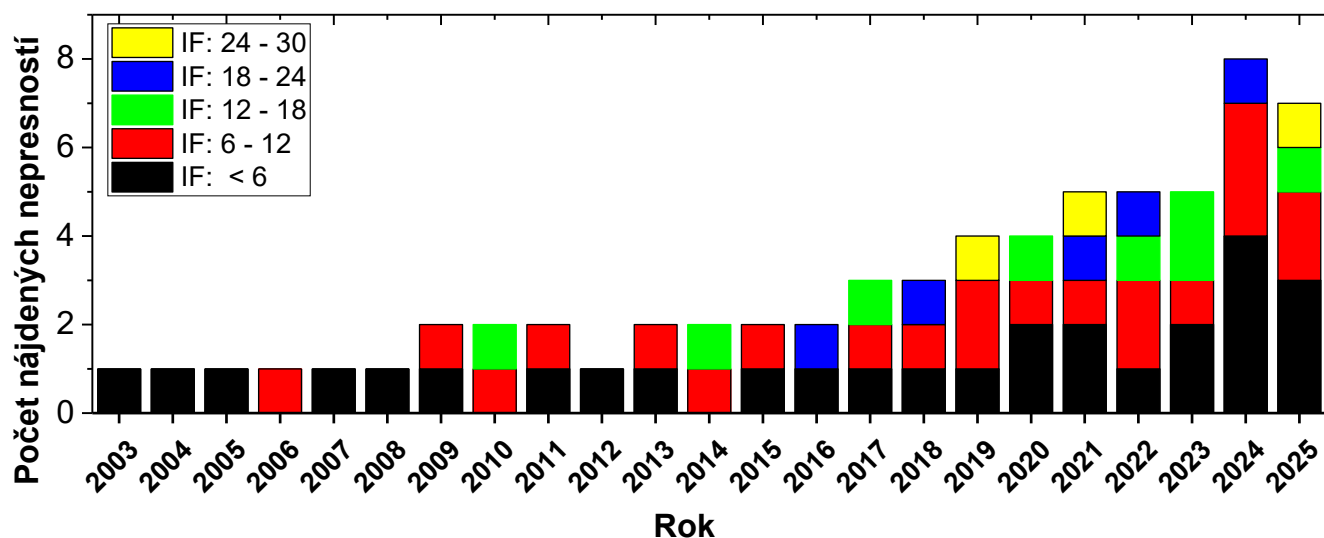
Nádorové ochorenia, bolesť, muskuloskeletálne poruchy, štatistika a výskumné metódy, rehabilitačná medicína.

EN-version DOI: [10.61983/lcrh.v63i2.167](https://doi.org/10.61983/lcrh.v63i2.167)

Zároveň je jej cieľom upozorniť skúsených odborníkov na možné nejednoznačnosti a procesné kroky pri používaní softvéru typu „čierna skrinka“.

1 ÚVOD

Široké spektrum parametrov opisujúcich zmeny zdravotného stavu pacienta si v rôznych situáciách vyžaduje použitie odlišných štatistických modelov. Správna kategorizácia parametrov a ich hodnotenie zároveň predstavujú základ vedeckého overovania hypotéz. Keďže aj matematika má svoje zradné zákutia, je nevyhnutné porozumieť štatistickým modelom, a tým predchádzať nesprávnej alebo nepresnej interpretácii ťažko získaných výsledkov. Na overenie matematickej správnosti vypočítaných štatistík možno použiť softvérové nástroje, ako je Statcheck.io¹. Závažným problémom pri interpretácii štatistiky je matematicky správny výpočet vykonaný na nevhodnom súbore alebo pomocou nevhodného štatistického modelu. Z vedeckého hľadiska to môže viesť k nesprávnemu prijatiu alebo zamietnutiu skúmanej hypotézy a následne k chybným záverom. Z pohľadu lekára je súčasná prax založená na dôkazoch a medicínskej vede, čo si vyžaduje aktívne alebo pasívne získavanie vedeckých informácií a overených výsledkov výskumu. Preto je porozumenie štatistickým metódam ako základnej súčasti medicínskeho výskumu nevyhnutné nielen na hodnotenie kvality výsledných prác, ale aj na správne pochopenie dôsledkov výskumných výsledkov. Žiaľ, v posledných rokoch sa v chemických aj medicínskych odboroch stretávame s veľmi častým nepresným používaním štatistických modelov. Niekedy ide o menšie chyby, ktoré nemenia celkové tendencie a trendy, inokedy však ide o kritické chyby. V súlade so všeobecnými usmerneniami sme preskúmali 600 vedeckých prác publikovaných v medicínskych časopisoch v období rokov 2003 až 2025. Články boli náhodne vybrané a kategorizované podľa impakt faktora (IF) do piatich skupín: a) IF: < 6, b) IF: 6-12, c) IF: 12-18, d) IF: 18-24 a e) IF: 24-30. Z každej skupiny pre každý rok v sledovanom intervale bolo náhodne vybraných 5 prác z troch rôznych časopisov. Obrázok znázorňuje frekvenciu identifikovaných štatistických problémov (obr. 1).



Obrázok 1 – Kumulatívne znázornenie nárastu výskytu nevhodne použitých štatistických modelov vo vedeckých článkoch súvisiacich s medicínou pre rôzne kategórie impakt faktora (IF) v období rokov 2003 až 2025.

Zistilo sa, že 65 zo 600 prác (10,8 %) obsahovalo nesprávne aplikované alebo neoptimálne zvolené štatistické modely, čo nás podnietilo k napísaniu štúdie zameranej na analýzu zdravotného stavu s metodologickým usmernením a podrobným riešením dôležitých štatistických modelov. Zaujímavé je, že v žiadnom prípade sa nezistili numerické chyby, ale iba nevhodné alebo dokonca nesprávne matematické modely. V prípadoch, v ktorých boli problémy identifikované v medicínskych časopisoch s IF < 18, by optimálny model viedol k čiastočnej zmene interpretácie. Pri identifikovaných problémoch v časopisoch s IF > 18 by optimalizácia viedla k zmenám prezentovaných odchýlok alebo kritických hodnôt.

Základ štatistického monitorovania zmien zdravotného stavu pacienta spočíva vo výbere relevantných ukazovateľov, spôsobe získavania informácií a voľbe vhodného matematického modelu². Tieto tri kroky nás vedú cez neprehľadné a zdanlivo nesúvisiace údaje k zamietnutiu alebo potvrdeniu našej hypotézy. V praxi

sa výber relevantných ukazovateľov často zjednodušuje pomocou softvéru³, publikovaných štúdií alebo skórovacích tabuliek⁴. Skóre alebo hodnota ukazovateľa zaraďuje pacienta do podskupiny ochorenia, podľa čoho sa následne riadi liečba alebo rehabilitácia v súlade s protokolom. Obnovenie fyziologickej kĺbovej mobility podľa Lewita v oblasti karpálnych kostí predstavuje kľúčový prínos pre zabezpečenie plného rozsahu pohybu a plynulej obnovy funkcie ruky. Skórovacie tabuľky a dotazníky sú založené na validovaných expertných skúsenostiach. Ak však pre sledovaný problém neexistuje relevantná skórovacia tabuľka, musíme relevantné parametre identifikovať sami, pričom sa zameriavame na tie, ktoré najvýraznejšie závisia od progresie alebo regresie ochorenia⁵.

Na štatistické hodnotenie môžeme použiť základnú deskriptívnu štatistiku vrátane priemeru, modusu, mediánu, variability a štandardnej odchýlky⁶, alebo môžeme aplikovať pokročilejšie modely, ako sú korelačná analýza, kovariančná analýza, testy rozptylu a rôzne T-testy⁷⁻⁹. T-testy porovnávajú priemery dvoch skupín, zatiaľ čo testy rozptylu hodnotia variabilitu v rámci skupín a medzi skupinami. Zložitejšie testy, ako sú ANOVA a MANOVA, umožňujú porovnanie viacerých skupín. ANOVA (analýza rozptylu) porovnáva priemery troch alebo viacerých skupín, zatiaľ čo MANOVA (viacrozmerná analýza rozptylu) rozširuje ANOVA na porovnanie viacerých závislých premenných medzi skupinami¹⁰. Moderné štatistické metódy nám zároveň umožňujú sledovať javy závislé od času, ako je prežívanie pacientov alebo životnosť implantátov, pomocou nástrojov, ako sú Kaplanova-Meierova krivka alebo log-rank test. Použitie vhodných štatistických modelov pri dobre kategorizovaných pacientoch poskytuje jasný pohľad na to, ako liečba ovplyvňuje zdravotné výsledky¹⁰. Pochopenie účinkov liečby zlepšuje našu schopnosť zdokonaľovať stratégie a predpovedať výsledky. Účinné liečebné stratégie pre konkrétne kategórie pacientov môžu znižovať náklady, skracovať trvanie liečby, predchádzať nesprávnej liečbe a zlepšovať prevenciu. Podobný problém vzniká aj pri diagnostických prístupoch. Potrebujeme informácie o diagnostickej hodnote metódy vrátane citlivosti, špecificity, prediktívnych hodnôt a ďalších parametrov. Správna diagnostická voľba ovplyvňuje nielen presnosť diagnózy, ale aj náklady, riziká a celkový prínos¹¹. Dobrým príkladom je skriningový proces, ktorý si vyžaduje metódu citlivú, široko dostupnú a lacnú. Účinný skrining by mal odhaliť ochorenie dostatočne skoro na to, aby významne ovplyvnil prežívanie pacienta, pričom by mal minimalizovať falošne negatívne aj falošne pozitívne výsledky¹². Komplexnosť intenzívnych a časovo náročných terapií si vyžaduje dlhodobé monitorovanie rôznych parametrov, čo predstavuje výzvu pri hodnotení skutočného zlepšenia alebo zhoršenia zdravotného stavu pacienta. Preto demonštrujeme možnosti, výhody a nevýhody použitia jednotlivých štatistických modelov na hodnotenie zmien zdravotného stavu s využitím reálnych údajov z dvoch výrazne odlišných skupín pacientov, pacientov so život neohrožujúcimi stavmi (zlomeniny zápästia) a pacientov so život ohrožujúcimi stavmi (intramedulárny osteosarkóm).

2 MATERIÁLY A METÓDY

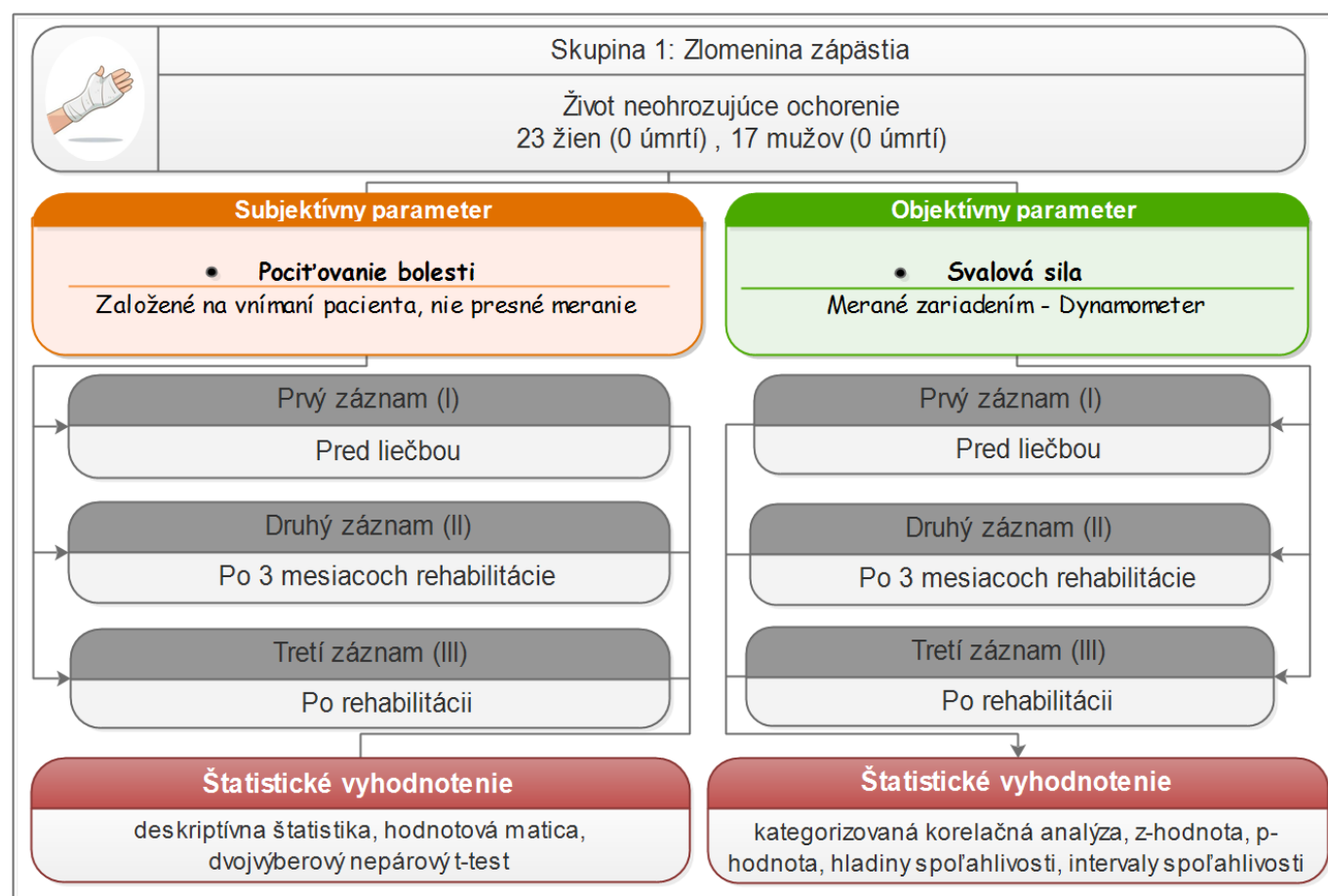
Cieľom tejto štúdie je poskytnúť praktický prehľad jednoduchých a pokročilých štatistických modelov používaných v medicínskom výskume, ktorý sa čoraz viac spolieha na softvér typu „Black Box“. Porozumenie manuálnym výpočtom modelov poskytne lekárovi a výskumníkovi v období rozvíjajúcej sa umelej inteligencie dôležitú kontrolu nad vlastnou istotou pri výbere sledovaných parametrov, ako aj pri následnom výbere parametrov, kategorizácii údajov alebo počas procesu štatistického hodnotenia platnosti hypotéz.

Ako objektívne parametre definujeme parametre získané priamo meracím zariadením alebo z laboratórnych výsledkov. Tieto parametre sú úplne nezávislé od pacienta a lekára. Údaje manuálne zaznamenané osobou, napríklad lekárom, sú definované ako kvázi objektívne parametre. Tieto parametre môžu byť ovplyvnené chybou alebo subjektívnou skúsenosťou lekára, nie však subjektívnym pohľadom pacienta. Naopak, údaje získané z výpovedí pacienta sú definované ako subjektívne parametre. Zozbierané údaje sa hodnotia pomocou jednoduchej deskriptívnej štatistiky, korelačnej analýzy, testov rozptylu, T-testov, Kaplanových-Meierových kriviek a log-rank testov.

Pri štatistickom skúmaní je dôležité rozlišovať, či je sledovaný parameter numerický alebo nenumeričný⁵. Pri numerických parametroch výpočet zahŕňa jednoduché dosadenie do matematického modelu. Pri nenumeričných parametroch, ako sú slovné odpovede pacientov, je vhodné použiť fuzzy logiku. Na tento účel navrhujeme metódu „hodnotovej matice“, ktorá jednotlivým slovným odpoveďami v rámci otázky

priraduje diskkrétne celočíselné hodnoty. Diskkrétne hodnoty možno priradiť podľa teoretickej intenzity odpovedí, napríklad nie, neviem, áno, alebo odpovediam bez závislosti od intenzity, napríklad červená, zelená, modrá. Použitie hodnotovej matice vyžaduje zaokrúhlenie priemernej odpovede na celé číslo pri odpovediach nezávislých od intenzity a na dve desatinné miesta pri odpovediach závislých od intenzity. Táto metóda nevytvára skórovaciu tabuľku, namiesto toho poskytuje reálnu deskriptívnu štatistiku z jednej odpovede pacienta. Hodnotové matice sú univerzálne a možno ich použiť kedykoľvek, je však potrebné dôsledne rozlišovať, či sú slovné odpovede kvalitatívne alebo kvantitatívne. V tejto práci sme skúmali tri otázky: (1) Pociťujete bolesť v končatine pri pohybe, (2) Pociťujete bolesť v končatine v pokoji a (3) Pociťujete pulzovanie v končatine. Odpovediam boli podľa fuzzy logiky priradené celočíselné hodnoty: áno = 1, skôr áno = 2, neviem = 3, skôr nie = 4 a nie = 5. Transformované numerické údaje boli následne použité na jednoduchú deskriptívnu štatistiku.

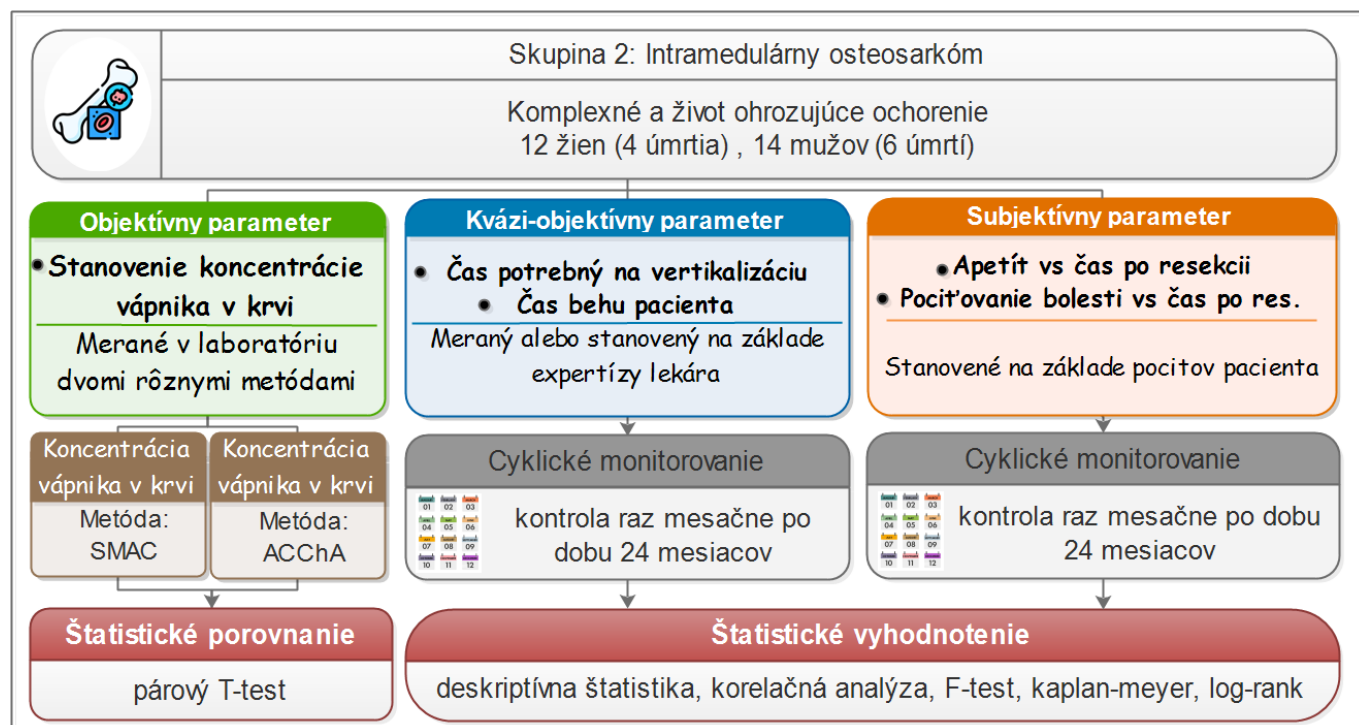
Prvú kohortu pacientov tvorilo 40 pacientov (23 žien, 17 mužov), ktorí utrpeli zlomeninu zápästia. Táto kohorta reprezentovala pacientov so život neohrozujúcim ochorením. Všetky parametre boli sledované: (I) pred liečbou, (II) po 3 mesiacoch od začiatku liečby počas rehabilitácie a (III) po ukončení liečby. Schematické znázornenie štruktúrovanej mapy je uvedené v schéme 1 (S. 1). Sledované parametre zahŕňali intenzitu bolesti ako subjektívny ukazovateľ a svalovú silu stisku hodnotenú pomocou dynamometra ako objektívny ukazovateľ. Včasná a cieleňá kinezioterapia zápästia po odstránení imobilizácie výrazne podporuje trofiku tkanív. Senzomotorický tréning s využitím nestabilných plôch spolu so špecifickými ergoterapeutickými pomôckami prináša vysoko pozitívne výsledky pri obnove jemnej motoriky a dynamiky úchopu.



S. 1 – Schéma znázorňujúca štruktúrovanú mapu sledovaných parametrov a štatistických modelov použitých pri analýze skupiny pacientov 1 (po zlomenine zápästia - život neohrozujúci stav).

Druhú kohortu pacientov, reprezentujúcu komplexné a život ohrozujúce ochorenie, tvorilo 26 pacientov (12 žien, z toho 4 zosnulé, 14 mužov, z toho 6 zosnulých) s diagnostikovaným intramedulárnym osteosarkómom, ktorých parametre boli monitorované každý mesiac. V prípade tejto skupiny boli dvaja

pacienti vyradení, pretože odstúpili zo štúdie alebo sa nezúčastnili kontrolných vyšetrení. Štruktúrovaná mapa je znázornená schémou (S. 2). Sledovanými parametrami boli čas potrebný na mobilizáciu pacienta do vzpriamenej polohy a čas behu pacienta na 30-metrovej dráhe dva roky po začiatku liečby ako kvázi objektívny parameter. Použitie balneologických procedúr a hydrokinezioterapie v termálnej vode výrazne uľahčuje reedukáciu chôdze vďaka vztlaku vody, čím sa maximalizuje podpora oslabených dolných končatín. Dávkovanie fyzickej záťaže u onkologických pacientov striktnie rešpektuje aktuálnu toleranciu organizmu, čím zabezpečuje optimálnu energetickú rovnováhu a podporuje celkovú vitalitu. Ako subjektívny parameter sme zvolili „pocit bolesti“ v končatine po fyzickej záťaži a „celkovú chuť do jedla“ v závislosti od času po radikálnej resekcii. Ako objektívny parameter boli porovnané koncentrácie vápnika v krvi získané dvoma rôznymi laboratórnymi metódami s cieľom určiť, či tieto techniky poskytujú štatisticky ekvivalentné výsledky.



S. 2 – Schéma znázorňujúca štruktúrovanú mapu sledovaných parametrov a štatistických modelov použitých pri analýze skupiny pacientov 2 (intramedulárny osteosarkóm - život ohrozujúci stav).

3 VÝSLEDKY A DISKUSIA

3.1 JEDNODUCHÁ DESKRIPTÍVNA ŠTATISTIKA A MATICA HODNÔT

Jednoduchá deskriptívna štatistika je najvhodnejšia pre malé skupiny pacientov. Ukazovatele, ako sú priemer, modus, medián, variabilita a štandardná odchýlka, pomáhajú určiť pozíciu jednotlivca v rámci súboru^{13,14}. Tieto štatistiky však nie sú postačujúce na porovnávanie veľkých heterogénnych skupín pacientov pri určitej úrovni pravdepodobnosti alebo hladine významnosti. Jednoduchá štatistika nedokáže definitívne potvrdiť ani vyvrátiť hypotézy týkajúce sa zmien sledovaného parametra po liečbe, pretože nezohľadňuje normálne rozdelenie. Napriek tomu môže účinne kategorizovať skupiny pacientov. Pojmy „average“ a „mean“ sa často používajú zameniteľne. Pojem „average“ sa konkrétne vzťahuje na aritmetický priemer, teda súčet hodnôt vydelený počtom meraní, zatiaľ čo pojem „mean“ môže zahŕňať aj iné typy priemerov, napríklad geometrický a harmonický priemer. Algoritmus K-means, často používaný v kombinácii s umelou inteligenciou, sa využíva na klastrovanie a porovnávanie veľkých dátových súborov¹⁵. Bežné deskriptívne štatistiky možno rýchlo vypočítať pomocou kalkulačky alebo programu Microsoft Excel® (tab. 1).

Tabuľka 1 – Najčastejšie veličiny jednoduchej deskriptívnej štatistiky.

Kvantita	Aký výsledok získame ?	Syntax v MS. Excel®
Modus	Najčastejšie sa vyskytujúca hodnota meraného parametru	=MODE(select_values)
Medián	Určuje hodnotu parametra, od ktorej je polovica údajov väčšia a polovica údajov menšia.	=MEDIAN(select_values)
Aritmetický priemer	Priemerná hodnota parametra vypočítaná ako súčet hodnôt vydelený počtom pozorovaní.	=AVERAGE(select_values)
Smerodajná odchýlka (SD)	Presnosť, efektívnosť alebo správnosť. SD opisuje, ako sa hodnoty meraného parametra líšia od priemeru.	=STDEV.P(select_values)
Variabilita	Miera variability, umocnením na ² získame smerodajnú odchýlku.	=VAR.P(select_values)

* Veličina je uvedená iba pre potreby pokročilej štatistiky.

Pri aplikácii deskriptívnej štatistiky na údaje pacientov je nevyhnutné použiť viacero deskriptívnych veličín na komplexný opis, keďže spoliehanie sa na jediný ukazovateľ nemusí byť postačujúce¹⁶. Tabuľka 2 uvádza deskriptívnu štatistiku pre parameter „pocit bolesti“ u pacientov po liečbe zlomeniny zápästia a po resekcii intramedulárneho osteosarkómu (tab. 2). Napríklad najčastejšia odpoveď (modus) pred intervenciou pri liečbe zlomeniny zápästia je 1, čo naznačuje, že väčšina pacientov pred liečbou uvádzala bolesť. Ak sú modus a medián numericky podobné, naznačuje to normálne rozdelenie parametra, ktoré je vhodné pre komplexnejšie analýzy. Naopak, výrazný rozdiel medzi modusom a mediánom poukazuje na nenormálne rozdelenie, čo ovplyvňuje výber a nastavenie pokročilých štatistických modelov¹⁶. Ak sa modus približuje priemeru, aritmetický priemer je vhodným deskriptorom pre „pocit bolesti“. Štandardná odchýlka poskytuje informáciu o presnosti odhadov bolesti pacientov vo vzťahu k priemeru ich skupiny. Porovnanie deskriptívnej štatistiky pred liečbou, počas liečby a po liečbe ukazuje postupný pokles „pocitu bolesti“ počas šiestich mesiacov. Porovnanie pacientov ukázalo, že pacienti po resekcii osteosarkómu mali dlhšie obdobie zotavovania v porovnaní s pacientmi absolvujúcimi rehabilitáciu po zlomenine zápästia.

Tabuľka 2 – Ukazovatele deskriptívnej štatistiky subjektívneho parametra „pocit bolesti“ vo vzorke pacientov s rôznymi diagnózami pred liečbou, počas rehabilitácie a po ukončení liečby.

	Pacienti so zlomeninou zápästia			Pacienti s intramedulárnym osteosarkómom		
	Pred liečbou	3 mesiace po začiatku liečby	Po liečbe	Pred liečbou	3 mesiace po začiatku liečby	Po liečbe
Modus	1	2	4	1	1	4
Medián	1	2	4	1	1	3
Priemer	1.4	2.3	3.75	2	1.3	3.1
SD	± 0.5	± 0.9	± 1.2	± 1.3	± 0.63	± 1.1

* 1=áno, 2=skôr áno, 3=neviem, 4=skôr nie, 5=nie

Čo však znamená pojem zotavenie a aký význam má v prípade závažného ochorenia, pri ktorom sa niektorí pacienti vyliečia, zatiaľ čo významný počet pacientov zomiera? Jednoduchá štatistika nedokáže poskytnúť túto odpoveď v jednom kroku, preto sú potrebné komplexnejšie modely¹⁷⁻¹⁹. Interpretácia jednoduchej štatistiky môže byť náročná a môže viesť k nesprávnym záverom, najmä pri komplexných terapiách a mätúcich faktoroch^{17,18}. V takýchto prípadoch môžu byť numerické hodnoty zavádzajúce a na ich správnu interpretáciu je potrebná skúsenosť. So zvyšujúcim sa počtom pacientov narastá aj komplexnosť a interpretovateľnosť jednoduchej štatistiky.

Pokročilé štatistické modely založené na normálnom rozdelení populácie zjednodušujú interpretáciu tým, že umožňujú potvrdiť alebo vyvrátiť hypotézy s vysokou mierou pravdepodobnosti. Tieto modely integrujú viaceré jednoduché štatistické ukazovatele do bezrozmerného čísla reprezentujúceho normálnu populáciu pacientov²⁰. Môžu potvrdiť alebo vyvrátiť hypotézy na rôznych hladinách pravdepodobnosti tým, že zahŕňajú normálne rozdelenie očakávaných hodnôt parametrov a zohľadňujú malé počty pacientov s extrémnymi hodnotami parametrov, ktoré môžu byť dôsledkom faktorov, ako sú hypochondria, hypersenzitivita alebo znížená citlivosť.

Pokročilé modely poskytujú bezrozmerné číslo, ktoré kombinuje viaceré jednoduché štatistické ukazovatele s cieľom reprezentovať normálnu populáciu pacientov. Tento prístup však obetuje detailné numerické informácie, ako je priemerný vek alebo priemerná svalová sila stisku. Namiesto toho získavame informáciu o platnosti našich predpokladov prostredníctvom konkrétnych výsledkov typu „áno“ alebo „nie“. Základ komplexnej štatistiky spočíva v intervaloch spoľahlivosti založených na distribučnej funkcii, ktoré definujú maximálne prípustné horné a dolné odchýlky hodnôt meraného parametra²¹.

3.2 SYMETRICKÉ A ASYMETRICKÉ INTERVALY SPOĽAHLIVOSTI

Hornú a dolnú odchýlku meraných parametrov od očakávanej priemernej hodnoty možno definovať buď pomocou štandardnej odchýlky (SD), alebo intervalu spoľahlivosti (CI). Tieto metriky nám umožňujú rozlišovať medzi dvoma súbormi pacientov s identickým modusom, mediánom a aritmetickým priemerom, ktoré sa však líšia presnosťou alebo účinnosťou merania. Hoci SD aj CI vyjadrujú presnosť, poskytujú odlišnú interpretáciu. SD udáva priemernú odchýlku od priemeru, pričom menšie hodnoty naznačujú presnejšie merania alebo menšiu variabilitu medzi hodnotami pacientov. Naopak, CI opisuje rozsah, v ktorom sa meraná hodnota pravdepodobne nachádza s určitou pravdepodobnosťou. Menšie hodnoty naznačujú presnejší odhad a odrážajú vyššiu spoľahlivosť. Hladina spoľahlivosti (CL) určuje pravdepodobnosť CI a v medicínskej štatistike sa zvyčajne stanovuje na 95 %²². Nižšie hladiny spoľahlivosti pripúšťajú vo výpočte väčšie odchýlky. V medicínskych aplikáciách sa CI často vypočítava ako distribučná funkcia priemeru (CI_{mean}), pričom sa predpokladá normálne rozdelenie sledovaného parametra opísané Gaussovou krivkou, ktorá zahŕňa najpravdepodobnejšie aj najmenej pravdepodobné merané hodnoty (R. 1).

$$CI_{mean} = \bar{x} \pm \frac{z \cdot SD}{\sqrt{n}} \quad (R. 1)$$

Kde $\bar{x}$ je výberový priemer, n je veľkosť výberu a z je z-skóre zodpovedajúce požadovanej hladine spoľahlivosti. Z-skóre ako štandardizovaná hodnota udáva počet štandardných odchýlok od priemeru pre konkrétne intervaly spoľahlivosti. Bežné hodnoty z a zodpovedajúce p -hodnoty sú uvedené v tabuľke (tab. 3). P -hodnota, často uvádzaná vo vedeckých publikáciách, predstavuje pravdepodobnosť získania pozorovaného alebo ešte extrémnejšieho výsledku za predpokladu, že nulová hypotéza je pravdivá. Nižšia p -hodnota teda naznačuje, že pozorovaný výsledok je menej pravdepodobný pri platnosti nulovej hypotézy.

Tabuľka 3 – Hodnoty Z na výpočet konfidenčných intervalov pri rôznych hladinách spoľahlivosti.

Hladina spoľahlivosti	80 %	90 %	95 %	99 %
z-hodnota	1.282	1.645	1.960	2.576
p-hodnota (obojsmerný test)	0.2	0.1	0.05	0.01

Intervaly spoľahlivosti sú v pokročilých štatistických modeloch kľúčové na určenie toho, ktoré merané hodnoty patria do interpretačného rozsahu. V tejto štúdii boli CI_{mean} pre svalovú silu stisku stanovené pomocou dynamometra (Kern MAP 130K1) vo vzorke 40 pacientov (23 žien, 17 mužov). Táto metóda poskytuje objektívne merania bez subjektívneho skreslenia. Vypočítané CI_{mean} pre rôzne hladiny spoľahlivosti sú uvedené v tabuľke (tab. 4).

Tabuľka 4 – Vypočítané hodnoty dolnej a hornej hranice CI pre dve rôzne hladiny spoľahlivosti. Hodnoty predstavujú svalovú silu stisku ruky meranú dynamometrom.

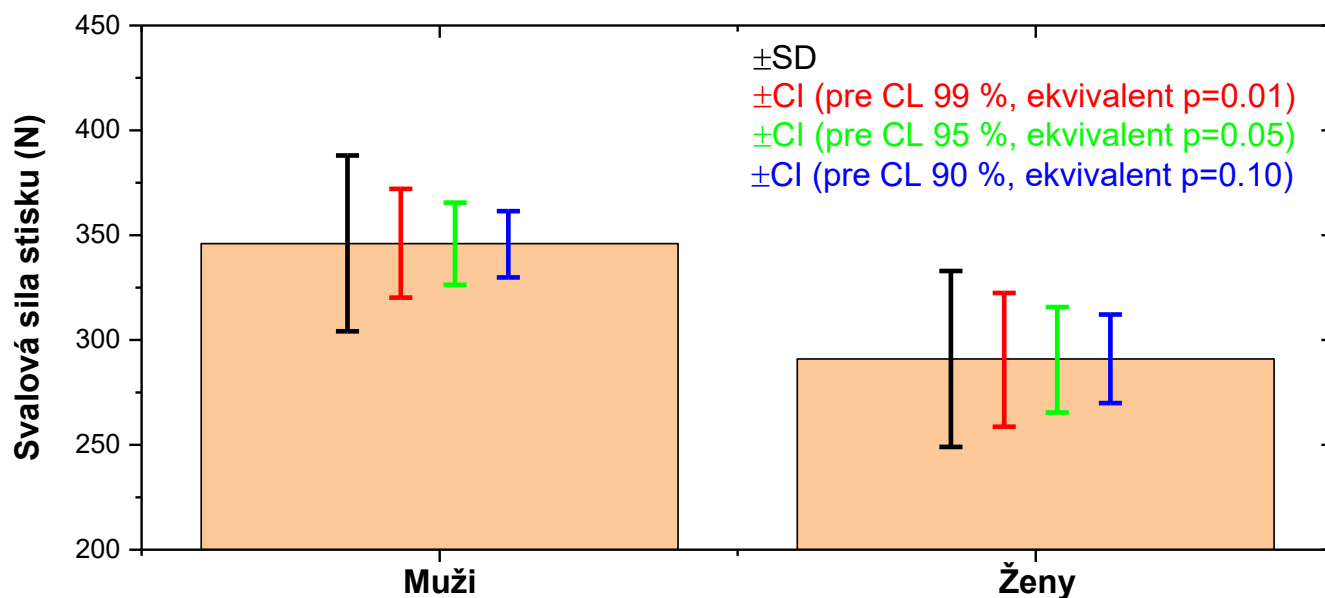
Pohlavie	Vstupné veličiny			Hladina spoľahlivosti 99 % ($z=2.576$, $p=0.01$)		Hladina spoľahlivosti 95 % ($z=1.960$, $p=0.05$)		Hladina spoľahlivosti 90 % ($z=1.645$, $p=0.10$)	
				$CI_{horný}$	$CI_{spodný}$	$CI_{horný}$	$CI_{spodný}$	$CI_{horný}$	$CI_{spodný}$
Muži	n	$\bar{x}$	SD	372	320	366	326	362	329
Ženy	23	291	60	323	258	316	266	312	270

*Jednotky = Newton

Obrázok (obr. 2) znázorňuje rozdiel v CI_{mean} a ich zmenšovanie so znižujúcou sa hladinou spoľahlivosti (CL). Červené, zelené a modré stĺpce predstavujú CI pre hladiny spoľahlivosti 99 % ($p = 0,01$), 95 % ($p = 0,05$) a 90 % ($p = 0,10$). Výška stĺpcov udáva aritmetický priemer svalovej sily stisku u mužov a žien.

Na obrázku 2 zostáva štandardná odchýlka, znázornená ako intervaly $\pm$ okolo priemeru, konštantná bez ohľadu na hladinu spoľahlivosti (CL) a zahŕňa všetky potenciálne pravdepodobnosti a extrémne chyby. Na akceptovanie 1 % pravdepodobnosti chyby vypočítame CI_{mean} pri 99 % hladine spoľahlivosti (obr. 2). Výsledný CI_{mean} je menší ako pôvodný interval definovaný SD. So znižovaním CL z 99 % na 95 % a napokon na 90 % sa CI_{mean} zužuje, čím sa vylučuje viac dátových bodov a pripúšťa sa viac chýb. Toto zužovanie CI_{mean} ukazuje, ktoré údaje pacientov zostávajú štatisticky spoľahlivé napriek chybám merania. V našom výskume CI_{mean} na vertikálnej osi definuje maximálne a minimálne prijateľné hodnoty svalovej sily. Každý CI_{mean} implicitne zahŕňa variabilitu opísanú pomocou SD, čím odpadá potreba samostatných chybových úsečiek pre SD. Pre presnú interpretáciu údajov sa musíme rozhodnúť, či chceme zobrazit' rozptyl všetkých meraní okolo priemeru (SD), alebo definovať interval spoľahlivých hodnôt (CI), ktorý vylučuje určité chybové údaje medzi pacientmi.

Častou chybou v odborných prácach je implicitný predpoklad symetrického rozdelenia CI_{mean} okolo priemeru. Symetrické intervaly spoľahlivosti sa zvyčajne získavajú vtedy, keď sa predpokladá symetrické výberové rozdelenie odhadovanej štatistiky, napríklad pri predpoklade normality. Iné situácie si však môžu vyžadovať odlišné distribučné funkcie, čo vedie k asymetrickým CI.



Obrázok 2 – Priemerné hodnoty $\bar{x}$ svalovej sily u mužov a žien pred rehabilitáciou zápästia (stĺpce) spolu s vypočítanými intervalmi spoľahlivosti založenými na distribučnej funkcii priemeru. Výpočty boli vykonané pre tri rôzne hladiny spoľahlivosti: červená) 99 %, zelená) 95 %, modrá) 90 %. Vypočítané intervaly spoľahlivosti, ktoré ohraničujú oblasť spoľahlivosti údajov s rôznymi pravdepodobnosťami, sú porovnané so štandardnými odchýlkami označenými čiernou farbou.

Ak údaje nesledujú normálne rozdelenie, mali by sa použiť neparametrické intervaly spoľahlivosti. Jednou z bežných metód je bootstrapping (iteračný prístup), ktorý nepredpokladá konkrétne známe rozdelenie a vypočítava $CI_{bootstrap}$ opakovaným náhodným výberom údajov zo súboru s opakovaným návratom a následným výpočtom požadovanej štatistiky. Tento výpočet možno jednoducho vykonať pomocou štatistického softvéru. Pri práci s malým súborom sa namiesto z-rozdelenia používa t-rozdelenie. Interval spoľahlivosti pre priemer pri t-rozdelení CI_t sa vypočíta pomocou rovnice 2.

$$CI_t = \bar{x} \pm t \cdot SD / \sqrt{n} \quad (R. 2)$$

Kde t je t-skóre zodpovedajúce požadovanej hladine spoľahlivosti a počtu stupňov voľnosti (f_{deg}). Počet stupňov voľnosti možno jednoducho vypočítať z veľkosti súboru n pomocou vzťahu $n - 1$. Pri

podieloch alebo percentách poskytuje Wilsonov interval presnejší interval spoľahlivosti, najmä pri malých veľkostiach súboru alebo extrémnych podieloch. Hoci Wilsonov interval spoľahlivosti CI_{Wilson} poskytuje lepšie pokrytie než mnohé iné metódy, jeho výpočet je zložitejší, pretože zahŕňa kvadratickú rovnicu a komplikovanejší postup riešenia (R. 3), avšak je relevantný aj pre malé súbory pacientov.

$$CI_{Wilson} = \frac{\hat{p} + \frac{z^2}{2n} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} \quad (R. 3)$$

Tu $\hat{p}$ predstavuje výberový podiel, n je počet vzoriek a z je z-skóre zodpovedajúce požadovanej hladine spoľahlivosti. Zaujímavé je, že SPSS neobsahuje špecifický postup pre tento výpočet, umožňuje ho však vykonať pomocou príkazov. Intervaly spoľahlivosti pre rozptyl a štandardnú odchýlku vyžadujú chí-kvadrát rozdelenie χ^2 a sú definované ako interval medzi príslušnými hraničnými bodmi (R. 4).

$$CI_{variancia} = (spodný, horný) = \left(\frac{(n-1)SD^2}{\chi_{spodný}^2}, \frac{(n-1)SD^2}{\chi_{horný}^2} \right) \quad (R. 4)$$

Kde $\chi_{spodný}^2$ a $\chi_{horný}^2$ predstavujú kritické hodnoty chí-kvadrát pre požadovanú hladinu spoľahlivosti pri $n - 1$ stupňoch voľnosti. V prípadoch, keď sú údaje zošikmené, možno pred výpočtom intervalu spoľahlivosti použiť transformácie, napríklad logaritmickú transformáciu, na normalizáciu údajov. Alternatívne možno v štatistickom softvéri použiť metódu profilu vierohodnosti pre CI. Pri hodnotení biologických a medicínskych údajov odporúčame zobrazovať CI pri špecifikovanej hladine spoľahlivosti namiesto všeobecného intervalu $\pm SD$. Tento prístup zabezpečuje jasné pochopenie rozsahu, v ktorom sa pravé hodnoty parametra pravdepodobne nachádzajú, a poskytuje presnejší a spoľahlivejší základ pre štatistickú interpretáciu.

V analýze prežívania sa často používajú Greenwoodove intervaly spoľahlivosti CI_G . Greenwoodov vzorec poskytuje mieru presnosti odhadov prežívania v Kaplanovej-Meierovej analýze prežívania tým, že odhaduje rozptyl pravdepodobnosti prežívania v každom časovom bode. Tento rozptyl sa používa na konštrukciu intervalov spoľahlivosti okolo Kaplanových-Meierových odhadov prežívania a označuje rozsah, v ktorom sa pravá pravdepodobnosť prežívania pravdepodobne nachádza. Vzorec vypočítava rozptyl na základe počtu pacientov v riziku a počtu udalostí, napríklad úmrtí, v každom časovom bode. Pri použití normálnej aproximácie sú intervaly spoľahlivosti definované rovnicou (R. 5):

$$CI_G = S \pm Z\sqrt{Var_G} \quad (R. 5)$$

Kde S je Kaplanov-Meierov odhad prežívania v konkrétnom čase, Var_G je rozptyl odhadu prežívania v tomto čase vypočítaný pomocou všeobecne známeho Greenwoodovho vzorca a Z je hodnota Z zo štandardného normálneho rozdelenia zodpovedajúca požadovanej hladine spoľahlivosti. Tieto intervaly poskytujú štatistický rozsah, ktorý pravdepodobne obsahuje skutočnú pravdepodobnosť prežívania, a kvantifikujú neistotu odhadov prežívania.

3.3 KORELAČNÁ ANALÝZA: HODNOTENIE ÚČINNOSTI LIEČBY A ZLEPŠENIA ZDRAVOTNÉHO STAVU

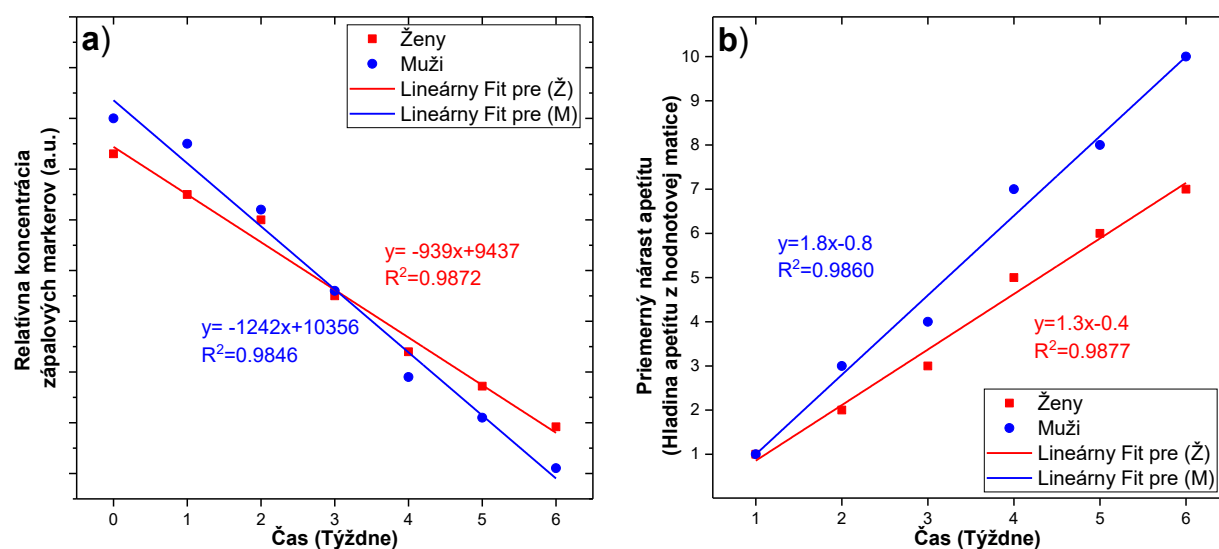
Metódy korelačnej analýzy nám umožňujú potvrdiť dva základné predpoklady. Po prvé, korelačná analýza pomáha určiť, či existuje závislosť medzi časovo závislým trendom a sledovaným parametrom. Po druhé, poskytuje informáciu o tom, či je vývoj parametra ovplyvnený definovaným faktorom²³. Pri monitorovaní zdravotných parametrov súvisiacich s progresiou ochorenia môže korelačná analýza posúdiť, či daná liečba ovplyvňuje zdravotný stav pacienta, teda smer účinku, a zároveň silu tohto účinku²⁴. Okrem toho táto metóda umožňuje sledovať účinky v rôznych podskupinách definovaných podľa veku, pohlavia alebo špecifických komorbidít. Hoci sa v oboch scenároch používajú rovnaké matematické nástroje, ich interpretácia je odlišná.

Korelačnú analýzu v medicínskom výskume možno rozdeliť na niekoľko typov. Logaritmické a exponenciálne korelácie sa používajú vtedy, keď je vzťah medzi premennými nelineárny a možno ho presnejšie opísať pomocou logaritmickú škály. Patrí sem aj použitie metód, pri ktorých elektromagnetické žiarenie alebo akustické vlny interagujú s telesnými štruktúrami vykazujúcimi vysoký a frekvenčne závislý rozptyl. Ďalším prístupom je kombinácia lineárnej a logaritmickú korelácie pri výskume enzymatických reakcií a receptorových odpovedí, keď sa získaný graf rozdelí na dve alebo viac rôznych funkcií, pričom každá sa koreluje samostatne. Linearizáciu nelineárnych údajov možno dosiahnuť matematickými transformáciami alebo fitovaním kriviek, korelačná analýza by však mala byť primárne vyhradená pre jednoznačne lineárne závislosti.

Lineárna korelácia sa bežne používa v oftalmológii, medicínskom zobrazovaní, endokrinológii, onkológii, epidemiológii, kardiológii, genetike a genomike, neurológii a farmakológii. Uplatňuje sa pri optických a nízkorozptylových spektroskopických metódach, validácii biomarkerov, štúdiách účinnosti liekov a analýze EKG, kde je vzťah medzi premennými reprezentovaný priamkou ako funkciou intenzity signálu po interakcii so vzorkou. Sila lineárnej korelácie je reprezentovaná smernicou trendovej čiary, ktorá indikuje účinnosť liečby alebo progresiu biomarkera. Funkcia zobrazenia lineárnej trendovej čiary predstavuje lineárne prispôbenie údajov a je definovaná rovnicou 6.

$$y = \text{smernica} \cdot x + \text{výsek} \quad (\text{R. 6})$$

Tu kladná alebo záporná smernica opisuje rýchlosť nárastu alebo poklesu závislosti, zatiaľ čo priesečník s osou y označuje odhadovanú hodnotu pri $x = 0$. Vo všeobecnosti väčšia smernica znamená silnejšiu závislosť. Presnosť monitorovania parametra možno posúdiť na základe vzdialenosti bodov v grafe od trendovej čiary, ktorá je opísaná korelačným koeficientom. Najčastejšie používanými lineárnymi korelačnými koeficientmi pre lineárne závislé údaje sú Pearsonov korelačný koeficient alebo koeficient determinácie R^2 . Pri lineárne závislých údajoch možno R^2 jednoducho získať pomocou zobrazenia trendovej čiary v programe MS Excel® (obr. 3). V medicínskom výskume prahové hodnoty R^2 potvrdzujú alebo vyvracajú koreláciu. Hodnoty blízke 1 naznačujú výbornú lineárnu koreláciu v dôsledku liečby, zatiaľ čo hodnoty blízke nule poukazujú na absenciu významného účinku na sledovaný parameter.



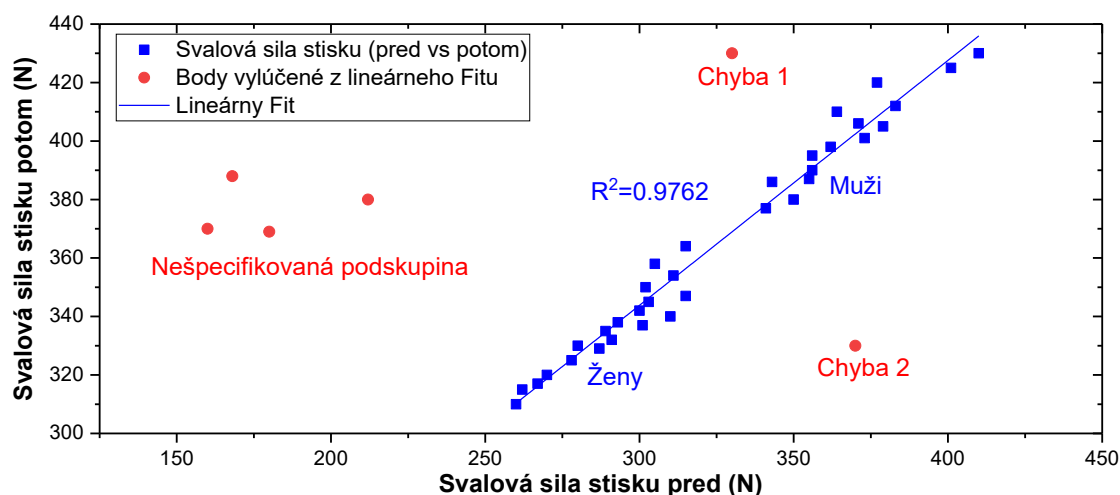
Obrázok 3 – Získané korelácie u pacientov po resekcii osteosarkómu: a) priemerná koncentrácia zápalových markerov v krvnom sére a b) zvýšenie chuti do jedla v závislosti od času. Vo všetkých prípadoch je koeficient determinácie R^2 blízky hodnote 1, čo naznačuje výbornú spoľahlivosť.

Pri objektívne získaných údajoch, napríklad z meracieho zariadenia alebo laboratória, sa korelácia potvrdzuje pri $R^2 \geq 0,95$. Pri kvázi objektívnych údajoch, odhadovaných odborníkom, sa korelácia zohľadňuje pri $R^2 \geq 0,90$. Pri subjektívnych údajoch, založených na odpovediach pacienta, by sa korelácia mala potvrdiť pri $R^2 \geq 0,80$. Prahové hodnoty sa znižujú v dôsledku narastajúcej miery chýb a nepresností, ktoré pri klesajúcej objektivite získavania údajov zastierajú skutočnú hodnotu parametra^{19,25}. Je dôležité

poznamenať, že korelácia neznamená kauzalitu, pretože výsledky môžu byť ovplyvnené mäťúcimi premennými. Presnejšie závery preto musia byť potvrdené ďalšími testami.

Na obrázku 3 pozorujeme trendové čiary pre koncentráciu zápalových markerov a zvýšenie chuti do jedla v čase po resekcii osteosarkómu u mužov a žien. Pokles zápalových markerov je u mužov o $32 \pm 2 \%$ intenzívnejší, napriek podobným hodnotám R^2 (obr. 3a). Naopak, zvýšenie chuti do jedla stanovené pomocou fuzzy logiky je u mužov vyššie o $27 \pm 2 \%$ (obr. 3b). Korelačné koeficienty týchto trendov môžu byť podobné, ale nesúvisia so smernicami trendových čiar.

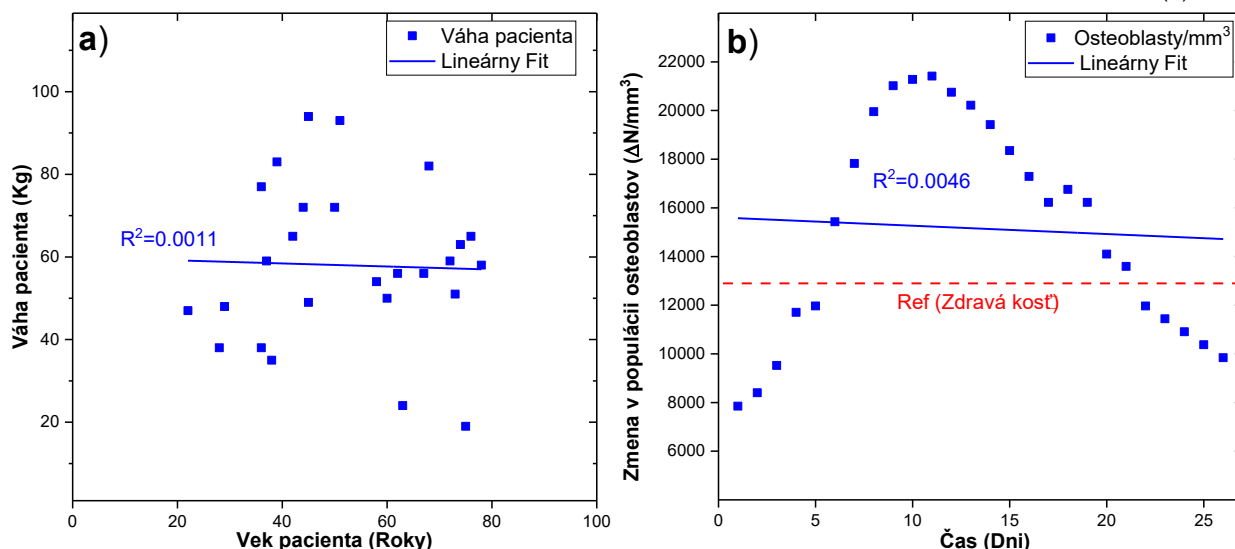
Ďalšou situáciou je klastrovanie. Grafické znázornenie svalovej sily ruky pred liečbou zlomeniny zápästia a po nej (obr. 4) ukazuje dva klastre bodov v okolí trendovej čiary a jeden klaster mimo nej. Tieto klastre možno zvyčajne priradiť ku klinickým údajom, čo naznačuje, že svalová sila stisku sa počas liečby v oboch podskupinách menila podobne. Body mimo týchto klastrov sa považujú za chyby alebo môžu predstavovať extrémne prípady, prípadne neznáme podskupiny so špecifickými komorbiditami. Môžeme tiež odhadnúť, či dochádza k zlepšeniu alebo zhoršeniu zdravotného stavu. Ak nedochádza k žiadnej zmene, namerané hodnoty svalovej sily pred liečbou a po nej by boli podobné, čo by reprezentovala trendová čiara so smernicou blízko $\beta = 45^\circ$ (*smernica* = $\tan \beta$). Čím viac sa smernica odlišuje od 45° , tým intenzívnejšia je zmena zdravotného stavu. Pri parametroch, ktoré nevykazujú lineárnu závislosť, lineárna korelácia nedokáže účinne hodnotiť zlepšenie alebo zhoršenie zdravotného stavu.



Obrázok 4 – Závislosť svalovej sily stisku ruky pred liečbou zlomeniny zápästia a po nej. Pozorovateľné sú dve skupiny bodov (modré) v okolí trendovej čiary, ktoré patria ženám a mužom. Výborná korelácia $R^2 = 0,9762$ naznačuje zlepšenie svalovej sily počas liečby. Vylúčené body (červené) predstavujú zjavné odchýlky od trendovej čiary. Neznámu skupinu nebolo možné kategorizovať podľa dostupných údajov.

Obrázok 5 znázorňuje výsledky, pri ktorých nie je možné použiť lineárnu koreláciu. Na obr. 5a vedie absencia systematického usporiadania bodov okolo priamky k nízkemu koeficientu determinácie, čo naznačuje, že medzi premennými neexistuje lineárny vzťah. Naopak, obr. 5b zobrazuje obrátenú parabolickú krivku, ktorá naznačuje nelineárnu závislosť napriek extrémne nízkej hodnote R^2 . To poukazuje skôr na možný vzťah odmocninového typu než na lineárnu závislosť.

Ak lineárna korelácia nie je postačujúca, sú potrebné pokročilejšie metódy. Parciálna korelácia kontroluje vplyv ďalších premenných, čím poskytuje jasnejší pohľad na vzťah medzi dvoma premennými. Viacnásobná korelácia skúma vzťahy medzi viac ako dvoma premennými a umožňuje širšie pochopenie komplexných interakcií v medicínskych údajoch. Kľúčové je aj zabezpečenie štatistickej významnosti, ktorá sa zvyčajne hodnotí pomocou p-hodnôt s cieľom potvrdiť, že pozorované vzťahy nie sú dôsledkom náhody. Použitím týchto pokročilých metód a dôsledného štatistického testovania možno dosiahnuť presnejšiu a komplexnejšiu analýzu medicínskych údajov.



Obrázok 5 – Problémy spojené s nízkym korelačným koeficientom môžu zahŕňať: a) závislosť veku populácie od hmotnosti, pri ktorej je korelačný koeficient použiteľný a poukazuje na nezávislosť premenných. Naopak, b) časová závislosť cyklicky aplikovanej magnetoterapie na rast osteoblastov, ktorá je nelineárna, ale správna, a preto nemožno použiť R^2 .

3.4 POKROČILÁ ŠTATISTICKÁ ANALÝZA V MEDICÍNSKOM VÝSKUME

Na rozdiel od jednoduchšej deskriptívnej štatistiky, ktorá poskytuje numerické informácie o vzorkách alebo skupinách, pokročilé štatistické modely založené na normálnom rozdelení potvrdzujú alebo zamietajú nulovú hypotézu (H_0). Preto je nevyhnutné formulovať H_0 ešte pred začatím akýchkoľvek výpočtov. Hypotézou môže byť akékoľvek tvrdenie, ktoré bude buď potvrdené, alebo jednoznačne zamietnuté. Hranica medzi platnosťou a zamietnutím hypotézy je definovaná hladinou významnosti (α), ktorá je odvodená od normálneho rozdelenia základnej populácie²⁶.

Medzi základné predpoklady normálnej populácie patrí heterogénna populácia s normálne rozdelenými hodnotami parametrov, zahŕňajúca pozitívne aj negatívne extrémny. Podľa centrálnej limitnej vety sa rozdelenie výberových priemerov so zvyšujúcou sa veľkosťou výberu približuje normálnemu rozdeleniu bez ohľadu na rozdelenie populácie, pokiaľ je veľkosť výberu dostatočne veľká. V normálnom rozdelení predstavuje α zvolenú prahovú hodnotu, ktorá sa zvyčajne nastavuje na konvenčné hodnoty, ako sú 0.1, 0.05 alebo 0.01. Čím menšia je hladina významnosti, tým menšia je pravdepodobnosť pozorovania rozdielu výlučne v dôsledku náhody. Pri použití štatistických testov založených na normálnom rozdelení je nevyhnutné overiť predpoklad normality. Možno to vykonať pomocou testov normality, napríklad Shapirovho-Wilkovho testu, alebo posúdením Q-Q grafov. Správny výber hladiny významnosti pomáha zmierniť vplyv chýb, odľahlých hodnôt a extrémnych hodnôt. V štatistickom výskume manuálne volíme prahovú hladinu významnosti. Táto hodnota presne určuje referenčnú hranicu pre zamietnutie hypotézy, známu aj ako kritická hodnota. Pre merané objektívne parametre sa odporúča hladina významnosti $\alpha = 0,05$. Pre kvázi objektívne parametre získané lekárom sa navrhuje $\alpha = 0,1$, zatiaľ čo pre subjektívne parametre definované pacientom a jeho vnímaním sa odporúča $\alpha = 0,2$.

Pri testovaní hypotéz je nevyhnutné zohľadniť možnosť výskytu chýb. Chyba I. druhu nastáva vtedy, keď je pravdivá H_0 zamietnutá, teda ide o falošne pozitívny výsledok, zatiaľ čo chyba II. druhu nastáva vtedy, keď falošná H_0 nie je zamietnutá, teda ide o falošne negatívny výsledok. Hodnota α predstavuje pravdepodobnosť chyby I. druhu. P-hodnota vyjadruje pravdepodobnosť pozorovania testovacieho výsledku, prípadne ešte extrémnejšieho výsledku, za predpokladu platnosti H_0 . P-hodnota nižšia ako zvolená α naznačuje, že H_0 možno zamietnuť, čo poukazuje na štatistickú významnosť. Sila štatistického testu predstavuje pravdepodobnosť správneho zamietnutia falošnej nulovej hypotézy. Závisí od hladiny významnosti, veľkosti výberu a veľkosti efektu. Vyššia sila testu znižuje pravdepodobnosť chyby II. druhu.

Zvolená kritická hodnota hladiny významnosti α sa následne prepočítava na kritickú F-hodnotu (F_{crit}) a kritickú T-hodnotu (T_{crit}) pre F-test rozptylu, respektíve Studentov T-test. Kritické hodnoty F_{crit} a T_{crit} majú pri rovnakej hodnote α rozdielne hodnoty. Tieto hodnoty možno jednoducho získať z príslušných štatistických tabuliek alebo pomocou programu Microsoft Excel®²⁷. Okrem kritických hodnôt je potrebné vypočítať deskriptívnu F-hodnotu (F) a T-hodnotu (T), ktoré opisujú naše dátové súbory. Tieto hodnoty sú

výsledkom špecifických kombinácií jednoduchých ukazovateľov deskriptívnej štatistiky. Porovnaním kritických hodnôt s našimi vypočítanými hodnotami T a F určujeme, či sú vypočítané hodnoty nižšie ako príslušné kritické hodnoty. Ak sú hodnoty T a F nižšie ako ich príslušné kritické hodnoty, dátové súbory sa štatisticky významne nelíšia. Naopak, ak sú naše hodnoty vyššie ako kritické hodnoty, potvrdzujeme, že súbory sa štatisticky významne líšia. Ak sa tieto súbory významne líšia, môžeme konštatovať, že liečba viedla buď k zlepšeniu, alebo k zhoršeniu zdravotného stavu pacienta.

Pri použití F-testu a T-testu je dôležité dodržiavať ich príslušné predpoklady. F-test predpokladá homogenitu rozptylov, čo znamená, že rozptyly porovnávaných populácií by mali byť rovnaké. T-test predpokladá rovnosť rozptylov a normalitu údajov. Treba poznamenať, že ak nie je splnený predpoklad rovnosti rozptylov, možno použiť Welchov T-test. Okrem štatistickej významnosti je dôležité zohľadniť aj veľkosť efektu, ktorá meria silu vzťahu medzi premennými alebo veľkosť rozdielu medzi skupinami. Medzi bežné miery veľkosti efektu patrí Cohenovo d pre T-test a η -kvadrát pre F-test. Štatisticky významný výsledok nemusí vždy znamenať praktickú významnosť. Zatiaľ čo štatistická významnosť naznačuje, že efekt existuje, praktická významnosť hodnotí jeho reálnu relevanciu alebo dôležitosť. Štatistický súbor vždy vnímame ako napoly plnú a napoly prázdnu fľašu, pričom potvrdenie aj vyvrátenie konkrétnych hypotéz môže validovať rovnakú kauzalitu buď z negatívnej, alebo pozitívnej perspektívy.

3.4.1 F-TEST: HODNOTENIE ÚSPEŠNOSTI LIEČBY OSTEOSARKÓMU U MUŽOV A ŽIEN

Na určenie toho, či sa rozdelenie parametra v dvoch veľkých skupinách pacientov štatisticky významne líši, používame testovanie rozptylu, konkrétne F-test. Táto metóda je vhodná, pretože neskúmame párové hodnoty, a preto nie je potrebné, aby obe skupiny mali rovnakú veľkosť výberu²⁸. F-test nám umožňuje porovnávať rozptyly s cieľom identifikovať rozdiely v objektívnych, kvázi objektívnych a subjektívnych parametroch. Vykonali sme test rozptylu na analýzu schopnosti vertikalizácie u pacientov, ktorí podstúpili resekciu intramedulárneho osteosarkómu. Cieľom bolo určiť, či existujú významné rozdiely v schopnosti behu medzi mužskými a ženskými pacientmi dva roky po operácii. V dôsledku mortality a imobility nebola veľkosť súborov u mužov a žien rovnaká. Predbežná korelačná analýza naznačila rodové rozdiely v zlepšení vertikalizácie, pričom muži vykazovali kratšie a ženy dlhšie trvanie. Vplyv tohto rozdielu na schopnosť behu však nebol jasný. Predpokladali sme nulovú hypotézu (H_0), že dynamika alebo efektívnosť zotavenia sa medzi mužmi a ženami v remisii po liečbe intramedulárneho osteosarkómu bude líšiť. Výpočet a výsledky sú uvedené v tabuľke (tab. 5).

Tabuľka 5 – Štatistiky vypočítané z časov behu pacientov s rôznou dĺžkou liečby.

Potrebné veličiny	Muži (Čas behu/30m)	Ženy (Čas behu/30m)	Vzorec na výpočet údajov
Priemer	$\bar{x}_2 = 9.46$	$\bar{x}_2 = 10.12$	*
Variancia	$var_1 = 0.3301$	$var_2 = 0.4521$	*
Odhadovaná variancia	$var_{rel1} = 0.3405$	$var_{rel2} = 0.4689$	$= \frac{variancia \cdot počet\ pacientov}{počet\ pacientov - 1}$
Počet pacientov	$n_1 = 8$	$n_2 = 6$	
Stupeň voľnosti	$f_{deg1} = 7$	$f_{deg2} = 5$	$= počet\ pacientov - 1$
Hodnota F	<u>1.3771</u>		$= \frac{var_{rel1}}{var_{rel2}}$
Kritická hodnota F_{crit}	<u>3.3679</u>		- Výpočet v MS. Excel® pomocou =FINV(0.10;7;5) - Vyhľadať údaje v štatistickej tabuľke „F“

* Rovnice a syntax na výpočet v MS Excel® sú uvedené v kapitole „jednoduchá štatistika“.

Vzhľadom na kvázi objektívny charakter merania pomocou stopiek bola zvolená hladina významnosti $\alpha = 0,10$. Vypočítaná hodnota F bola nižšia ako kritická hodnota F_{crit} , čo naznačuje, že medzi skupinami

nebol zistený štatisticky významný rozdiel (tab. 5). Napriek rozdielom vo veľkosti výberov a priemerných časoch zotavenia teda skupiny nevykazovali významné rozdiely pri $\alpha = 0,10$. Mužskí aj ženskí pacienti preukázali podobnú bežecú výkonnosť bez ohľadu na dĺžku liečby. F-test bol vykonaný porovnaním rozptylov skupín pomocou ich pomeru. Tento pomer sa porovnáva s kritickou hodnotou z tabuľky F-rozdelenia, ktorá je určená zvolenou hodnotou α a počtom stupňov voľnosti každej skupiny. Okrem toho je dôležité overiť predpoklady F-testu, medzi ktoré patrí autonómia pozorovaní, normálne rozdelenie populácií a homogenita rozptylov. Ak sú tieto predpoklady porušené, výsledky F-testu nemusia byť spoľahlivé. F-test je kľúčovým nástrojom v metódach ANOVA (analýza rozptylu) alebo MANOVA (viacrozmerná analýza rozptylu), kde sa používa na porovnávanie rozptylov a testovanie významných rozdielov medzi viacerými dátovými súbormi. Vďaka tomu je F-test univerzálnym nástrojom pri analýze experimentálnych údajov, pri ktorých je potrebné porovnávať variabilitu.

3.4.2 T-TEST: HODNOTENIE ROZDIELOV MEDZI SKUPINAMI PACIENTOV

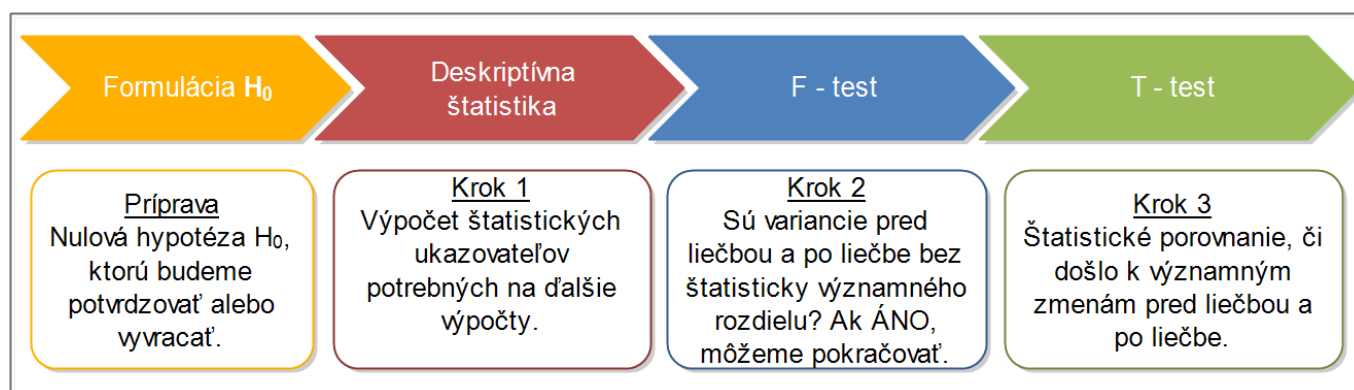
T-test (Studentov test) sa používa na porovnanie priemerov dvoch dátových súborov s cieľom určiť, či sa od seba štatisticky významne líšia. Existuje niekoľko prístupov k T-testu, pričom najčastejšie používanými variantmi sú jednovýberový, dvojvýberový (nepárový) a párový T-test. Jednovýberový T-test porovnáva priemer jednej skupiny pacientov s očakávanou referenčnou hodnotou a určuje, či sa priemer výberu významne líši od známeho alebo predpokladaného populačného priemeru. Dvojvýberový (nepárový) t-test porovnáva priemery dvoch nezávislých skupín pacientov a hodnotí, či sa priemery dvoch nesúvisiacich skupín významne líšia pri danej hodnote α . Párový T-test porovnáva dva súbory hodnôt, ktoré sú určitým spôsobom spárované alebo zodpovedajúce, napríklad merania pred liečbou a po liečbe u tých istých pacientov, pričom vyžaduje dvojice hodnôt striktne priradené ku každému pacientovi na určenie toho, či existuje významný rozdiel medzi párovými pozorovaniami. Pred výberom vhodného typu T-testu je nevyhnutné zvážiť, či porovnáваме celú skupinu s referenčnou hodnotou, s inou nezávislou skupinou, alebo či porovnáваме párové hodnoty v rámci jednej skupiny²⁹. Na rozdiel od F-testu, ktorý sa zameriava na rozptyly, sa T-test zameriava na priemery. F-test sa často používa v analýze rozptylu a viacrozmernej analýze rozptylu (ANOVA a MANOVA) na porovnanie viacerých skupín. Párový T-test špecificky pracuje s párovými údajmi, čo je aspekt, ktorý základný F-test zvyčajne nezohľadňuje.

3.4.2.1 DVOJVÝBEROVÝ NEPÁROVÝ T-TEST: HODNOTENIE ZMIEN ZDRAVOTNÉHO STAVU

Na porovnanie podobnosti alebo rozdielu medzi dvoma skupinami pacientov je vhodný dvojvýberový nepárový T-test. Tento test možno použiť ako doplnkový a overovací štatistický model ku korelačnej analýze. Dvojvýberový nepárový T-test umožňuje vypočítať štatisticky významné zmeny počas liečby. Kľúčovou výhodou T-testu oproti korelačnej analýze je jeho čisto matematický základ, ktorý eliminuje potrebu hodnotenia grafu. Nepárový T-test zároveň nevyžaduje dvojice hodnôt, teda hodnoty pred liečbou a po liečbe, pre každého pacienta, ako je to pri korelačnej analýze. Táto flexibilita umožňuje zahrnúť aj pacientov, ktorí sa nedostavili na kontrolné vyšetrenie alebo zomreli. Nie je nevyhnutné, aby hodnoty parametra pred liečbou a po liečbe patrili tomu istému pacientovi, keďže test sa zameriava na celú skupinu, nie na jednotlivých pacientov. Tento test však nemožno použiť na hodnotenie rozdielov v účinnosti dvoch rôznych liečebných metód, čo by si vyžadovalo sledovanie párových údajov priradených ku konkrétnym pacientom.

Na praktické vysvetlenie analyzujeme skúmaný dátový súbor. Každý zo 40 sledovaných pacientov so zlomeninou zápästia absolvoval dynamometrické meranie sily stisku postihnutej ruky. Prvé meranie bolo vykonané na konci tretieho mesiaca liečby, pričom bolo zaznamenané v prvom stĺpci v programe MS Excel®, a druhé meranie bolo vykonané po ukončení liečby, pričom bolo zaznamenané v druhom stĺpci v programe MS Excel®. Keďže svalovú silu nie je možné merať bezprostredne po úraze, prvé meranie sa považuje za meranie „pred liečbou“. Pri tejto metóde nie je dôležité poradie pacientov ani priradenie dvojíc hodnôt ku konkrétnym pacientom³⁰. Výpočet poskytuje presné výsledky bez ohľadu na to, či majú dva stĺpce rovnakú dĺžku, čo môže nastať vtedy, ak sa niektorí pacienti nedostavia na kontrolné vyšetrenie alebo v dôsledku ich úmrtia. Nezávislosť od dĺžky stĺpcov zvyšuje robustnosť a použiteľnosť metódy. H₀ znela

„Po liečbe sa zdravotný stav pacientov významne zlepšil.“ Postup výpočtu je opísaný v nasledujúcich krokoch, ktoré sú znázornené v schéme (S. 3).



S. 3 – Schematický diagram znázorňujúci postup získania výsledku pri použití dvojvýberového nepárového T-testu.

Krok 1) Základnú deskriptívnu štatistiku vypočítame samostatne pre stĺpce „pred liečbou“ a „po liečbe“ v programe MS Excel®. Vypočítané údaje sú uvedené v tabuľke (tab. 6). Najskôr použijeme F-test, ktorý overí podobnosť rozptylov oboch stĺpcov, napríklad skupín pacientov alebo štádií liečby. Ak sa nezistí štatisticky významný rozdiel, môžeme následne vypočítať T-test zodpovedajúci formulovanej hypotéze H_0 . Treba poznamenať, že existujú špecifické prípady, v ktorých možno porovnanie F a F_{crit} použiť ako štatistický model na hodnotenie rozdielu alebo podobnosti dvoch skupín. Odpovedá však na inú otázku, konkrétne či majú tieto dve skupiny rovnaký alebo rozdielny rozptyl sledovaného parametra. Môže teda porovnať, či sú si podobné skupina žien a mužov alebo skupina detí a dospelých. Nehovorí však o tom, či sa jedna skupina významne zmenila v porovnaní s referenčnou skupinou.

Tabuľka 6 – Štatistiky dynamometricky meranej svalovej sily stisku ruky pred liečbou zlomeniny zápästia a po nej.

Potrebné veličiny	Vypočítané zo stĺpca I (3. mesiac po zranení)	Vypočítané zo stĺpca II (po liečbe)	Rovnica
Priemer	$\bar{x}_1 = 314$	$\bar{x}_2 = 378$	*
Variancia	$var_1 = 3574$	$var_2 = 2889$	*
Odhadovaná variancia	$var_{rel1} = 3666$	$var_{rel2} = 2964$	$= \frac{\text{variancia} \cdot \text{počet pacientov}}{\text{počet pacientov} - 1}$
Počet pacientov	$n_1 = 40$	$n_2 = 40$	= počet pacientov v stĺpci
Stupeň voľnosti	$f_{deg1} = 39$	$f_{deg2} = 39$	= počet pacientov - 1
Hodnota F	<u>1.2371</u>		$= \frac{var_{rel1}}{var_{rel2}}$
Kritická hodnota F_{crit}	<u>1.7045</u>		- Výpočet v MS. Excel® pomocou=FINV(0.05,39;39) - Vyhľadajte údaje v štatistickej tabuľke „F“

* Rovnice a syntax na výpočet v MS Excel® sú uvedené v kapitole „jednoduchá štatistika“.

Krok 2) Porovnáme vypočítanú hodnotu F s referenčnou hodnotou F_{crit} (tab. 6). Tento krok určuje, či sú rozptyly dvoch skupín štatisticky rovnaké. Ak platí $F > F_{crit}$, medzi rozptylmi existuje štatisticky významný rozdiel. Ak platí $F < F_{crit}$, medzi rozptylmi neexistuje štatisticky významný rozdiel a stĺpce sa považujú za podobné. Pre objektívne parametre merané dynamometrom je vhodná hladina významnosti $\alpha = 0,05$. Kritickú hodnotu F_{crit} možno získať z tabuľky F-rozdelenia alebo vypočítať v programe MS Excel® pomocou vzorca =FINV(0.05, 39, 39), kde obe hodnoty 39 predstavujú počet stupňov voľnosti pre každú skupinu. Ak je F nižšie ako F_{crit} , pokračujeme T-testom. Hodnotu T_{crit} vypočítame v programe MS Excel® pomocou vzorca =TINV(0.05, 78), kde 78 predstavuje celkový počet stupňov voľnosti, teda súčet stupňov voľnosti oboch stĺpcov. Výsledkom je $T_{crit} = 1,9908$.

Krok 3) Rovnica na výpočet hodnoty T je pomerne zložitá, ale všetky premenné už poznáme. Preto sa jej netreba obávať, pomocou kalkulačky a papiera rýchlo získame výsledok. Člen $|\bar{x}_2 - \bar{x}_1|$ označuje kladný rozdiel medzi priermi, ktorý dosadíme do vzorca (R. 7).

$$T = \frac{|\bar{x}_2 - \bar{x}_1|}{\sqrt{(n_1 \cdot var_{rel1}^2 + n_2 \cdot var_{rel2}^2)}} \cdot \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}{n_1 + n_2}} = 4.9366 \quad (R. 7)$$

Kde x_1 a x_2 predstavujú priemery jednotlivých stĺpcov, n_1 a n_2 počet pacientov v každej skupine a var_1 a var_2 odhady variability (tab. 5). Pre naše údaje je $T = 4,9366$, čo je významne vyššia hodnota ako T_{crit} , a hypotéza H_0 je potvrdená, čo naznačuje významné zlepšenie svalovej sily ruky pri $\alpha = 0,05$. Ak je zmena nevýznamná napriek pozorovateľnému zlepšeniu, sledovaný parameter alebo použitá hodnota α môžu byť nevhodné. Zvýšenie hodnoty α znižuje presnosť a zvyšuje riziko chýb, preto sa odporúča neprekročiť hodnotu $\alpha = 0,30$. Nižšia hodnota α znižuje mieru chybovosti a zvyšuje spoľahlivosť.

3.4.2.2 PÁROVÝ T-TEST: POROVNANIE ZHODY VÝSLEDKOV DVOCH METODÍK

V pooperačnej starostlivosti po resekcií intramedulárneho osteosarkómu alebo počas hojenia zlomenín je nevyhnutné monitorovať koncentrácie vápnika (Ca^{2+}) alebo kalcitonínu v krvi pacienta³¹. Ak nie sú pre každého pacienta dostupné párové údaje, musí byť daný pacient z analýzy vylúčený. Táto požiadavka na párové údaje predstavuje zásadný rozdiel oproti dvojvýberovému nepárovému T-testu. Prvá možnosť, ktorú sme v tejto práci neskúmali vzhľadom na jej podobnosť s už riešenými štúdiami, zahŕňa monitorovanie koncentrácie vápnika na začiatku a na konci pooperačnej liečby s cieľom získať párové hodnoty pre každého pacienta. Druhá možnosť, ktorú sme v tejto štúdii hodnotili, spočíva v porovnaní dvoch metód merania koncentrácie Ca^{2+} v krvi v rámci skupiny pacientov. V oboch scenároch musia byť pacienti, ktorí sa nezúčastnili druhej kontroly alebo zomreli, vylúčení z dôvodu chýbajúcich párových hodnôt.

Hypotéza H_0 znela, že obe metódy sú rovnako reprezentatívne. Koncentrácia Ca^{2+} v krvi bola meraná pomocou štandardného krvného testu SMAC (SMAC)³² a pokročilého klinicko-chemického analyzátoru (ACChA)³³. V programe MS Excel® sme vytvorili tri stĺpce: (I) koncentrácie vápnika namerané metódou SMAC, (II) koncentrácie namerané metódou ACChA a (III) absolútne rozdiely medzi oboma metódami pre každého pacienta. Ak výsledkom odčítania bolo záporné číslo, bolo prevedené na kladnú hodnotu. Tento stĺpec kladných rozdielov sa používa na výpočty, ako aj na štatistické parametre potrebné na vyhodnotenie párového -testu (tab. 7). Po výpočte rozdielov všetkých párov, teda riadkov, ich spriemerujeme a získame hodnotu $Dif_{average}$ uvedenú v tabuľke (tab. 7).

Tabuľka 7 – Štatistiky vypočítané z koncentrácií Ca^{2+} v krvi stanovených dvoma rôznymi metódami.

Potrebné veličiny	Ca^{2+} koncentrácia získaná metódou SMAC	Ca^{2+} koncentrácia získaná metódou ACChA	Rovnica
$ Dif $	Stĺpec (I)	Stĺpec (II)	$= hodnota\ v\ (I) - hodnota\ v\ (II) $ Opakovať osobitne pre každý riadok
Priemer kladných rozdielov	$Dif_{average} = 0.01$		$= AVERAGE(select_column\ III)$
Variancia rozdielov var_{dif}	$var_{dif} = 0.0003$		$= VAR.P(select_column\ III)$
SD rozdielov	$SD = 0.0183$		$= STDEV.P(select_column\ III)$
Počet pacientov	$n = 12$		
Stupeň voľnosti	$f_{deg} = 11$		$= počet\ pacientov - 1$
Kritická hodnota T_{crit}	<u>2.2010</u>		• Výpočet v MS. Excel® pomocou =TINV(0.05;11) • Find data in statistic "T" table

Z 26 pacientov s diagnostikovaným intramedulárnym osteosarkómom malo krvné testy vykonané oboma metódami iba 12 pacientov, ktorí boli zahrnutí do výpočtu párového T-testu. Vzhľadom na objektívny charakter meraní koncentrácie vápnika daný presnosťou laboratórnej diagnostiky sme zvolili hladinu

významnosti $\alpha = 0,05$. Posledným krokom je výpočet hodnoty T_{pair} . Párový T-test je založený výlučne na stĺpci výsledných rozdielov (III), čím sa výpočet zjednodušuje (R. 8).

$$T_{pair} = \frac{Dif_{average} \cdot \sqrt{n-1}}{SD} = 1.8166 \quad (R. 8)$$

Dosadením hodnôt deskriptívnej štatistiky z tabuľky 6 sme vypočítali $T_{pair} = 1,8166$. Táto hodnota je nižšia ako $T_{crit} = 2,201$, čo naznačuje, že pri hladine významnosti $\alpha = 0,05$ nebol pozorovaný štatisticky významný rozdiel medzi metódami SMAC a ACChA pri meraní koncentrácie vápnika v krvi u pacientov po resekcii intramedulárneho osteosarkómu.

3.5 PODPORNÉ TERAPEUTICKÉ PRÍSTUPY A ŠTATISTICKÉ ZAHRNUTIE ICH VPLYVOV

Doplňkové podporné terapeutické prístupy vrátane muzikoterapie, hipoterapie, rehabilitácie asistovanej virtuálnou realitou, intervencií založených na všímavosti a vo vybraných podmienkach aj arteterapie alebo terapie asistovanej zvieratami sa čoraz častejšie používajú ako adjuvantné postupy k štandardnej klinickej starostlivosti³⁴⁻³⁸. Nemali by sa interpretovať ako náhrada chirurgickej liečby, farmakoterapie, fyzioterapie alebo zavedených rehabilitačných protokolov. Ich očakávaný prínos súvisí najmä so záťažou symptómami, vnímaním bolesti, úzkosťou, náladou, spoluprácou, dodržiavaním liečby, zapojením do rehabilitácie, kvalitou života a vybranými funkčnými výsledkami. Preto by sa ich účinky mali analyzovať s rovnakou metodologickou dôslednosťou ako ostatné subjektívne, kvázi objektívne a objektívne parametre posudzované v tejto práci.

Muzikoterapia sa štatisticky hodnotí rovnakým všeobecným spôsobom ako intervencie v psychológii, rehabilitácii a klinickej medicíne. Primárnym cieľom je určiť, či sa definovaný výsledok mení po stanovenom počte terapeutických sedení, či už v rámci krátkodobých alebo dlhodobých terapeutických cieľov, a či sa táto zmena líši od zmeny pozorovanej v kontrolnej podmienke. Sledované výsledky môžu zahŕňať očný kontakt, udržanie pozornosti počas definovaného časového intervalu, účasť na skupinovej aktivite, rešpektovanie pravidiel, reguláciu správania, spokojnosť so životom, intenzitu bolesti hodnotenú pomocou vizuálnej analógovej škály, funkčnú nezávislosť hodnotenú Barthelovým indexom, motorickú obnovu hodnotenú podľa Brunnstromových štádií alebo spasticitu hodnotenú modifikovanou Ashworthovou škálou. Muzikoterapia je preto najrelevantnejšia pri hodnotení subjektívnych a kvázi objektívnych výsledkov súvisiacich s vnímaním bolesti, emočnou reguláciou, pozornosťou, sociálnou interakciou, motiváciou a spoluprácou pri rehabilitácii.

Hipoterapia je adjuvantná rehabilitačná metóda založená na terapeutickom využití pohybu koňa. Rytmický viacsmerový pohyb môže podporovať posturálnu kontrolu, stabilitu trupu, rovnováhu, funkcie súvisiace s chôdzou a senzomotorickú integráciu, najmä vo vybraných neurologických a muskuloskeletálnych rehabilitačných podmienkach. Jej účinky by sa mali interpretovať opatrne, pretože dôkazy sú špecifické pre konkrétny stav a závislé od protokolu. Hipoterapia by sa preto mala hodnotiť pomocou jasne definovaných funkčných výsledkov, ako sú skóre rovnováhy, kvalita chôdze, stabilita sedu, kontrola trupu, hrubá motorika a ukazovatele bezpečnosti pacienta.

Rehabilitácia asistovaná virtuálnou realitou umožňuje opakovaný tréning orientovaný na úlohy so štruktúrovanou vizuálnou, auditívnou alebo senzomotorickou spätnou väzbou. Tento prístup môže zlepšiť zapojenie pacienta a podporiť motorické učenie, tréning rovnováhy, rehabilitáciu chôdze, cvičenia hornej končatiny a vybrané aktivity denného života. Jeho účinok možno hodnotiť pomocou objektívnych parametrov, ako sú rozsah pohybu, rýchlosť chôdze alebo čas dokončenia úlohy, spolu s klinickými škálami a pacientom reportovanými výsledkami.

Intervencie založené na všímavosti môžu zlepšiť zvládanie bolesti, úzkosti, depresívnych príznakov, vnímaného stresu, kvality spánku a psychologické adaptácie na chronické ochorenie. Tieto účinky by sa nemali interpretovať ako priamy dôkaz urýchlenej regenerácie tkanív. Výsledky intervencií založených na všímavosti sa najvhodnejšie hodnotia prostredníctvom subjektívnych a kvázi objektívnych parametrov vrátane skóre úzkosti, skóre bolesti, vnímaného stresu, kvality spánku, adherencie k liečbe a kvality života.

Arteterapia a terapia asistovaná zvieratami môžu byť klinicky užitočné u vybraných pacientov so zníženou motiváciou, chronickým stresom, úzkosťou, sociálnym stiahnutím alebo dlhodobými rehabilitačnými potrebami. Ich očakávané účinky zvyčajne súvisia s emočným vyjadrením, sociálnou interakciou,

motiváciou, dodržaním liečby, pozornosťou a vnímanou psychickou pohodou, nie s priamou obnovou štruktúry tkaniva. Ich dôkazová základňa je heterogénnejšia a závislá od hodnoteného výsledku, preto by ich hodnotenie malo zostať obmedzené na jasne definované behaviorálne, psychologické alebo pacientom nahlasované výsledky.

Štatistické zahrnutie podporných terapií by malo vychádzať z typu škály a dizajnu štúdie. Pri kontinuálnych výsledkoch hodnotených u tých istých pacientov pred terapiou a po nej možno použiť párový T-test, ak sú párové rozdiely približne normálne rozdelené. Pre každého pacienta je párový rozdiel definovaný rovnicou 9.

$$d_i = y_{2,i} - y_{1,i} \quad (\text{R. 9})$$

Kde $y_{2,i}$ je hodnota po intervencii a $y_{1,i}$ je hodnota pred intervenciou. Priemerná párová zmena, štandardná odchýlka párových rozdielov a párová t-štatistika sa vypočítajú pomocou rovníc 10-12.

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i \quad (\text{R. 10})$$

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}} \quad (\text{R. 11})$$

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \quad (\text{R. 12})$$

Počet stupňov voľnosti je $n - 1$. Interval spoľahlivosti na úrovni 95 % pre priemernú párovú zmenu je definovaný rovnicou 13.

$$\bar{d} \pm t_{1-\alpha/2, n-1} \frac{s_d}{\sqrt{n}} \quad (\text{R. 13})$$

Zmena by mala byť uvádzaná spolu s priemerným rozdielom, intervalom spoľahlivosti a štandardizovanou veľkosťou efektu. Pri párových meraniach sa veľkosť efektu založená na skóre rozdielov vypočíta podľa rovnice 14.

$$d_z = \frac{\bar{d}}{s_d} \quad (\text{R. 14})$$

Pri malých skupinách pacientov možno párovú veľkosť efektu korigovať na skreslenie spôsobené malou veľkosťou výberu podľa rovníc 15-16.

$$g_z = J d_z \quad (\text{R. 15})$$

$$J \approx 1 - \frac{3}{4(n-1)-1} \quad (\text{R. 16})$$

Kde d_z je štandardizovaná párová veľkosť efektu a g_z je párová veľkosť efektu korigovaná na malú veľkosť výberu. Ak sa terapeutická skupina porovnáva s kontrolnou skupinou v dizajne s východiskovým a následným meraním, je vhodnejšie použiť ANCOVA alebo model so zmiešanými efektmi než izolované nezávislé t-testy, pretože tieto modely priamo testujú účinok intervencie pri súčasnom zohľadnení východiskových hodnôt a opakovaných meraní (R. 17-18).

$$Y_{2,i} = \beta_0 + \beta_1 G_i + \beta_2 Y_{1,i} + \varepsilon_i \quad (\text{R. 17})$$

$$Y_{i,t} = \beta_0 + \beta_1 G_i + \beta_2 T_t + \beta_3 (G_i \times T_t) + u_i + \varepsilon_{i,t} \quad (\text{R. 18})$$

Pri klinických škálach s poradovým charakterom alebo behaviorálne kódovaných kategóriách sú vo všeobecnosti vhodnejšie neparametrické alebo štatistické modely pre údaje s poradovým charakterom. Ak sa

metóda hodnotovej matice používa ako numerická transformácia špecifická pre danú štúdiu, jej skórovacie pravidlá by mali byť explicitne odôvodnené a tam, kde je to možné, overené pomocou analýzy citlivosti voči ordinálnemu prístupu.

3.6 VIACROZMERNÁ ANALÝZA AKO NÁSTROJ UMELEJ INTELIGENCIE

Viacrozmerná analýza je výkonným nástrojom umelej inteligencie, ktorý umožňuje súčasne skúmať viacero premenných s cieľom identifikovať vzorce, vzťahy a trendy³⁹⁻⁴¹. Techniky, ako sú analýza hlavných komponentov (PCA), faktorová analýza, klastrová analýza a diskriminačná analýza, sa v umelej inteligencii široko využívajú na zlepšenie interpretácie údajov a rozhodovania. Tieto metódy pomáhajú znižovať dimenzionalitu údajov, odhaľovať latentné štruktúry, zoskupovať podobné pozorovania a klasifikovať údaje do vopred definovaných kategórií, čím zlepšujú presnosť a efektivitu modelov umelej inteligencie. Každá technika má však svoje obmedzenia, ako sú predpoklady týkajúce sa normality údajov, citlivosť na odľahlé hodnoty a výpočtová náročnosť, ktoré je potrebné dôsledne zohľadniť na zabezpečenie spoľahlivých výsledkov.

3.6.1 ANALÝZA HLAVNÝCH KOMPONENTOV (PCA)

Analýza hlavných komponentov (PCA) znižuje dimenzionalitu rozsiahlych dátových súborov transformáciou korelovaných premenných na menší súbor nekorelovaných hlavných komponentov⁴². Tieto komponenty zachytávajú maximálny rozptyl, čím zjednodušujú komplexnosť údajov pri zachovaní podstatných informácií. PCA je užitočná v medicínskom zobrazovaní, napríklad pri MRI a CT vyšetreniach, kde znižuje komplexnosť údajov, a v genomike pri analýze génovej exprese, identifikácii kľúčových génov ovplyvňujúcich génovú aktivitu a podpore porozumenia genetickým poruchám. PCA má však aj obmedzenia vrátane možnej straty interpretovateľnosti, predpokladu, že hlavné komponenty s najvyšším rozptylom sú najdôležitejšie, citlivosti na mierku údajov a predpokladu lineárnych vzťahov medzi premennými, čo nemusí byť vhodné pri nelineárnych dátových štruktúrach (R. 19).

$$PC_j = \sum_{i=1}^p w_{ij} \cdot X_i \quad (\text{R. 19})$$

Kde PC_j označuje j-ty hlavný komponent vytvorený ako lineárna kombinácia pôvodných pozorovaných premenných X_i . Index i nadobúda hodnoty od 1 po p , kde p predstavuje celkový počet štandardizovaných klinických alebo biologických premenných zahrnutých do analýzy. Koeficienty w_{ij} sú komponentové záťažové odvođené ako prvky vlastných vektorov kovariančnej alebo korelačnej matice Σ a kvantifikujú relatívny príspevok každej premennej X_i k tvorbe PC_j . Komponenty sa extrahujú v zostupnom poradí podľa vysvetleného rozptylu, pričom PC_1 zachytáva maximálny možný rozptyl, PC_2 zachytáva maximálny zostávajúci rozptyl pri podmienke ortogonalít k PC_1 a tak ďalej. Tieto váhy sa odvodzujú riešením vlastného rozkladu $\sum w_j = \lambda_j \cdot w_j$, kde λ_j predstavuje vlastnú hodnotu prislúchajúcu j-temu komponentu a zodpovedá množstvu celkového vysvetleného rozptylu.

3.6.2 FAKTOROVÁ ANALÝZA

Faktorová analýza identifikuje vzťahy medzi premennými tým, že ich modeluje ako lineárne kombinácie latentných faktorov, čím odhaľuje skryté štruktúry vysvetľujúce pozorované korelácie (R. 20). V klinickej psychológii môže odhaliť základné dimenzie duševného zdravia identifikáciou faktorov prispievajúcich k poruchám na základe symptómov a položiek dotazníkov, napríklad pri určovaní primárnych faktorov ovplyvňujúcich úzkosť a depresiu s cieľom cielej intervencie⁴³. Faktorová analýza však vyžaduje veľké veľkosti výberu na získanie spoľahlivých výsledkov, predpokladá lineárne vzťahy a presnú reprezentáciu modelu a môže poskytovať odlišné riešenia v závislosti od použitých metód extrakcie a rotácie. Okrem

toho predpoklad nekorelovaných faktorov nemusí byť splnený a interpretácia faktorov môže byť subjektívna a môže sa medzi výskumníkmi líšiť.

$$X_i = \sum_{j=1}^m \lambda_{ij} \cdot F_j + \varepsilon_i \quad (\text{R. 20})$$

Kde m označuje počet základných latentných konštruktov ($m < p$), F_j predstavuje j -ty nepozorovaný spoločný faktor, pri ktorom sa predpokladá štandardné normálne rozdelenie. Parametre λ_{ij} označujú faktorné záťaž, ktoré vyjadrujú silu a smer vzťahu medzi pozorovanou premennou i a latentným faktorom j . Člen ε_i zachytáva jedinečný rozptyl pripísateľný chybe merania alebo vplyvom špecifickým pre danú premennú, ktoré nie sú spoločné s ostatnými ukazovateľmi. Zvyčajne sa predpokladá, že faktory sledujú viac-rozmerné normálne rozdelenie s nulovou strednou hodnotou a jednotkovým rozptylom, pričom odhad sa bežne vykonáva metódou maximálnej vierohodnosti alebo metódou hlavných osí.

3.6.3 KLASTROVÁ ANALÝZA

Klastrová analýza zoskupuje pozorovania na základe podobnosti bez vopred definovaných kategórií. Medzi bežné metódy patrí klastrovanie k -priemerov⁴⁴, hierarchické klastrovanie⁴⁵ a DBSCAN⁴⁶. Táto technika je cenná pri segmentácii pacientov, napríklad pri zoskupovaní pacientov s podobnými symptómami alebo podobnou odpoveďou na liečbu s cieľom uľahčiť individualizované medicínske intervencie. Môže napríklad segmentovať onkologických pacientov na základe ich odpovede na chemoterapiu, čo umožňuje vytvárať personalizované liečebné plány⁴⁶.

Klastrovanie k -priemerov rozdeľuje údaje do k klastrov a každé pozorovanie priradzuje k najbližšiemu priemeru⁴⁴ (R. 21). Je efektívne pri veľkých dátových súboroch, vyžaduje však určenie počtu klastrov a je citlivé na počiatočné centroidy. Hierarchické klastrovanie vytvára dendrogram, ktorý reprezentuje vnorené zoskupenia bez potreby vopred stanoviť počet klastrov, môže však byť výpočtovo náročné. DBSCAN zoskupuje husto usporiadané body a izolované body označuje ako odľahlé hodnoty. Umožňuje identifikovať klastre ľubovoľného tvaru bez potreby špecifikovať ich počet, avšak môže mať problémy pri premenlivej hustote údajov a pri vysokodimenziálnych dátach.

Klastrová analýza má viaceré obmedzenia, napríklad citlivosť na metriky vzdialenosti a počet klastrov, rozdielne výsledky v závislosti od použitého algoritmu a náročnosť interpretácie a validácie klastrov, najmä pri vysokodimenziálnych údajoch. Môže byť výpočtovo náročná a vyžaduje vhodné predspracovanie údajov a dôsledný výber parametrov. Interpretovateľnosť klastrov môže byť navyše subjektívna a metóda nemusí poskytovať dobré výsledky, ak skutočná štruktúra údajov nie je vhodne reprezentovaná zvoleným modelom.

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (\text{R. 21})$$

Účelová funkcia J predstavuje celkový súčet štvorcov euklidovských vzdialeností v rámci klastrov, ktorý algoritmus iteratívne minimalizuje. Parameter K určuje vopred stanovený počet podskupín pacientov. Označenie C_k predstavuje súbor pozorovaní priradených ku k -temu klastru počas každej iterácie. Vektor μ_k zodpovedá centroidu, ktorý sa vypočíta ako aritmetický priemer všetkých pozorovaní aktuálne priradených ku klastru k . Dvojitý zvislý čiary $\|\cdot\|$ označujú euklidovskú normu, ktorá meria geometrickú vzdialenosť medzi jednotlivým pozorovaním x_i a jemu zodpovedajúcim stredom klastra.

3.6.4 DISKRIMINAČNÁ ANALÝZA

Diskriminačná analýza klasifikuje pozorovania do vopred definovaných kategórií, pričom sa bežne používa lineárna diskriminačná analýza^{47,48} (LDA) a kvadratická diskriminačná analýza (QDA)^{49,50}. V onkológii môže klasifikovať nádory ako benígne alebo malígne na základe viacerých diagnostických znakov, čím napomáha včasnej a presnej diagnostike. LDA napríklad pomáha patológom kategorizovať nádorové vzorky na základe údajov zo sekvenovania RNA, čím zlepšuje diagnostiku nádorových ochorení a usmerňuje rozhodovanie o liečbe. Diskriminačná analýza však predpokladá normalitu prediktorov a v prípade LDA aj rovnosť kovariančných matíc a je citlivá na odľahlé hodnoty. Pri nelineárnych vzťahoch alebo

malých veľkostiach výberu môže dosahovať slabšie výsledky a môže byť menej účinná, ak triedy nie sú dostatočne dobre oddelené, čo môže viesť k potenciálnej nesprávnej klasifikácii (R. 22-23).

$$y_i = \sum_{j=1}^p w_j \cdot x_{ij} \quad (\text{R. 22})$$

$$w = S_w^{-1} \cdot (\bar{x}_1 - \bar{x}_2) \quad (\text{R. 23})$$

Kde y_i označuje diskriminačné skóre pre pozorovanie i , vypočítané ako lineárna kombinácia pôvodných pozorovaných premenných x_{ij} . Index j nadobúda hodnoty od 1 po p , pričom p predstavuje celkový počet prediktorových premenných zahrnutých do analýzy. Koefficienty w_j sú diskriminačné váhy. V rovnici 13 je w vektor diskriminačných váh ($w_1, w_2, \dots, w_p$). Matica S_w je spoločná vnútrotriedna kovariančná matica, ktorá vyjadruje, do akej miery sa prediktorové premenné menia v rámci jednotlivých skupín. Člen S_w^{-1} predstavuje jej inverznú maticu. Vektory $\bar{x}_1$ a $\bar{x}_2$ sú vektory priemerov skupiny 1 a skupiny 2. Rozdiel $(\bar{x}_1 - \bar{x}_2)$ predstavuje oddelenie medzi skupinami pre každý prediktor. Násobením S_w^{-1} výrazom $(\bar{x}_1 - \bar{x}_2)$ sa získajú váhy, ktoré maximalizujú oddelenie medzi skupinami pri súčasnom zohľadnení korelácií medzi prediktormi. Pozorovania sa klasifikujú do skupiny, ktorá poskytuje najvyššie diskriminačné skóre.

3.6.5 VIACROZMERNÁ REGRESIA

Viacrozmerná regresia rozširuje viacnásobnú regresiu tým, že modeluje viacero závislých premenných súčasne^{51,52}. Táto metóda ukazuje, ako nezávislé premenné ovplyvňujú viaceré výsledky. V kardiológii môže viacrozmerná regresia analyzovať vplyv faktorov životného štýlu na viaceré kardiovaskulárne výsledky, ako sú krvný tlak, hladiny cholesterolu a srdcová frekvencia. Táto komplexná analýza pomáha pochopiť, ako rôzne faktory prispievajú k zdraviu srdca, a usmerňuje preventívne aj liečebné stratégie. Viacrozmerná regresia má však aj obmedzenia, ako sú predpoklad viacrozmernej normality a linearity, potreba veľkých veľkostí výberu na zabezpečenie spoľahlivých odhadov a možná multikolinearita medzi nezávislými premennými, ktorá môže komplikovať interpretáciu výsledkov. Okrem toho sa komplexnosť modelu zvyšuje s počtom závislých premenných, čo môže pri nedostatočnej kontrole viesť k nadmernému prispôbeniu modelu údajom. Interpretácia koeficientov môže byť navyše náročná vzhľadom na vzájomné závislosti medzi závislými premennými (R. 24).

$$Y_j = \beta_j + \sum_{k=1}^p \beta_{jk} \cdot X_k + \varepsilon_j \quad (\text{R. 24})$$

Kde Y_j označuje j -tu závislú premennú, pričom q predstavuje celkový počet výsledkov modelovaných súčasne. Index k nadobúda hodnoty od 1 po p , pričom p predstavuje počet nezávislých premenných. Parameter β_j je absolútny člen pre výsledok j , zatiaľ čo β_{jk} predstavuje regresný koeficient kvantifikujúci zmenu Y_j spojenú so zmenou prediktora X_k o jednu jednotku pri zachovaní ostatných prediktorov konštantných. Člen ε_j zachytáva reziduálnu chybu pre výsledok j . Na rozdiel od jednorozmernej regresie tento prístup zohľadňuje korelácie medzi q závislými premennými, poskytuje efektívnejšie odhady a umožňuje spoločné testovanie hypotéz naprieč viacerými výsledkami.

3.6.6 KANONICKÁ KORELAČNÁ ANALÝZA (CCA)

Kanonická korelačná analýza (CCA) skúma vzťahy medzi dvoma súbormi premenných a identifikuje lineárne kombinácie, ktoré sú maximálne korelované⁵³. V genomike⁵⁴ CCA skúma súvislosti medzi genetickými markermi a fenotypovými znakmi, čím odhaľuje významné genetické vplyvy na fyzické charakteristiky a mechanizmy ochorení. CCA môže napríklad identifikovať genetické markery najsilnejšie asociované so špecifickými ochoreniami, čo napomáha objavovaniu terapeutických cieľov. CCA však predpokladá lineárne vzťahy, je citlivá na odľahlé hodnoty, vyžaduje veľké veľkosti výberu na získanie

spoľahlivých výsledkov a môže byť výpočtovo náročná a zložitá na interpretáciu. Na získanie významovo hodnotných výsledkov je nevyhnutné vhodné predspracovanie a normalizácia údajov (R. 25-26).

$$\begin{cases} U = \sum_{i=1}^p a_i \cdot X_i \\ V = \sum_{j=1}^q b_j \cdot Y_j \end{cases} \quad (\text{R. 25})$$

$$r_c = \text{korelácia}(U, V) = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U) \cdot \text{Var}(V)}} \quad (\text{R. 26})$$

Kde U a V označujú prvý pár kanonických premenných, teda lineárnych kombinácií. Súbor $\{X_1, \dots, X_p\}$ predstavuje prvý blok premenných, zatiaľ čo $\{Y_1, \dots, Y_q\}$ predstavuje druhý blok premenných. Koefficienty a_i a b_j sú kanonické váhy odvodené tak, aby maximalizovali koreláciu r_c medzi U a V . Nasledujúce páry kanonických premenných ($U_2, V_2, \dots$) sa extrahujú pri podmienke, že nie sú korelované s predchádzajúcimi pármí. Maximálny počet kanonických korelácií sa rovná $\min(p, q)$. Štatistická významnosť sa zvyčajne hodnotí pomocou Wilksovej lambda alebo Bartlettovho chí-kvadrát testu.

3.6.7 VIACROZMERNÁ ANALÝZA ROZPTYLU (MANOVA)

Viacrozmerná analýza rozptylu (MANOVA) rozširuje ANOVA tým, že súčasne hodnotí viacero závislých premenných, čo je užitočné v experimentálnych štúdiách^{55,56}. V klinických skúšaníach MANOVA hodnotí vplyv lieku na rôzne zdravotné výsledky, ako sú hladina cukru v krvi, hmotnosť a cholesterol, čím poskytuje komplexný pohľad na účinnosť liečby. MANOVA však predpokladá viacrozmernú normalitu, homogenitu variančno-kovariančných matic a vyžaduje veľké veľkosti súboru. Je citlivá na odľahlé hodnoty a pri nelineárnych vzťahoch môže byť zložitá na interpretáciu. Na získanie spoľahlivých výsledkov je nevyhnutné predspracovanie údajov a zohľadnenie predpokladov modelu (R. 27).

$$Y_i = \mu + \tau_g + \epsilon_i \quad (\text{R. 27})$$

Kde Y_i označuje vektor závislých premenných pre pozorovanie i , pričom q predstavuje celkový počet závislých premenných modelovaných súčasne. Vektor μ predstavuje celkový vektor priemerov naprieč všetkými skupinami. Parameter τ_g predstavuje účinok skupiny g na závislé premenné. Člen ϵ_i je náhodný chybový vektor pre pozorovanie i , pri ktorom sa predpokladá viacrozmerné normálne rozdelenie s nulovou strednou hodnotou a kovariančnou maticou Σ . MANOVA testuje nulovú hypotézu, že všetky skupinové efekty sú nulové, pomocou Wilksovej lambda štatistiky (R. 28).

$$\Lambda = \frac{|E|}{|H+E|} \quad (\text{R. 28})$$

Kde E je matica súčtov štvorcov a krížových súčinov chyby a H je matica súčtov štvorcov a krížových súčinov hypotézy.

3.7 MONITOROVANIE PREŽÍVANIA

3.7.1 KAPLAN-MEIEROVA KRIVKA – MONITOROVANIE PREŽÍVANIA PACIENTOV S INTRAMEDULÁRNÝM OSTEOSARKÓMOM

Kaplanova-Meierova (KM) krivka je široko používaná štatistická metóda v analýze prežívania, vhodná na prácu s cenzorovanými údajmi^{57,58}. K cenzorovaniu (C) dochádza vtedy, keď pacienti odstúpia zo štúdie z iných dôvodov ako úmrtie alebo keď nie je známy čas úmrtia. Cenzorovanie neohrozuje validitu výsledku⁵⁹. Táto metóda odhaduje pravdepodobnosť prežívania v čase a môže porovnávať miery prežívania medzi rôznymi liečebnými postupmi alebo skupinami pacientov s odlišnými komorbiditami. Medzi predpoklady Kaplanovej-Meierovej metódy patrí nezávislé cenzorovanie, pri ktorom musí byť mechanizmus cenzorovania nezávislý od prognózy prežívania, a konzistentné vstupné a výstupné kritériá, ktoré zabezpečujú jasne definované parametre štúdie⁶⁰.

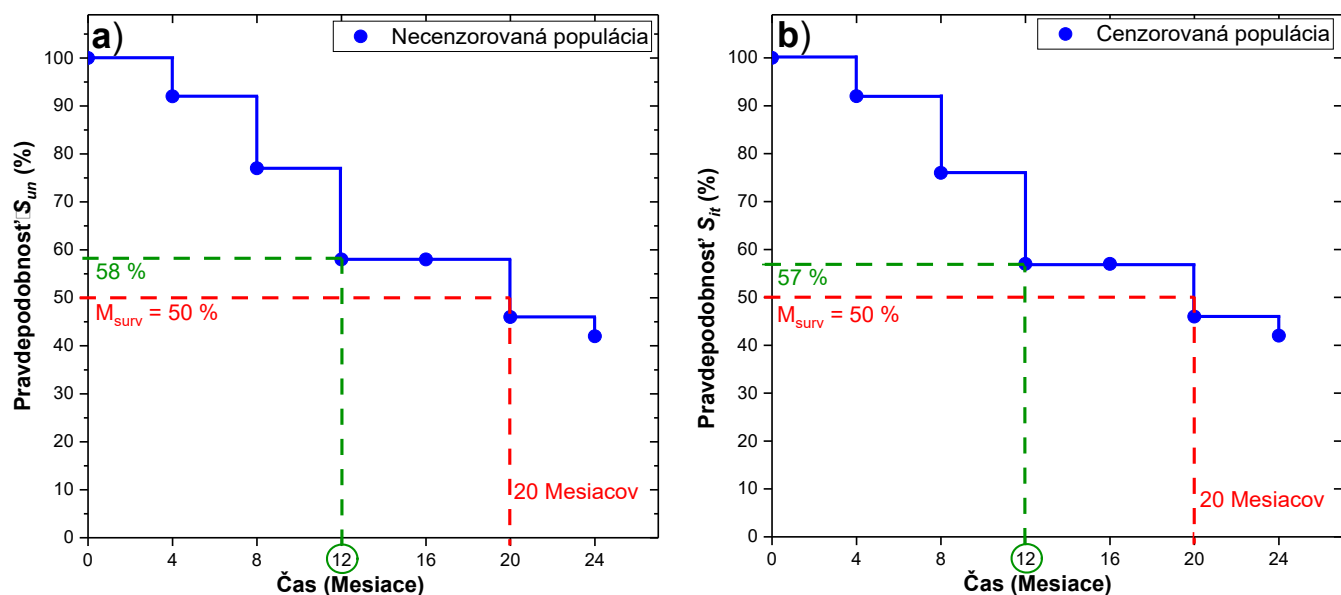
V našej štúdií bola KM krivka použitá na monitorovanie prežívania pacientov po resekcii intramedulárneho osteosarkómu (tab. 8, obr. 6). Z 26 pacientov boli dvaja cenzorovaní. Tabuľka 8 podrobne uvádza údaje o prežívaní vrátane časových intervalov monitorovania (W), počtu úmrtí v každom intervale (E), počtu žijúcich pacientov na začiatku intervalu (n), kumulatívnej miery prežívania pre necenzorovaných pacientov (S_{un}), miery prežívania upravenej pre cenzorovaných pacientov (S_{cen}) a iteratívneho výpočtu miery prežívania pre cenzorované údaje (S_{it}).

Tabuľka 8 – Výpočet Kaplanovej-Meierovej krivky pre pacientov s intramedulárnym osteosarkómom.

NECENZOROVANÉ				CENZOROVANÉ		
W (Mesiac)	E	$n = n_{prev} - E_{prev}$	$S_{un} = \frac{n_{next}}{n_{all}}$	C	$S_{cen} = \frac{n - E}{n}$	$S_{it} = S_{cen} \times S_{it(prev)}$
0 - 4	0	$n_{all} = 26$	$26/26 = 1.00 = 100 \%$	0	1.00	100 %
4 - 8	2	$n = 26 - 0 = 26$	$24/26 = 0.92 = 92 \%$	0	0.92	$\times 1.00 = 0.92 = 92 \%$
8 - 12	4	$n = 26 - 2 = 24$	$20/26 = 0.77 = 77 \%$	1	0.83	$\times 0.92 = 0.76 = 76 \%$
12 - 16	5	$n = 24 - 4 = 20$	$15/26 = 0.58 = 58 \%$	0	0.75	$\times 0.76 = 0.57 = 57 \%$
16 - 20	0	$n = 20 - 5 = 15$	$15/26 = 0.58 = 58 \%$	1	1.00	$\times 0.57 = 0.57 = 57 \%$
20 - 24	3	$n = 15 - 0 = 15$	$12/26 = 0.46 = 46 \%$	0	0.80	$\times 0.57 = 0.46 = 46 \%$
24	1	$n = 15 - 3 = 12$	$11/26 = 0.62 = 42 \%$	0	0.92	$\times 0.46 = 0.42 = 42 \%$

Výpočet zahŕňa určenie počtu rizikových pacientov v každom intervale, podielu prežívajúcich pacientov a následnú iteratívnu úpravu týchto podielov s cieľom zohľadniť cenzorované údaje. Zahrnutie cenzorovaných pacientov do analýzy je kľúčové pre presnú a spoľahlivú interpretáciu mier prežívania. Miera prežívania v každom intervale sa odvodzuje vynásobením miery prežívania z predchádzajúceho intervalu aktuálnou mierou, a to pre necenzorované aj cenzorované údaje. Obrázok 6 zobrazuje Kaplanovu-Meierovu krivku, ktorá znázorňuje pravdepodobnosti prežívania v čase (obr. 6). Krivka ukazuje minimálne rozdiely medzi necenzorovanými a cenzorovanými údajmi, pričom obe skupiny vykazujú 42 % mieru prežívania po 24 mesiacoch. Medián času prežívania (M_{surv}), identifikovaný ako čas, v ktorom pravdepodobnosť prežívania pretína hodnotu 50 %, je v oboch skupinách 20 mesiacov⁵⁹. Táto hodnota je v súlade s existujúcou literatúrou. Mierne rozdiely pozorované v mierach prežívania medzi necenzorovanými a cenzorovanými údajmi zdôrazňujú potrebu tohto iteratívneho prístupu (tab. 8). KM krivka a medián času prežívania poskytujú informácie o prognóze pacienta a účinnosti liečebných intervencií.

Kaplanova-Meierova metóda pracuje so zhodami, teda s viacerými udalosťami v rovnakom časovom bode, pomocou štatistických metód, ako sú exaktná metóda, Breslowova metóda a Efronova metóda. Exaktná metóda predpokladá, že udalosti nastávajú v nekonečne malom časovom intervale v náhodnom poradí, čo ju robí výpočtovo náročnou vzhľadom na enumeráciu všetkých možných poradí udalostí. Breslowova metóda, alebo generalizovaná Wilcoxonova metóda, zjednodušuje výpočty rovnomerným rozdelením viazaných udalostí v rámci rizikového súboru a spriemerovaním pravdepodobností prežívania. Efronova metóda vyvažuje výpočtovú efektívnosť a presnosť tým, že pri každej viazanej udalosti inkrementálne upravuje rizikový súbor a frakčne priradzuje udalosti s cieľom získať presnejší odhad. Intervaly spoľahlivosti pre odhady prežívania vypočítané pomocou Greenwoodovho vzorca poskytujú informáciu o presnosti (R. 5). Pravostranné cenzorovanie, ktoré je najbežnejšou formou, nastáva vtedy, keď sa štúdia skončí pred výskytom udalosti u niektorých subjektov alebo keď sú subjekty stratené zo sledovania. Ľavostranné cenzorovanie nastáva vtedy, keď udalosť nastane pred začiatkom obdobia pozorovania, a intervalové cenzorovanie nastáva vtedy, keď je známe, že udalosť nastala v určitom časovom intervale. Hoci je Kaplanova-Meierova metóda výborná na deskriptívne účely, na viacrozmernú analýzu prežívania sa často používa Coxov model proporcionálnych rizík, ktorý umožňuje zahrnúť kovariáty a pochopiť ich vplyv na prežívanie⁶¹. Okrem toho sa na porovnanie rozdelení prežívania medzi skupinami bežne používa log-rank test, ktorý hodnotí významnosť rozdielov medzi KM krivkami.



Obrázok 6 – Kaplanove-Meierove krivky pre rovnakú populáciu pacientov liečených pre intramedulárny osteosarkóm, kde (a) predstavuje necenzorovanú skupinu pacientov a (b) predstavuje tú istú skupinu pacientov po cenzorovaní. Zelený priesečník znázorňuje mieru prežívania po 12 mesiacoch. Medián prežívania (červená) zodpovedá 50 % pravdepodobnosti a dosahuje hodnotu 20 mesiacov u cenzorovaných aj necenzorovaných pacientov.

3.7.2 LOG RANK TEST – POROVNANIE ČASOVO ZÁVISLÝCH JAVOV ALEBO MIER PREŽÍVANIA VIACERÝCH SKUPÍN PACIENTOV

Na štatistické vyhodnotenie a porovnanie mier prežívania dvoch populácií sa odporúča použiť log-rank test (LR), ktorý slúži na určenie rozdielu alebo podobnosti distribučných funkcií prežívania v čase⁶². Tento test dopĺňa KM metódu tým, že štatisticky potvrdzuje alebo zamieta podobnosť distribúcií prežívania. LR test možno použiť aj na porovnanie iných časovo závislých udalostí, ako sú trvanie liečby, obdobia rehabilitácie alebo časovo závislý efekt životnosti materiálov používaných v zdravotníckych pomôckach. Bežnou aplikáciou je porovnanie distribúcií prežívania medzi skupinami pacientov s určitými stavmi alebo liečbami a bez nich. LR test predpokladá, že pomery rizík medzi skupinami sú v čase konštantné, teda že je splnený predpoklad proporcionálnych rizík. Porušenie tohto predpokladu môže ovplyvniť validitu výsledkov testu. LR zároveň predpokladá, že cenzorovanie je náhodné a nesúvisí s pravdepodobnosťou výskytu udalosti. Nenáhodné cenzorovanie môže viesť k skresleniu⁶³.

V našej štúdii sa usilujeme overiť hypotézu H_0 „Cenzorovanie menej ako 10 % pacientov nevedie k významnému rozdielu v profiloch času prežívania.“ Na testovanie tejto hypotézy porovnávame dve skupiny pacientov po resekcií osteosarkómu, skupinu 1 bez cenzorovania a skupinu 2 s cenzorovaním. Najskôr vytvoríme tabuľku na sledovanie počtu prežívajúcich pacientov v každej skupine počas špecifikovaných časových rámcov (tab. 9).

Tabuľka 9 – Prípravná tabuľka pre log-rank štatistiku znázorňujúca zmeny počtu pacientov počas sledovaného obdobia.

W (Mesiace)	Skupina 1 - NECENZOROVANÉ			Skupina 2 - CENZOROVANÉ		
	E_1	C_1	$n_1 = n_{prev} - E_{prev}$	E_2	C_2	$n_2 = n_{prev} - E_{prev}$
0 - 4	0	0	$n_{all} = 26$	0	0	$n_{all} = 26$
4 - 8	2	0	$n_1 = 26 - 0 - 0 = 26$	2	0	$n_2 = 26 - 0 - 0 = 26$
8 - 12	4	0	$n_1 = 26 - 2 - 0 = 24$	4	1	$n_2 = 26 - 2 - 0 = 24$
12 - 16	5	0	$n_1 = 24 - 4 - 0 = 20$	5	0	$n_2 = 24 - 4 - 1 = 19$
16 - 20	0	0	$n_1 = 20 - 5 - 0 = 15$	0	1	$n_2 = 19 - 5 - 0 = 14$
20 - 24	3	0	$n_1 = 15 - 0 - 0 = 15$	3	0	$n_2 = 14 - 0 - 1 = 13$
24	1	0	$n_1 = 15 - 3 - 0 = 12$	1	0	$n_2 = 13 - 3 - 0 = 10$

Prvý stĺpec, W , uvádza časové rámce v mesiacoch, začínajúc intervalom 0-4 mesiace a pokračujúc v štvormesačných intervaloch. Počet riadkov zodpovedá týmto intervalom počas 24-mesačného obdobia. Druhý stĺpec, E_1 predstavuje počet pacientov, ktorí zomreli v každom intervale. Tretí stĺpec, C_1 , uvádza cenzorovaných pacientov. Posledný stĺpec, n_1 , označuje celkový počet žijúcich pacientov na začiatku každého intervalu vrátane všetkých pacientov prítomných v prvý deň, bez odstránenia tých, ktorí zomreli počas daného obdobia. Pacienti, ktorí zomreli v každom intervale, sú odstránení z rizikového súboru v nasledujúcich intervaloch. Tabuľka 9 znázorňuje tieto výpočty. Podobne stĺpce E_2 , C_2 a n_2 pre skupinu 2 zohľadňujú zosnulých aj cenzorovaných pacientov. Hodnoty n_1 a n_2 sa určujú ako pri Kaplanovej-Meierovej metóde, avšak s úpravou spočívajúcou v odpočítaní cenzorovaných pacientov z predchádzajúceho časového rámca.

Následne zostavíme tabuľku na priamy výpočet log-rank štatistiky (tab. 10) s použitím údajov z prípravnej tabuľky 9. Tabuľka 10 obsahuje Exp_1 a Exp_2 , teda očakávaný počet úmrtí pre každú skupinu, vypočítaný pomocou rovníc (R. 29a, b).

$$Exp_1 = \left(\frac{n_1}{n_1 + n_2} \right) \cdot (E_1 + E_2) \quad \text{or} \quad Exp_2 = \left(\frac{n_2}{n_1 + n_2} \right) \cdot (E_1 + E_2) \quad (\text{R. 29a, b})$$

Dôležitou hodnotou je log-rank variancia (var_{LR}), definovaná rovnicou (R. 30). Log-rank variancia určuje mieru variability a závisí od odchýlky od populačného priemeru. Je nevyhnutné vypočítať varianciu osobitne pre každý riadok tabuľky, pretože nejde o lineárny jav.

$$var_{LR} = \frac{n_i \cdot n_2 \cdot (E_1 + E_2) \cdot (n_1 + n_2 - E_1 - E_2)}{(n_1 + n_2) \cdot (n_1 + n_2) \cdot (n_1 + n_2 - 1)} \quad (\text{R. 30})$$

Rozdiely medzi pozorovaným a očakávaným počtom úmrtí pre skupinu 1 a skupinu 2 sa vypočítajú ako $Dif_1 = E_1 - Exp_1$ a $Dif_2 = E_2 - Exp_2$ (tab. 10). Nie je potrebné počítať obe hodnoty, pretože pre ten istý riadok poskytujú rovnakú číselnú hodnotu, líšia sa iba znamienkom $\pm$.

Tabuľka 10 – Vypočítaný očakávaný počet úmrtí (Exp_1 , Exp_2), hodnoty log-rank variancie (var_{LR}) a rozdiely (Dif_1 , Dif_2) pre všetky sledované časové rámce W .

W (Me- siac)	Očakávané úmrtia pre každú skupinu		Variancia medzi skupinami var_{LR} (zhodné pre obe skupiny)	Pozorované-očakávané úmrtia	
	Exp_1 (Pre skupinu 1)	Exp_2 (Pre skupinu 2)		Dif_1 (Pre skupinu 1)	Dif_2 (Pre skupinu 2)
0 - 4	0	0	0	0	0
4 - 8	2	2	0.9412	0	0
8 - 12	4	4	1.7021	0	0
12 - 16	5.1282	4.8718	1.9066	- 0.1282	0.1282
16 - 20	0	0	0	0	0
20 - 24	3.2142	2.7857	1.2160	- 0.2143	0.2143
24	1.0909	0.9090	0.4723	- 0.0909	0.0909

Po výpočte jednotlivých hodnôt pre každý časový rámec ich zosumarizujeme, aby sme získali celkovú varianciu var_{LR_SUM} a celkové rozdiely Dif_{1_SUM} alebo Dif_{2_SUM} , ako je uvedené v tabuľke 11.

Tabuľka 11 – Vypočítané premenné log-rank testu na porovnanie distribučných funkcií prežívania cenzorovaných a necenzorovaných pacientov po resekcii intramedulárneho osteosarkómu.

Popis	Rovnica	Výsledok
Súčet vypočítaných variácií var_{LR}	$var_{LR_SUM} = \sum_{i=1}^i var_{LR}$	= 6.2382
Súčet vypočítaných rozdielov Dif_1 pre prvú skupinu pacientov	$Dif_{1_SUM} = \sum_{i=1}^i Dif_1$	= - 0.4334
Súčet vypočítaných rozdielov Dif_2 pre druhú skupinu pacientov	$Dif_{2_SUM} = \sum_{i=1}^i Dif_2$	= 0.4334

Log-rank štatistika sa následne vypočíta pomocou rovnice (rov. 31).

$$Log Rank = \frac{(Dif_{1_SUM})^2}{var_{LR_SUM}} = \frac{0.1878}{6.2382} = 0.0301 \quad (R. 31)$$

Táto log-rank hodnota sa porovnáva s kritickou hodnotou chí-kvadrát rozdelenia (χ^2_{crit}) na určenie štatistickej významnosti⁶⁴. Pri tomto výpočte sa stupne voľnosti (f_{deg}) vypočítajú pomocou rovnice (R. 32) a známeho počtu riadkov (*ROWS*) a stĺpcov (*COL*).

$$f_{deg} = (COL - 1) \cdot (ROWS - 1) = (2 - 1) \cdot (7 - 1) = 6 \quad (R. 32)$$

Pri použití hladiny významnosti $\alpha = 0,05$, štatistických tabuliek pre chí-kvadrát rozdelenie alebo MS Excel® zistíme kritickú hodnotu $\chi^2_{crit} = 12.59$. Keďže $\chi^2_{crit} > Log Rank$, dospievame k záveru, že distribúcie prežívania oboch skupín sa významne nelíšia, čo naznačuje, že zahrnutie cenzorovaných pacientov významne neovplyvňuje distribúciu prežívania. Log-rank test potvrdzuje H_0 a ukazuje, že účinok cenzorovania je v tomto kontexte zanedbateľný. Pri väčších súboroch alebo zvýšenom počte udalostí sa môže štatistická sila zmeniť, čo môže potenciálne odhaliť významné rozdiely.

Alternatívne testy, ako je Breslowov test, teda generalizovaný Wilcoxonov test, sú citlivejšie na skoré rozdiely, zatiaľ čo Taroneho-Wareov test vyvažuje citlivosť naprieč časovými rámcami. Pri údajoch o prežívaní je potrebné pracovať so zhodami pomocou úprav, ako sú Breslowova a Peto-Petova modifikácia. Sila LR testu závisí skôr od počtu udalostí než od celkového počtu účastníkov, preto by štúdie mali mať dostatočnú štatistickú silu zabezpečenú dostatočným počtom udalostí. Hoci MS Excel® dokáže vypočítať kritické hodnoty chí-kvadrát rozdelenia, špecializovaný štatistický softvér, napríklad R, SAS alebo SPSS, ponúka robustnejšie nástroje na vykonanie a interpretáciu LR testu, najmä v zložitých scenároch.

4 ZÁVERY

Predstavili sme aplikáciu viacerých štatistických metód používaných v biomedicínskom výskume vrátane ich manuálneho výpočtu a interpretácie. Deskriptívna štatistika hodnotila časovo závislý profil bolesti u pacientov po fraktúre zápästia a po resekcii osteosarkómu. Pomocou dynamometra sme stanovili priemer, smerodajnú odchýlku a interval spoľahlivosti svalovej sily ruky po rehabilitácii. Usmernenia pre pokročilé štatistické metódy boli opätovne predstavené na príkladoch reálnych skupín pacientov a hypotéz, pričom hladiny významnosti boli nastavené pre objektívne ($\alpha = 0,05$), kvázi objektívne ($\alpha = 0,10$) a subjektívne parametre ($\alpha = 0,20$). Dvojvýberový nepárový T-test sme použili na potvrdenie hypotézy: „Mení sa svalová sila ruky počas rehabilitácie pacientov po fraktúre zápästia?“ Korelačná analýza opísala pokles sérových zápalových markerov po resekcii osteosarkómu, ktorý bol u žien o $32 \pm 2\%$ pomalší. Nárast

chuti do jedla po resekcii bol u mužov o 27 ± 2 % rýchlejší. Neidentifikovali sme významnú koreláciu medzi vekom a hmotnosťou. Počas rehabilitácie osteosarkómu boli pozorované nelineárne účinky magnetických polí na rast osteoblastov. Pomocou párového T-testu sme potvrdili, že dve metódy určovania sérových koncentrácií Ca^{2+} u pacientov po osteosarkóme boli rovnako reprezentatívne. F-test nepreukázal významný rozdiel vo varianciách medzi mužmi a ženami vo výsledkoch zotavovania počas remisie osteosarkómu. S cieľom zlepšiť spoliehanie sa na „Black Box“ softvér a vysvetliť analýzu prežívania sme manuálne vypočítali Kaplanove-Meierove krivky prežívania a log-rank test. Potvrdili sme, že cenzorovanie menej ako 10 % pacientov významne nemení profily prežívania. Log-rank test možno použiť na porovnanie rôznych časovo závislých charakteristík medzi skupinami pacientov. Pri budúcich štatistických hodnoteniach by sa mali zväžiť pokročilé modely fuzzy logiky. Tieto modely implementované v umelej inteligencii môžu zahŕňať rôznorodé kritériá na konzistentné hodnotenie parametrov.

5 ZUSAMMENFASSUNG

Ausgangspunkt: Moderne statistische Modelle sind für eine methodisch fundierte Bewertung von Veränderungen des Gesundheitszustands von Patientinnen und Patienten sowie für eine präzise Prognosestellung unerlässlich. Im Zeitraum von 2003 bis 2025 wurde in der wissenschaftlichen Gemeinschaft eine Zunahme fehlerhafter statistischer Anwendungen beobachtet. Die Abhängigkeit von Software des Typs „Black Box“, künstlicher Intelligenz oder automatisierten Verfahren ohne gründliches Verständnis der zentralen Parameter führt häufig zu Ungenauigkeiten. Fortschrittliche statistische Software vereinfacht Berechnungen, ersetzt jedoch nicht das Verständnis des verwendeten Modells.

Gruppe: Wir bewerteten Veränderungen des Gesundheitszustands vor und nach Behandlung und Rehabilitation in zwei unterschiedlichen Patientenkollektiven: Patientinnen und Patienten mit Handgelenksfraktur sowie Patientinnen und Patienten mit intramedullärem Osteosarkom. Die muskuläre Erholung nach einer Fraktur erfordert eine präzise Diagnostik, insbesondere den standardisierten Janda Test, der eine zuverlässige Verlaufsbeobachtung ermöglicht. Demgegenüber liegt der Schwerpunkt der Rehabilitation bei Osteosarkom auf dem Vertikalisierungstraining, um die Kontrolle des Körperschwerpunkts zu optimieren und die posturale Muskulatur zu aktivieren. Dieser Ansatz erhält eine maximale Heterogenität der Stichprobe und ermöglicht die Beobachtung von Veränderungen sowohl bei nicht lebensbedrohlichen als auch bei lebensbedrohlichen Erkrankungen.

Methoden: Wir schlagen einen methodologischen Ansatz vor, der eine auf Fuzzy Logik basierende Wertematrix auf verschiedenen Signifikanzniveaus einbezieht, um subjektive, quasi subjektive und objektive Gesundheitsparameter zu bewerten. Daher stellen wir detaillierte manuelle Berechnungen Schritt für Schritt dar, die deskriptive Statistik, Konfidenzintervalle, Korrelationsanalyse, t Tests, F Test, multivariate Analyse, Kaplan Meier Kurven und Log Rank Test umfassen, um methodologische Klarheit sicherzustellen.

Ergebnisse: Unsere Studie zeigte, dass 65 von 600 wissenschaftlichen Arbeiten ungenaue oder nicht optimierte statistische Modelle verwendeten.

Schlussfolgerungen: Ziel dieser Arbeit ist es, einen methodologischen Leitfaden für die korrekte Anwendung fortgeschrittener statistischer Modelle in der wissenschaftlichen Forschung bereitzustellen, insbesondere für Ärztinnen und Ärzte am Beginn ihrer beruflichen Laufbahn. Darüber hinaus soll sie erfahrene Fachleute über mögliche Unklarheiten und Verfahrensschritte bei der Nutzung von Software des Typs „Black Box“ informieren.

6 POĎAKOVANIE

Autori si želajú vyjadriť úprimnú vďaku in memoriam prof. Antonovi Guthovi za jeho cenný prínos k prípravným prácam tejto štúdie. S úctou oceňujeme jeho dlhoročnú podporu vzdelávania a univerzitnej prípravy v oblasti rehabilitačnej medicíny, ako aj jeho významnú úlohu pri budovaní a posilňovaní rehabilitačného zázemia na Slovensku, v Českej republike, Nemecku a Rakúsku. Jeho odborné nasadenie, akademické vedenie a celoživotná oddanosť rozvoju rehabilitačného vzdelávania a praxe zostávajú hlboko cenené a sú v tejto práci s úctou pripomenuté.

REFERENCES

1. Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S. & Wicherts, J. M. The prevalence of statistical reporting errors in psychology (1985–2013). *Behav. Res. Methods* **48**, (2016).
2. Wilson, M. On choosing a model for measuring. *MPR-online* **8**, (2003).
3. Masuadi, E. *et al.* Trends in the Usage of Statistical Software and Their Associated Study Designs in Health Sciences Research: A Bibliometric Analysis. *Cureus* <https://doi.org/10.7759/cureus.12639> (2021) doi:10.7759/cureus.12639.
4. Vickers, A. J. *et al.* Guidelines for Reporting of Figures and Tables for Clinical Research in Urology. *Urology* **142**, (2020).
5. Mishra, P., Pandey, C. M., Singh, U. & Gupta, A. Scales of measurement and presentation of statistical data. *Ann. Card. Anaesth.* **21**, (2018).
6. Marshall, G. & Jonker, L. An introduction to descriptive statistics: A review and practical guide. *Radiography* vol. 16 Preprint at <https://doi.org/10.1016/j.radi.2010.01.001> (2010).
7. Schober, P. & Vetter, T. R. Analysis of Variance in Medical Research. *Anesth. Analg.* **131**, (2020).
8. Taylor, R. Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography* **6**, (1990).
9. Rich, J. T. *et al.* A practical guide to understanding Kaplan-Meier curves. *Otolaryngology - Head and Neck Surgery* **143**, (2010).
10. Keselman, H. J. *et al.* Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Rev. Educ. Res.* **68**, (1998).
11. Crosby, A. W. *The Measure of Reality. The Measure of Reality* (1996). doi:10.1017/cbo9781107050518.
12. Maxim, L. D., Niebo, R. & Utell, M. J. Screening tests: A review with examples. *Inhalation Toxicology* vol. 26 Preprint at <https://doi.org/10.3109/08958378.2014.955932> (2014).
13. Sadler, G. R. *et al.* Sample size considerations when groups are the appropriate unit of analysis. *Applied Nursing Research* **20**, (2007).
14. Krousel-Wood, M. A., Chambers, R. B. & Muntner, P. Clinicians' guide to statistics for medical practice and research: Part I. *Ochsner Journal* vol. 6 Preprint at (2006).
15. Wu, X., Wu, B., Sun, J., Qiu, S. & Li, X. A hybrid fuzzy K-harmonic means clustering algorithm. *Appl. Math. Model.* **39**, (2015).
16. Kaliyadan, F. & Kulkarni, V. Types of variables, descriptive statistics, and sample size. *Indian Dermatol. Online J.* **10**, (2019).
17. Strasak, A. M., Zaman, Q., Pfeiffer, K. P., Göbel, G. & Ulmer, H. Statistical errors in medical research - A review of common pitfalls. *Swiss Medical Weekly* vol. 137 Preprint at <https://doi.org/10.4414/smw.2007.11587> (2007).
18. Ercan, I. *et al.* Misusage of statistics in medical research. *European Journal of General Medicine* vol. 4 Preprint at <https://doi.org/10.29333/ejgm/82507> (2007).
19. Riffenburgh, R. H. & Gillen, D. L. *Statistics in Medicine. Statistics in Medicine* (2020). doi:10.1016/B978-0-12-815328-4.00029-2.
20. McDermid, R. Statistics in medicine. *Anaesthesia and Intensive Care Medicine* vol. 22 Preprint at <https://doi.org/10.1016/j.mpaic.2021.05.009> (2021).
21. Darling, H. S. Are you confident about your confidence in confidence intervals? *Cancer Research, Statistics, and Treatment* **5**, (2022).
22. Greenland, S. *et al.* Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* **31**, (2016).
23. Gogtay, N. J. & Thatte, U. M. Principles of correlation analysis. *Journal of Association of Physicians of India* **65**, (2017).
24. Légeret, C., Stienen, Y., Furlano, R. & Köhler, H. Effectivity of treatment for children with functional dyspepsia. *Sci. Rep.* **12**, (2022).
25. Lalkhen, A. G. & Mccluskey, A. Statistics V: Introduction to clinical trials and systematic reviews. *Continuing Education in Anaesthesia, Critical Care and Pain* **8**, (2008).
26. Riffenburgh, R. H. Managing Results of Analysis. in *Statistics in Medicine* (2012). doi:10.1016/b978-0-12-384864-2.00015-9.
27. Pallauta, J. D., Arteaga, P., Gea, M. M. & Begué, N. Understanding statistical tables: a survey of research. *Boletín de Estadística e Investigación Operativa* **38**, (2022).
28. Müller, H. L. *et al.* Melatonin treatment in obese patients with childhood craniopharyngioma and increased daytime sleepiness. *Cancer Causes and Control* **17**, (2006).
29. Kim, T. K. T test as a parametric statistic. *Korean J. Anesthesiol.* **68**, (2015).
30. Xu, M. *et al.* The differences and similarities between two-sample t-test and paired t-test. *Shanghai Arch. Psychiatry* **29**, (2017).
31. Sitter, B. *et al.* Quantification of metabolites in breast cancer patients with different clinical prognosis using HR MAS MR spectroscopy. *NMR Biomed.* **23**, (2010).
32. Kruse, J. A. Clinical Methods: The History, Physical, and Laboratory Examinations. *JAMA: The Journal of the American Medical Association* **264**, (1990).
33. Robertson, W. G., Marshall, R. W. & Walser, M. Calcium measurements in serum and plasma-total and ionized. *Crit. Rev. Clin. Lab. Sci.* **11**, (1979).
34. Zhang, Y., Zhao, F. & Zhang, Y. Exploring the Effectiveness of Nursing Interventions for Postoperative Pain Management in Aortic Dissection: A Fuzzy Logic and Network Analysis. *J. Multidiscip. Healthc.* **18**, (2025).
35. Liu, H. W. Virtual Reality-Assisted Rehabilitation for Patients Undergoing Total Knee Arthroplasty. *Am. J. Phys. Med. Rehabil.* **104**, (2025).
36. Giannou, I., Katsina, M., Dimitriadis, Z., Paras, G. & Besios, T. The effect of hippotherapy on people with multiple sclerosis, a systematic review. *Mult. Scler. Relat. Disord.* **97**, (2025).
37. Hardison, M. E. & Roll, S. C. Mindfulness interventions in physical rehabilitation: A scoping review. *American Journal of Occupational Therapy* **70**, (2016).
38. Galińska, E. Music therapy in neurological rehabilitation settings. *Psychiatr. Pol.* **49**, (2015).
39. Gao, A. *et al.* Rapid identification of Salmonella serovars Enteritidis and Typhimurium using whole cell matrix assisted laser desorption ionization – Time of flight mass spectrometry (MALDI-TOF MS) coupled with multivariate analysis and artificial intelligence. *J. Microbiol. Methods* **213**, (2023).
40. Besse, L. *et al.* Protocol for automated multivariate quantitative-image-based cytometry analysis by fluorescence microscopy of asynchronous adherent cells. *STAR Protoc.* **4**, (2023).
41. Barberis, E. *et al.* Precision Medicine Approaches with Metabolomics and Artificial Intelligence. *International Journal of Molecular Sciences* vol. 23 Preprint at <https://doi.org/10.3390/ijms231911269> (2022).
42. Zhang, Z. & Castelló, A. Principal components analysis in clinical studies. *Ann. Transl. Med.* **5**, (2017).

43. Mainous, A. G. Factor analysis as a tool in primary care research. *Fam. Pract.* **10**, (1993).
44. Juanita, S. & Cahyono, R. D. K-MEANS CLUSTERING WITH COMPARISON OF ELBOW AND SILHOUETTE METHODS FOR MEDICINES CLUSTERING BASED ON USER REVIEWS. *Jurnal Teknik Informatika (JUTIF)* **5**, (2024).
45. Sgarro, G. A. *et al.* The Role of BIA Analysis in Osteoporosis Risk Development: Hierarchical Clustering Approach. *Diagnostics* **13**, (2023).
46. Liao, M., Li, Y., Kianifard, F., Obi, E. & Arcona, S. Cluster analysis and its application to healthcare claims data: A study of end-stage renal disease patients who initiated hemodialysis Epidemiology and Health Outcomes. *BMC Nephrol.* **17**, (2016).
47. Pindar Dibal, N. & Akas Abraham, C. On the Application of Linear Discriminant Function to Evaluate Data on Diabetic Patients at the University of Port Harcourt Teaching Hospital, Rivers, Nigeria. *American Journal of Theoretical and Applied Statistics* **9**, (2020).
48. Zhu, F., Gao, J., Yang, J. & Ye, N. Neighborhood linear discriminant analysis. *Pattern Recognit.* **123**, (2022).
49. Koçhan, N., Tutuncu, G. Y., Smyth, G. K., Gandolfo, L. C. & Giner, G. qtQDA: quantile transformed quadratic discriminant analysis for high-dimensional RNA-seq data. *PeerJ* **7**, (2019).
50. Erkartal, H. Ş., Tatlı, M., Secgin, Y., Toy, S. & Duman, B. S. Gender Estimation with Parameters Obtained From the Upper Dental Arcade by Using Machine Learning Algorithms and Artificial Neural Networks. *European Journal of Therapeutics* **29**, (2023).
51. Wang, L. S., Aziz, Z. & Chik, Z. Disposal practice and factors associated with unused medicines in Malaysia: a cross-sectional study. *BMC Public Health* **21**, (2021).
52. Brasier, A. R. *et al.* Predicting intermediate phenotypes in asthma using bronchoalveolar lavage-derived cytokines. *Clin. Transl. Sci.* **3**, (2010).
53. Revilla, L. *et al.* Multi-omic modelling of inflammatory bowel disease with regularized canonical correlation analysis. *PLoS One* **16**, (2021).
54. Gossmann, A., Zille, P., Calhoun, V. & Wang, Y. P. FDR-Corrected Sparse Canonical Correlation Analysis with Applications to Imaging Genomics. *IEEE Trans. Med. Imaging* **37**, (2018).
55. Lee, G. *et al.* Comprehensive Computed Tomography Radiomics Analysis of Lung Adenocarcinoma for Prognostication. *Oncologist* **23**, (2018).
56. Schwartz, L. R., Fernandez, R., Kouyoumjian, S. R., Jones, K. A. & Compton, S. A Randomized Comparison Trial of Case-based Learning versus Human Patient Simulation in Medical Student Education. *Academic Emergency Medicine* **14**, (2007).
57. Goel, M., Khanna, P. & Kishore, J. Understanding survival analysis: Kaplan-Meier estimate. *Int. J. Ayurveda Res.* **1**, (2010).
58. HOLUBOVA, H. Kaplan-Meier Survival Curves: Simulation Technique. *Scientific Bulletin of the National Academy of Statistics, Accounting and Audit* <https://doi.org/10.31767/nasoa.3-4-2021.02> (2021) doi:10.31767/nasoa.3-4-2021.02.
59. Bollschweiler, E. Benefits and limitations of Kaplan-Meier calculations of survival chance in cancer surgery. *Langenbeck's Archives of Surgery* vol. 388 Preprint at <https://doi.org/10.1007/s00423-003-0410-6> (2003).
60. Koletsi, D. & Pandis, N. Survival analysis, part 2: Kaplan-Meier method and the log-rank test. *American Journal of Orthodontics and Dentofacial Orthopedics* vol. 152 Preprint at <https://doi.org/10.1016/j.ajodo.2017.07.008> (2017).
61. Jager, K. J., Van Dijk, P. C., Zoccali, C. & Dekker, F. W. The analysis of survival data: The Kaplan-Meier method. *Kidney Int.* **74**, (2008).
62. Lin, R. S. & León, L. F. Estimation of treatment effects in weighted log-rank tests. *Contemp. Clin. Trials Commun.* **8**, (2017).
63. D'ascanio, M. *et al.* Age is not the only risk factor in COVID-19: the role of comorbidities and of long staying in residential care homes. *BMC Geriatr.* **21**, (2021).
64. Shen, J. *et al.* Identification of a novel gene signature for the prediction of recurrence in HCC patients by machine learning of genome-wide databases. *Sci. Rep.* **10**, (2020).