O POPRAWNYM STOSOWANIU STATYSTYKI W ROZPRAWACH DOKTORSKICH - SPOJRZENIE NAUKOWCA

Emilia Wołowiec-Korecka, Politechnika Łódzka

W nauce statystyka odgrywa rolę trudną do przecenienia. Jest nie tylko zestawem narzędzi matematycznych, lecz przede wszystkim sposobem myślenia o danych, który pozwala na rzetelne wyciąganie wniosków i ocenę wiarygodności wyników badań. W przypadku rozpraw doktorskich – dokumentów o szczególnej wadze w rozwoju młodego naukowca – poprawne stosowanie metod statystycznych ma znaczenie podwójne: decyduje o jakości i sile argumentów przedstawionych w pracy oraz stanowi świadectwo dojrzałości badawczej doktoranta, a więc także kompetencji, które będzie on wykorzystywał w dalszej karierze naukowej lub zawodowej. Poniższe rozważania mają na celu pomóc w dwóch momentach kluczowych dla jakości rozprawy: (1) na etapie planowania i budowy bazy danych oraz (2) na etapie analizy, interpretacji i raportowania.

Znaczenie statystyki w procesie badawczym

Według klasycznej definicji, statystyka jest nauką o badaniu prawidłowości występujących w zjawiskach masowych [1]. Jej główną funkcją jest umożliwienie badaczowi przejścia od danych empirycznych – często nieuporządkowanych, pełnych zmienności i niepewności – do wniosków, które mogą być uogólnione na szerszą populację. W tym sensie statystyka pełni rolę języka, w którym naukowcy komunikują swoje obserwacje i odkrycia.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

W badaniach prowadzonych na potrzeby rozprawy doktorskiej prawidłowości rzadko wyłaniają się same; najczęściej potrzebują uważnie skomponowanego projektu badania, skrupulatnie zgromadzonych danych i właściwie zastosowanych metod analitycznych. W wielu dziedzinach – od inżynierii materiałowej, przez biologię i medycynę, po ekonomię eksperymentalną – doktorant dysponuje dziś narzędziami obliczeniowymi, które pozwalają „skonstruować" i „uruchomić" niemal dowolny test lub model. Komputer przyjmie wiele, lecz to badacz decyduje, czy używa metody właściwej i czy raportuje wyniki w sposób przejrzysty. Zatem w przypadku dysertacji doktorskiej poprawne użycie metod statystycznych ma dodatkowy wymiar: pokazuje, że kandydat do stopnia naukowego doktora potrafi świadomie planować badania, krytycznie analizować wyniki i interpretować je w kontekście literatury. Statystyka staje się tu narzędziem legitymizującym wnioski – recenzenci rozprawy zwracają szczególną uwagę na to, czy autor nie tylko zna techniki, ale również stosuje je adekwatnie do rodzaju danych i postawionych pytań badawczych.

Projekt badania: pytania, hipotezy, wielkość próby, moc

Rozważania nad rozprawą doktorską warto rozpocząć od określenia populacji badawczej. Zazwyczaj termin „populacja" jest definiowany dość powierzchownie, jako „cała zbiorowość interesująca badacza", co w praktyce na samym początku drogi młodego naukowca wprowadza go w błąd. Tymczasem w dysertacjach doktorskich **populacja to bardzo często mechanizm zjawiska** generujący obserwacje a nie zbiorowość obiektów, które udało się zbadać [1]. Ta zmiana perspektywy ma istotne konsekwencje: jeśli analizowana jest np. trwałość warstw azotowanych na powierzchni stalowych części maszyn – z tej perspektywy nie chodzi o pisanie pracy „o 120 próbkach ze stali X z laboratorium Y", tylko o **mechanizmach** odpowiedzialnych za kształtowanie właściwości warstwy w określonych warunkach procesu. Jeśli badanie dotyczy skuteczności interwencji klinicznej – sensem pracy nie jest pisanie o „tych 141 pacjentów z naszej kliniki", lecz o **procesie chorobowym i odpowiedzi na leczenie** w definiowalnej populacji pacjentów.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

Właściwe określenie populacji na pierwszym etapie prac sprawia, że kolejne kroki będą o wiele łatwiejsze. Łatwiej będzie przejść do zasadniczej części pracy, to jest sądu o mechanizmie zjawiska (populacji), czyli **hipotezy statystycznej**, która będzie testowana na podstawie próby. Hipoteza zerowa H_0 zwykle stwierdza brak efektu (różnic/zależności), zaś alternatywna H_1 – jego obecność. Dobrą praktyką w pracach badawczych jest **formułowanie hipotez merytorycznych i ich odpowiedników statystycznych**. Przykład (techniczny) takiego zestawu mógłby być następujący:

1. **Hipoteza merytoryczna:** Obróbka cieplno-chemiczna stali narzędziowych metodą azotowania w temperaturze T_2 w porównaniu z T_1 zwiększa średnią twardość powierzchni o co najmniej Δ* w sposób istotny technologicznie.
2. Odpowiednik statystyczny może przyjąć postać hipotezy jednostronnej:

   ***H_0*:** Obróbka cieplno-chemiczna stali narzędziowych metodą azotowania w temperaturze T_2 nie zwiększa średniej twardości powierzchni w porównaniu z T_1 ($\mu_{T_2} \leq \mu_{T_1}$).

   ***H_1*:** Obróbka cieplno-chemiczna stali narzędziowych metodą azotowania w temperaturze T_2 zwiększa średnią twardość powierzchni w porównaniu z T_1 ($\mu_{T_2} > \mu_{T_1}$).

Ustalenie hipotez badawczych jest pierwszą sugestią, że czas rozpocząć rozważania nad doborem testu statystycznego. Czy porównywane będą dwie grupy badawcze, czy może więcej? Czy grupy badawcze stanowią układ prób niezależnych czy układ prób zależnych? A może badana będzie jedna grupa, lecz potrzebne jest jej porównanie z grupą kontrolną? Odpowiedź na te i podobne pytania automatyczne zawęża katalog testów statystycznych możliwych do zastosowania. Drugą wskazówką jest liczebność próby, ponieważ statystyka rozróżnia testy statystyczne również ze względu na liczebność obiektów badanych. Liczne źródła literaturowe podkreślają, że dopasowanie testu statystycznego do rodzaju danych (ilościowych, porządkowych lub nominalnych) oraz spełnianych przez nie warunków (niezależności obserwacji a w przypadku danych ilościowych dodatkowo normalności rozkładu i homogeniczności wariancji) jest kluczowym elementem poprawnej analizy

Copyright © StatSoft Polska 2025, info@StatSoft.pl

danych [5]. Prawdą jest jednak, że **planowanie wielkości próby** sprawia wiele problemów młodym naukowcom, zwłaszcza, że w rzeczywistości badawczej mogą pojawić się niezależne od badacza uwarunkowania narzucające liczebność próby (np. pojedynczy obiekt do badań kosztuje kilka tysięcy złotych lub dane zjawisko badawcze zaistniało niewielką ilość razy). W warunkach, kiedy obiekty badawcze są stosunkowo łatwo dostępne, wielkość próby można wyznaczyć określając oczekiwaną **wielkość efektu** (z literatury/pilotażu), **wariancję**/**rozrzut**, akceptowany **poziom istotności α** i **pożądaną moc 1−β**. W rozprawach doktorskich często spotyka się sytuację, w której autorzy nie uzasadniają wyboru liczebności próby ani nie dyskutują ograniczeń wynikających z jej struktury. Tymczasem wielkość zbiorowości badanej, sposób jej doboru (losowy lub nielosowy), a także potencjalne źródła błędu systematycznego mają bezpośredni wpływ na możliwość uogólnienia wyników, dlatego poprawne podejście wymaga refleksji nad reprezentatywnością i jasnego opisania kryteriów doboru danych. Podsumowując, w rozprawie doktorskiej należy pokazać, że autor nie tylko zna nazwy testów, lecz potrafi uzasadnić ich wybór. Nadużywanie testu t-Studenta czy analizy wariancji (ANOVA) bez sprawdzenia założeń jest poważnym uchybieniem metodologicznym.

W badaniach technicznych często bardziej sensowne jest projektowanie pod kątem wykrycia **efektu minimalnie istotnego technologicznie (MIT)** niż „jakiejkolwiek różnicy". Jego odpowiednikiem w badaniach klinicznych będzie **MCID** (ang. *minimal clinically important difference*).

Przygotowanie bazy danych

Bazę do zbierania danych najlepiej zorganizować w postaci tabeli; można to zrobić ręcznie, wykorzystać program Excel lub arkusz danych pakietu Statistica. Korzystne jest przyjęcie reguły, że każda zmienna (płeć pacjenta, ciśnienie, temperatura, itd.) to **kolumna** tabeli, zaś każda badana jednostka/obiekt to **wiersz**. Pytania wielokrotnego wyboru (np. obecność objawów, rodzaje defektów) najwygodniej jest umieścić w osobnych kolumnach binarnych (Tak=1, Nie=0). Ważne jest rozróżnienie sytuacji typu **„brak"** (tj. brak danych dla badanego

Copyright © StatSoft Polska 2025, info@StatSoft.pl

obiektu/jednostki) od **„nie występuje”** (tj. dane zjawisko dla badanego obiektu nie wystąpiło). W programie Statistica obowiązuje konwencja, że 0 oznacza **stwierdzony brak zjawiska**, zaś puste pole to **brak informacji**. Wprowadzanie kodów typu „bd”, „NA” może spowodować utrudnienie analizy w gotowych programach do analizy. W przypadku Excela należy zauważyć, że jest on wygodny na etapie wprowadzania danych, lecz równocześnie jest bardzo wrażliwy na stosowanie zagnieżdżonych nagłówków, scalonych komórek i stosowanie kodów słownych zamiast słowników wartości.

Choć niewielu badaczy to czyni, warto rozważyć prowadzenie **słownika danych**, który ułatwi ich weryfikację; jest to szczególnie pomocne przy złożonych projektach, gdzie badania trwają latami. Słownik danych również może mieć formę tabeli, w której w kolejnych wierszach zostaną zanotowane: nazwa zmiennej, typ, skala, możliwe wartości, zasady kodowania, reguły walidacji (Tabela 1). Poza weryfikacją danych istotne jest również zweryfikowanie ich spójności, np. przy gromadzeniu danych o dacie zabiegu medycznego oraz daty kontroli po zabiegu, właściwy warunek weryfikacji mógłby być następujący : „data kontroli” ≥ „data zabiegu”. W momencie analizy bardzo ważne jest **„zamrożenie bazy”**, ponieważ dopisywanie kolejnych przypadków po rozpoczęciu analizy jest proszeniem się o problemy.

Tabela. 1. Przykładowy słownik danych w badaniu.

nazwa danej	skala	dopuszczalne wartości	jednostka	kodowanie	reguły walidacji
twardość	liczbowa ilorazowa	300 ÷ 900	HV	liczba	0 ≤ twardość ≤ 1200 HV
wykształcenie	jakościowa porządkowa	podstawowe, średnie, wyższe	-	literami: P, Ś, W	wartość jest którymś z wariantów

Analiza, interpretacja i raportowanie wyników badań

Przed przejściem do weryfikacji hipotez, warto pokusić się o precyzyjną i przejrzystą statystykę opisową oraz wizualizację danych, gdyż często wskazywanym uchybieniem w rozprawach doktorskich jest nieczytelna prezentacja danych a w szczególności: brak

Copyright © StatSoft Polska 2025, info@StatSoft.pl

opisowych statystyk, brak wizualizacji lub stosowanie wykresów, które wprowadzają w błąd. Ta część pracy powinna prezentować wyniki w sposób umożliwiający samodzielną ocenę jakości danych przez czytelnika, dlatego przed rozpoczęciem opisywania testów, warto te dane opisać:

- dla cech jakościowych warto przedstawić tabelę podsumowującą dane zawierające kolumny takie jak: zmienna, warianty zmiennej, liczba pacjentów, procent,
- dla cech ilościowych warto podać średnią ($\bar{x}$), medianę, odchylenie standardowe (SD, *Standard Deviation*), wartości min-max, opcjonalnie błąd standardowy (SE, *Standard Error*) i przedziały ufności dla średniej (CI, *Confidence Interval*), jeśli dla rozprawy ważne jest precyzyjne oszacowanie wartości przeciętnej,
- na wykresach warto przedstawić histogramy i wykresy gęstości dla rozkładów empirycznych; wykresy pudełkowe mają zastosowanie dla porównań, wykresy rozrzutu – dla zależności,
- lepiej unikać zbędnej redundancji tekst–tabela–wykres; wiele czasopism naukowych podkreśla, że każde medium powinno wnosić coś innego (np. tabela – wartości, wykres – rozkład) i ma to zastosowanie także do prac doktorskich.

Godny rozważenia jest postulat prof. A. Sokołowskiego, że stosowanie zapisu $\bar{x} \pm SE$, gdzie błąd średni informuje o ile przeciętnie różni się średnia od nieznanej wartości przeciętnej w populacji, jest bardziej sensowne od powszechnie stosowanego zapisu $\bar{x} \pm SD$. Jednakże mając na uwadze przywiązanie recenzentów tej konwencji zapisu, warto być również przygotowanym na wyjaśnienie tak przyjętej notacji.

O weryfikacji hipotez powiedziano i napisano wiele [2-8], mimo to właściwy wybór testu może sprawiać problemy badaczowi stojącemu u progu kariery naukowej. Podsumowując literaturę tematu, poniżej zestawiono listę typowych zadań badawczych oraz zestaw wyborów testów z praktycznym komentarzem. Zestaw ten wynika z doświadczeń Autorki i nie jest opracowaniem kompleksowym, jest raczej punktem startu do własnych poszukiwań.

Zadanie badawcze: dwie grupy badawcze, badana zmienna: ilościowa

Copyright © StatSoft Polska 2025, info@StatSoft.pl

- Rozważ **test t Studenta dla dwóch średnich** (z korektą na nierówność wariancji — wersja Welcha), gdy rozkłady są zbliżone do normalnych a zbiorowość badana *n* umiarkowana/duża.
- Rozważ test **Manna–Whitneya** jako alternatywę nieparametryczną (test operuje na rangach i jest sensowny przy skali co najmniej porządkowej, najlepiej ilorazowej/przedziałowej).
- Uwagi: w pracy raportuj różnicę średnich/median, *p*, **d** (Cohena) i 95% CI.

Zadanie badawcze: więcej niż dwie grupy badawcze, badana zmienna: ilościowa

- Rozważ test **jednoczynnikowej ANOVA**; jeśli wariancje są nierówne a liczebności różne – rozważ test **Welcha**.
- Wykonaj **testy post-hoc** z korektą na wielokrotność (test Tukey'a, HSD, test Gamesa–Howella przy nierówności wariancji).
- Rozważ jako alternatywę test **Kruskala–Wallisa** (wtedy raportuj różnice w parach na rangach z korektą).

Zadanie badawcze: zależności między zmiennymi ilościowymi

- Rozważ **korelację Pearsona** (zawsze obejrzyj wykres rozrzutu).
- Dla danych, których rozkłady odbiegają od normalnych rozważ test **Spearmana** lub **Kendalla**.
- Pamiętaj, że „brak korelacji" nie oznacza „brak związku" (istnieje możliwość, że badany jest związek nieliniowy).

Zadanie badawcze: zmienne jakościowe vs. jakościowe

- Rozważ **test niezależności χ^2** w tabeli kontyngencji a przy małych oczekiwanych częstościach - **dokładny test Fishera**.
- Raportuj **miary siły związku** (współczynnik V Craméra, współczynnik kontyngencji φ) i CI.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

Ponieważ przy bardzo dużej wielkości próby nawet mały efekt bywa „istotny statystycznie”, dlatego lepiej planować i raportować wartości **efekt + przedział ufności**, oprócz samej wartości *p*.

Warto pamiętać, że nie każde odchylenie od normalności dyskwalifikuje test *t* lub metodę ANOVA, zwłaszcza przy liczebności $n > 30–50$ (działa wtedy centralne twierdzenie graniczne). Ważniejsze często są:

- **symetria/ograniczenie wpływu skrajnych obserwacji** (rozważ transformacje, metody odporne, trimming),
- **homogeniczność wariancji** (sprawdź wykresy reszt a test Levene'a traktuj pomocniczo),
- **niezależność** obserwacji (najczęściej łamana przez zagnieżdżenia/powtórzenia).

W rozprawach doktorskich sensowne bywa testować **równoważność** (TOST) albo **nie gorszą jakość** względem metody standardowej. Wymaga to **z góry zdefiniowanych granic równoważności** opartych na przesłankach merytorycznych (technologicznych/klinicznych). Zaletą tego podejścia jest bezpośrednia odpowiedź na pytanie praktyczne: „Czy nowa metoda jest co najmniej tak dobra jak stara w granicach, które uznajemy za istotne?”

Jednym z częstszych błędów w analizie statystycznej jest traktowanie wartości *p* jako jedynego wyznacznika istotności wyników. Autorzy często utożsamiają „$p < 0,05$” z „odkryciem naukowym”, ignorując znaczenie wielkości efektu czy przedziałów ufności. Tymczasem w pracy doktorskiej, gdzie chodzi o głębsze zrozumienie badanego zjawiska, takie uproszczenie jest niewystarczające. Drugim poważnym problemem jest pomijanie wielkości efektu (ang. *effect size*). Nawet jeśli wynik jest statystycznie istotny, może nie mieć znaczenia praktycznego. Recenzenci coraz częściej oczekują, że doktorant nie tylko wskaże istotność, lecz także oceni siłę związku czy różnicy. Statystyka sama w sobie nie daje pełnej odpowiedzi i jej **wyniki muszą być interpretowane w świetle literatury, teorii i praktyki badawczej**. Dlatego dobrym zwyczajem jest odnoszenie wyników statystycznych do wcześniejszych badań, a nie traktowanie ich jako autonomicznego potwierdzenia hipotez.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

Promotorzy rozpraw doktorskich biorą szczególną odpowiedzialność w zakresie kształtowania świadomości statystycznej doktorantów. To oni powinni sygnalizować dostrzeżone pomyłki przy stosowaniu metod analizy, a następnie wskazywać właściwą drogę lub zachęcać do konsultacji ze specjalistami. Recenzenci z kolei powinni oceniać nie tylko poprawność formalną, lecz także spójność i adekwatność metod. Choć zrozumiałe jest, że pełen entuzjazmu Kandydat do stopnia naukowego doktora pragnie wszystkie etapy pracy wykonać samodzielnie podczas realizacji rozprawy doktorskiej, w praktyce najlepiej sprawdza się współpraca oparta na triadzie doktorant-promotor-statystyk, w której: (1) doktorant opisuje problem merytoryczny i ograniczenia logistyczne, (2) promotor zapewnia spójność celów i dba, by pytania badawcze były istotne naukowo, (3) statystyk/metodolog pomaga dobrać adekwatne metody, oszacować liczebności, zaprojektować bazę, przygotować plan analizy i weryfikację założeń. Rozmowa i „burza trzech mózgów" przed rozpoczęciem zbierania danych prawdopodobnie oszczędzi tygodnie „ratowania" analizy po ich zgromadzeniu.

Podsumowanie

Statystyka w nauce to sposób myślenia o danych, strukturze badań i wiarygodności wniosków, w którym narzędzia do weryfikacji wyników badań są tylko jednym z elementów. Wynika stąd, że stosowanie statystyki w rozprawie doktorskiej powinno zaczynać się już na etapie planowania eksperymentu i stanowić **spójny łańcuch decyzji**: od definicji populacji i jednostki, przez konstrukcję bazy i plan analizy, aż po interpretację i raportowanie. Rzetelna, przejrzysta analiza statystyczna wzmacnia nie tylko siłę wniosków, ale i wiarygodność doktoranta jako badacza. Rozprawa, w której statystyka służy rozumieniu mechanizmu zjawiska – a nie wyłącznie „udowadnianiu różnic" – ma większą szansę stać się trwałym wkładem w naukę i praktykę.

Copyright © StatSoft Polska 2025, info@StatSoft.pl

Podziękowania

Autorka składa serdeczne podziękowania dr. Januszowi Wątrobie za merytoryczną weryfikację tekstu, która przyczyniła się do ulepszenia niniejszej publikacji.

Literatura

1. Sokołowski A. (2024) O poprawnym stosowaniu statystyki w publikacjach naukowych – spojrzenie statystyka, Konferencja Zastosowania statystyki i data mining w badaniach naukowych, 17.10.2024.
2. Stanisz A. Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 1. Statystyki podstawowe. StatSoft, Kraków 2006.
3. Stanisz A. Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 2. Modele liniowe i nieliniowe. StatSoft, Kraków 2007.
4. Stanisz A. Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3. Analizy wielowymiarowe. StatSoft, Kraków 2007.
5. Francuz P., Mackiewicz R. Liczby nie wiedzą, skąd pochodzą. Przewodnik po metodologii i statystyce nie tylko dla psychologów. Wydawnictwo KUL, Lublin 2007.
6. Maliński M. Wybrane zagadnienia statystyki matematycznej w Excelu i pakiecie Statistica. Wydawnictwo Politechniki Śląskiej, Gliwice 2015.
7. Bogacki M. Metody planowania eksperymentów. Politechnika Poznańska, Poznań 2018.
8. Rabiej M. Statystyka z programem Statistica. Helion, Gliwice 2012.

Copyright © StatSoft Polska 2025, info@StatSoft.pl