



Sameer Hanna-Juma M.D., Ph.D.

Sameer Hanna-Jumma, MB CHB (*Bachelor of Medicine, Bachelor of Surgery*), MSc (*Medical Sciences*), FRCA (*Fellowship of the Royal College of Anaesthetists*), FCAI (*Fellowship of The College of Anaesthetists of Ireland*), EDIC (*European Diploma in Intensive Care Medicine*) – starszy asystent w Klinice Anestezjologii i Intensywnej Terapii University Hospitals of Leicester NHS Trusts, w Wielkiej Brytanii. Zatrudniony jako lekarz specjalista na tamtejszym oddziale Anestezjologii i Intensywnej Terapii.

Poniższy rozdział jest polskim tłumaczeniem tekstu *Statistics in Anaesthesia: part one, Anaesthesia Tutorial of the Week 302, 20th January 2014*, który pierwotnie ukazał się w World Federation of Societies of Anesthesiologists. Link do oryginalnego tekstu: <http://bit.ly/1GUflog> [dostęp 17.12.2014]

Praca udostępniona na licencji *Creative Commons Attribution-NonCommercial 3.0 Unported*. Aby zapoznać się z jej warunkami, odwiedź stronę: <http://creativecommons.org/licenses/by-nc/3.0/>.

Podstawy biostatystyki

Wstęp

Statystyka to nauka zajmująca się analizowaniem i liczbowym opisem zjawisk masowych. Rozdział ten obejmuje podstawową wiedzę dotyczącą zastosowania statystyki w praktyce klinicznej. Zostaną w nim omówione następujące zagadnienia:

- rodzaje danych,
- zbieranie danych,
- prezentacja danych,
- wnioskowanie statystyczne.

Rodzaje danych

Dane to obserwacje uzyskane z pomiarów cech zmiennych. Zmienne są wielkościami, które mogą przybierać różne wartości. Pierwszym etapem poprawnej pracy z danymi jest określenie, z jakim typem zmiennych mamy do czynienia. Umożliwi to ich poprawny opis oraz wybór odpowiednich testów statystycznych. Dane dzielimy na dwie główne kategorie: ilościowe i jakościowe. W większości badań spotykamy kombinacje obu tych typów. Dane ilościowe są łatwe do analizowania i stosunkowo obiektywne, natomiast dane jakościowe zapewniają bardziej szczegółowy opis badanej próby statystycznej.

Dane jakościowe

Są to zmienne, które nie posiadają wartości numerycznych. Zazwyczaj służą do opisywania znaczenia innych zmiennych – stanowią ich nazwy lub etykiety. Zauważ, że taka etykieta może składać się również z numeru, jak np. w skalach medycznych takich jak ASA (*American Society of Anaesthesiology*, używana do klasyfikowania pacjenta według ryzyka operacyjnego – przyp. tłum.). Kategoria ASA-1 nie oznacza realnej wartości liczbowej równej 1, tylko umowną etykietę nadawaną pacjentowi w dobrej kondycji fizycznej. Dane jakościowe mogą być nominalne lub porządkowe, co zostało przedstawione w tabeli 1.

Tab. 1. Dane nominalne i dane porządkowe

Dane nominalne	Dane porządkowe
<ul style="list-style-type: none"> • Poszczególne kategorie są rozłączne. • Nie posiadają realnej wartości liczbowej. Nazywane binarnymi, jeżeli są jedynie dwie możliwe opcje (np. kobieta, mężczyzna). • Zazwyczaj używa ich się do opisu swoich cech próby. • Przykłady: grupa krwi, typ operacji. 	<ul style="list-style-type: none"> • Dają się uporządkować w sensowny sposób. • Odzwierciedlają rzeczywisty, jednak niemożliwy do zmierzenia stosunek wielkości. • Zazwyczaj używane w ankietach do analizy nienumerycznych zmiennych. • Przykładowo: stopień NYHA, skala ASA, skale napięcia bólu.

Zauważ, że w skalach bólowych, gdzie brak bólu = 0, lekki ból = 1, umiarkowany ból = 2, ciężki ból = 3, umiarkowany ból jest większy niż lekki, jednak nie oznacza to, że jest od niego dwukrotnie silniejszy.

Dane ilościowe

Są to prawdziwie zmienne liczbowe. Rozróżnia się wiele typów danych ilościowych, zazwyczaj jednak nie mają one wpływu na wybór testu statystycznego. Zmienne ilościowe mogą być dyskretne lub ciągłe.

Tab. 2. Zmienne dyskretne i ciągłe

Dyskretne	Ciągłe
<ul style="list-style-type: none"> • Zmieniają się skokowo. • Np. liczba zębów, liczba złamanych kości. 	<ul style="list-style-type: none"> • Przyjmują dowolne wartości z określonego przedziału liczbowego. • Np. wzrost człowieka, poziom kreatyniny w surowicy, stężenie hemoglobiny.

Zbieranie danych

Istnieją różne sposoby zbierania danych. Najbardziej powszechne z nich to prosta próba losowa oraz losowanie warstwowe.

Prosta próba losowa

O prawdziwej próbie losowej mówimy, kiedy każda jednostka danej populacji ma równe szanse zostać włączona do badania. Jest to najbardziej realistyczna reprezentacja populacji generalnej. W badaniach medycznych trudno jednak osiągnąć całkowicie losowy dobór próby. Na przykład, jeśli badasz wpływ monitorowania parametrów hemodynamicznych na wynik ratunkowej laparotomii, powinieneś teoretycznie włączyć do badania wszystkich pacjentów na świecie przechodzących tę procedurę.

Losowanie warstwowe (ze stratyfikacją)

Próba losowa jest w nim podzielona na podgrupy zwane warstwami. Robi się tak, gdy istnieje potrzeba stworzenia grupy będącej dokładniejszą reprezentacją populacji generalnej (uwzględniając jej niejednorodność – przyp. tłum.). Tym samym redukuje się wpływ czynników mogących zaburzyć badanie. We wcześniej przytoczonym przykładzie, próba statystyczna mogłaby zostać podzielona na dwie podgrupy – palaczy oraz osoby niepalące. Następnie pacjenci byliby losowo przydzielani do grupy bez urządzenia monitorującego parametry hemodynamiczne (kontrolnej) lub do grupy z monitorowaniem parametrów krążenia (badanej). Kontrolując ilość palaczy w obu grupach, eliminujemy wpływ palenia papierosów na wynik laparotomii.

Innymi schematami doboru próby są: losowanie grupowe, losowanie wielostopniowe i losowanie wielofazowe.

Prezentacja danych

Podczas prowadzenia badań gromadzimy duże ilości surowych danych. Dane te można przedstawić na trzy odrębne sposoby: za pomocą tabel, graficznie (jako wykresy) oraz parametrycznie (dzięki statystykom opisowym). Każdy ze sposobów prezentacji posiada swoją przydatność, zalety oraz wady.

Tabele

Tabele licznosci oraz tabele zbiorcze są powszechnie stosowane do prezentowania danych w badaniach klinicznych. Mogą być używane do przedstawienia wszystkich typów zmiennych, w tym danych nominalnych, porządkowych oraz ilościowych. Dane ciągłe muszą wpieryw zostać podzielone na przedziały liczbowe. Poniższe przykłady pokazują wyniki punktacji ASA oraz poszczególne grupy wiekowe pewnej badanej próby.

Tab. 3. Przykład tabeli – kategorie ASA

Kategoria ASA	Liczność (ilość pacjentów n=390)
ASA-1	60
ASA-2	230
ASA-3	65
ASA-4	32
ASA-5	3

Tab. 4. Przykład tabeli – grupy wiekowe

Kategoria wiekowa (lata)	Liczność	Skumulowana liczność
20-40	24	24
41-60	210	234
61-80	126	360
81-100	30	390

Tabela częstości względnej jest kolejnym sposobem na zaprezentowanie powyższych danych. Liczby bezwzględne zastępuje się w nich wartościami procentowymi.

Tab. 5. Przykład tabeli – tabela procentowa

Kategoria ASA	Względna częstość (% pacjentów w każdej kategorii; n=390)
ASA-1	15,30
ASA-2	58,90
ASA-3	16,60
ASA-4	8,20
ASA-5	0,77

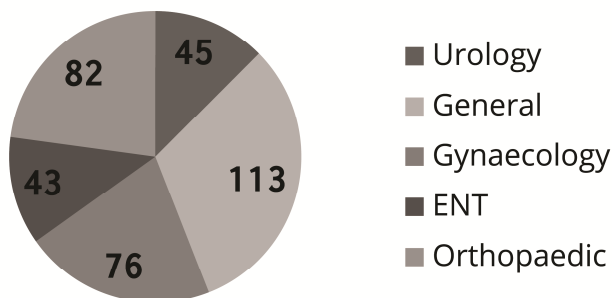
Inne typy tabel licznosci to tabele przekrojowe oraz kategoryzacyjne.

Prezentacja graficzna (wykresy)

Wykresy i diagramy są bardzo ważnymi sposobami prezentowania dużych zbiorów danych. Stanowią potężne narzędzie do podkreślenia relacji między nimi. Poniższe typy znajdują częste zastosowanie w statystyce klinicznej.

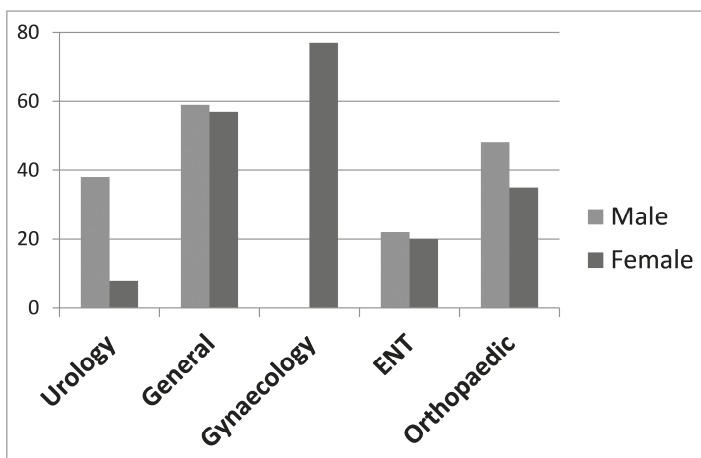
- **Wykresy kołowe.** Wykresy kołowe pokazują proporcje między podgrupami składającymi się na grupę badaną. Pełny diagram reprezentuje całość grupy statystycznej, podczas gdy rozmiar każdej z podgrup jest odpowiedni do jej licznosci. Wykresy kołowe są używane do prezentowania danych porządkowych i nominalnych. Mogą być przydatne, kiedy chodzi nam o uwydatnienie nierównowagi wewnątrz badanej grupy lub udokumentowanie potencjalnych czynników zakłócających. Przykład z ryciny 1 pokazuje ilość oddziałów chirurgicznych uczestniczących w badaniu nad użytecznością kliniczną nadgłośniowych urządzeń do udrażniania dróg oddechowych.

Surgical Groups



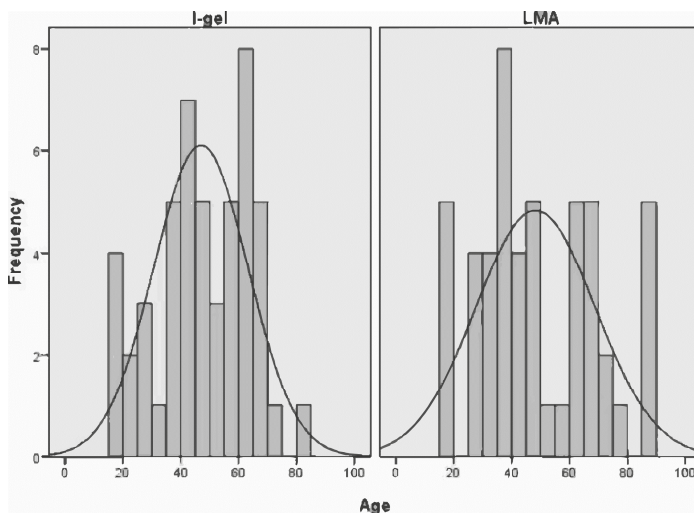
Ryc 1. Wykres kołowy

- **Wykresy słupkowe.** Są używane do porównania odmiennych typów danych. Oś x jest zazwyczaj bezwymiarowa, podczas gdy oś y reprezentuje licznosc poszczególnych kategorii. Każda kategoria może być przedstawiona jako pojedyncza grupa lub może zostać podzielona na podgrupy. Przykładowo na rycinie 2 pokazano w słupkach te same dane co poprzednio. Zwróć uwagę, że tym razem obecny jest także podział według płci.



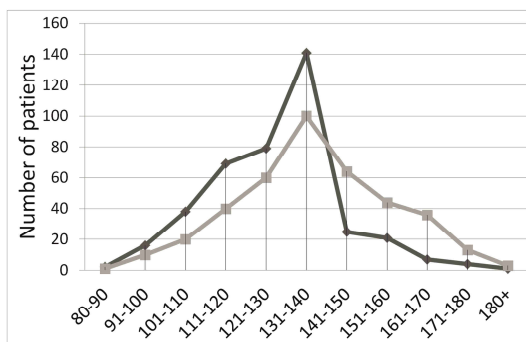
Ryc. 2. Wykres słupkowy

- **Histogramy.** To wyspecjalizowane wykresy słupkowe używane w celu prezentacji danych w przedziałach. Dane ilościowe, a w szczególności zmienne ciągłe, są uprzednio dzielone na przedziały w celu włączenia ich do tabel liczości. Histogramy są użyteczne do wizualizacji sposobu rozkładu danych. Ma to istotne znaczenie przy wyborze odpowiednich testów statystycznych. Na rycinie 3 możemy wyraźnie zauważyć, że dane nie mają rozkładu normalnego. Aby tak było (w uproszczeniu – przyp. tłum.), środek każdego słupka powinien leżeć na krzywej rozkładu normalnego. Do analizy statystycznej tych danych należy więc użyć specjalnych testów nieparametrycznych. Histogramy są również przydatne do demonstracji statystyk opisowych takich jak średnia, mediana i odchylenie standardowe.



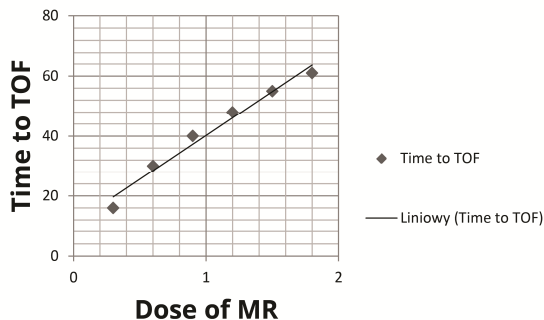
Ryc. 3. Histogramy

- **Krzywe częstości (liczności).** Są bardzo podobne do histogramów, ale nie posiadają typowej reprezentacji słupkowej. Ich zastosowania są również zbliżone. Mają jedną znaczącą przewagę nad histogramami – mogą być użyte do porównania rozkładów dwóch lub więcej serii na jednym wykresie. Na rycinie 4 porównano ciśnienia skurczowego krwi tętniczej w grupie kontrolnej i badanej. W grupie badanej zauważ mniejszą liczbę pacjentów z ciśnieniem powyżej 140 mmHg.



Ryc. 4. Krzywe częstości

- **Krzywe liczebności skumulowanej.** Wykres przedstawia wartości liczbowe na osi poziomej, z równoczesnym przedstawieniem skumulowanej liczebności, skumulowanej liczebności względnej lub skumulowanej liczebności procentowej na osi pionowej. Tego typu diagramy są przydatne do ustalania jaka część grupy znajduje się poniżej lub powyżej pewnej wartości progowej.
- **Wykresy rozrzutu.** Są użyteczne do analizowania związków pomiędzy dwoma zmiennymi. Mogą być również użyte do obliczenia stopnia współzależności zmiennych za pomocą współczynnika korelacji. Przykład przedstawiony na rycinie 5 demonstruje związek pomiędzy dawką nowego leku zwiotczającego (ang. *muscle relaxant*, MR), podanego w mg/kg masy ciała, a czasem bodźca poczwórnego (ang. *Train-of-Four*, TOF) [rodzaj monitoringu funkcji mięśni, używany w anestezjologii – przyp. tłum.]. Dane wskazują na wprost proporcjonalną zależność dawki od czasu TOF.



Ryc. 5. Wykres rozrzutu

- Histogramy są powszechnie stosowane do przedstawiania sposobu rozkładu danych, co jest pomocne w wyborze odpowiednich testów statystycznych.
- Wykresy rozrzutu stanowią ważną pomoc wizualną przy ustalaniu stopnia korelacji w analizie wieloczynnikowej.

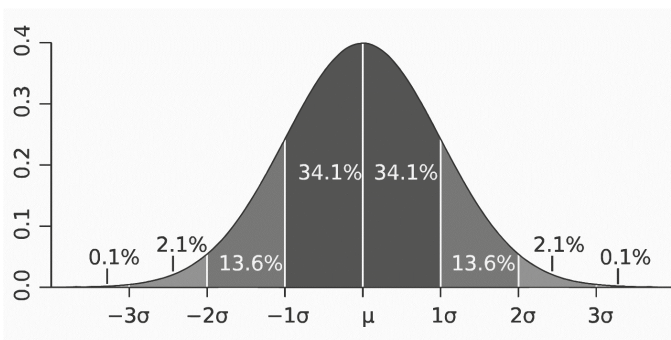
Statystyki opisowe

Pomimo że do przedstawienia statystyk opisowych używamy wykresów oraz tabel, to w naukowych badaniach medycznych interesują nas głównie ich wartości liczbowe. Głównym celem statystyk opisowych jest scharakteryzowanie posiadanej próby statystycznej, a nie wyciąganie wniosków dotyczących całej populacji. Należy więc pamiętać, że choć statystyki opisowe to potężne narzędzie dla podsumowywania różnic pomiędzy zmiennymi, używając ich ryzykujemy przeoczenie pewnych ważnych szczegółów. Statystyki opisowe jako metoda mogą być użyte zarówno do analizy jednoczynnikowej (pojedyncza zmienna), jak i wieloczynnikowej. Charakterystyczne dla analizy jednoczynnikowej są: miary rozkładu, tendencja centralna (miary położenia) oraz miary zmienności.

Rozkład

W zastosowaniach analizy statystycznej na potrzeby służby zdrowia zasadnicze znaczenie ma rozkład normalny (Gaussa). Jest to najczęściej spotykany typ rozkładu danych biomedycznych (takich jak wzrost, waga czy ciśnienie skurczowe) u zdrowych osobników. Graficzna prezentacja wyników (zwłaszcza histogramy) ułatwiają ustalenie typu rozkładu zmiennych, ale istnieje również wiele formalnych testów statystycznych, używanych celem klarownego określenia, czy zebrane dane spełniają warunki rozkładu normalnego. Takimi testami są na przykład testy Shapiro Wilka czy D'Agostino-Pearsona. Normalny (parametryczny) rozkład charakteryzuje się pojedynczym (unimodalnym) szczytem z dwoma symetrycznymi stokami po obu stronach.

Każda z miar tendencji centralnej (średnia, dominanta i mediana) w rozkładzie normalnym jest sobie równa i reprezentowana przez punkt maksymalnej częstości. Rozkład danych jest jednakowy po każdej ze stron i odpowiada odchyleniu standardowemu (ang. *standard deviation*, SD). Dwa parametry (średnia i SD) wystarczają do pełnego opisu tej funkcji (Fig 6.)



Ryc. 6. Krzywa rozkładu normalnego

68% wartości zmiennej znajduje się w odległości jednego odchylenia standardowego (1 SD) od średniej, **95%** w obrębie dwóch odchylenia standardowych (2 SD) i **99,7%** w odległości trzech odchylenia standardowych (3 SD) od średniej.

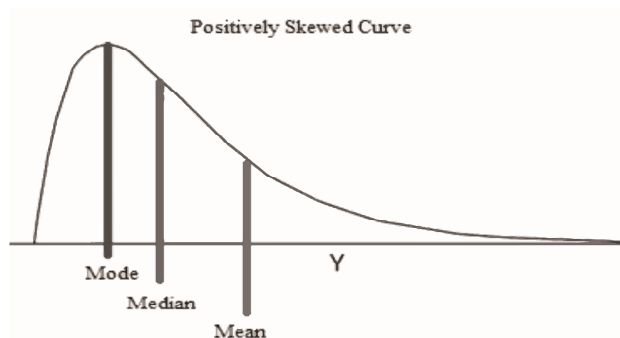
Większość badań medycznych skupia się na wartościach z przedziału **średnia \pm 2SD** (czyli 95% populacji)

Przy danych nieparametrycznych, wartości zmiennej nie są rozłożone równo wokół punktu tendencji centralnej. Wartości mogą być skupione wyłącznie po jednej stronie, a po drugiej być rzadkie – taki rozkład nazywamy rozkładem skośnym (rycina 7). Dane mogą mieć również więcej niż jeden szczyt (tzw. rozkład multimodalny). Kurtoza to jeszcze jeden termin opisujący stopień koncentracji wartości względem szczytu krzywej. Dla rozkładu normalnego kurtoza równa się 0. Krzywa, która posiada bardziej stromy szczyt i smukłe „ogony” ma kurtozę dodatnią, podczas gdy kurtozę ujemną posiada krzywa spłaszczona wzdłuż osi x. W sytuacjach, kiedy rozkład wydaje się odbiegać od normalnego, zawsze bezpieczniej jest użyć nieparametrycznych testów statystycznych.

Tendencja centralna

Tendencja centralna rozkładu to szacunkowy środek rozkładu danych. Trzy główne miary jej położenia to dominanta, mediana oraz średnia arytmetyczna:

- **Średnia** to suma wszystkich wartości podzielona przez ich liczbę (tzw. średnia arytmetyczna). Sprawdza się jako miara tendencji centralnej danych parametrycznych, lecz nie powinna być używana w wypadku zmiennych porządkowych czy nominalnych;



Ryc. 7. Rozkład skośny dodatni. Zauważ różnice pomiędzy wartością dominantą, medianą oraz średnią arytmetyczną

- **Mediana** to wartość środkowa danych. Oznacza to, że po uszeregowaniu rosnąco wszystkich obserwacji powyżej i poniżej niej znajdzie się ich jednakowa liczba. Preferowana do wyliczeń tendencji centralnej dla danych nieparametrycznych, gdyż jest niezależna od obserwacji skrajnych;
- **Dominanta** (wartość modalna) wskazuje na najczęściej występującą wartość w zbiorze danych. Nie jest najlepszym sposobem przedstawiania tendencji centralnej, ale dla danych porządkowych i nominalnych nie ma innego sposobu jej wyliczenia.

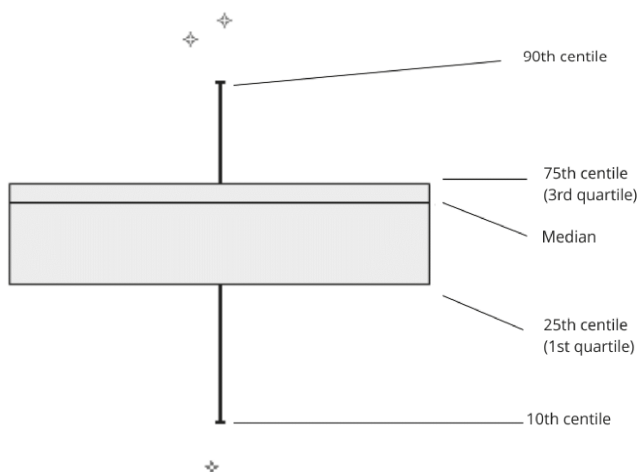
Miary zmienności

Rozstęp to najprostszy sposób pomiaru rozproszenia danych, ale posiadający jednocześnie ograniczone użycie w praktyce. Definiowany jest jako różnica między najwyższą i najniższą wartością w zbiorze danych.

Wariancja oraz odchylenie standardowe (SD) to główne wartości używane do pomiaru rozkładu danych parametrycznych. Pomiaru te są bardziej dokładne, ponieważ obejmują wyliczenia z użyciem wszystkich obserwacji w danej próbie. Wariancja to średnia arytmetyczna z podniesionych do kwadratu różnic pomiędzy wynikami obserwacji a średnią dla danej grupy. Wartości tych różnic są podniesione do potęgi, ponieważ inaczej zniósłoby się wzajemnie. Odchylenie standardowe to pierwiastek kwadratowy z wariancji.

Kwantyle to główne miary rozkładu zbioru danych nieparametrycznych. Określają nam wartości podziału badanej grupy w założonych propor-

cjach. Przy podziale na cztery części mówimy o kwartylach. Pierwszy kwartył to wartość zmiennej większa od 25% wartości danych. Drugi kwartył (mediana) to wartość dzieląca zbiór na dwie równe części. Analogicznie trzeci kwartył to wartość większa od 75% pomiarów. Kwantyle są zwykle pokazane za pomocą wykresu ramka-wąsy (Ryc. 8).



Ryc. 8. Wykres ramka – wąsy

Ramka w tym przypadku obejmuje środkowe 50% obserwowanych wartości w próbie. *Wąsy* opisują 10 i 90 percentyl, a gwiazdki to wartości odstające. Zauważ, że mediana nie jest na środku ramki, co wskazuje na odchylenie od rozkładu normalnego.

Techniki wnioskowania statystycznego

Wnioskowanie statystyczne to proces wyciągania logicznych wniosków ze znanych lub przyjętych za prawdziwe przesłanek przy założeniu, że badana próba statystyczna odzwierciedla populację generalną. Musimy mieć na uwadze, iż cały proces obarczony jest ryzykiem popełnienia błędu, którego prawdopodobieństwo jesteśmy jednak w stanie dokładnie określić. Wnioskowanie statystyczne dzielimy na opisowe (średnia w próbie statystycznej odzwierciedla średnią w populacji generalnej) i analityczne (ocena zależności między różnymi zmiennymi w próbie statystycznej, które można generalizować na całą interesującą nas popula-

cję generalną). Dwa głównie pojęcia we wnioskowaniu statystycznym to próba statystyczna oraz prawdopodobieństwo.

Próba statystyczna

Jest to podgrupa populacji generalnej wykorzystana podczas analizy. Na przykład w badaniu dotyczącym pooperacyjnej oceny hemodynamicznej pacjentów poddanych laparotomii próbą statystyczną są pacjenci, którzy zostali przeanalizowani w danym przedziale czasowym. Populacją generalną będą zaś wszyscy pacjenci wymagający laparotomii. Obliczenie wymaganej wielkości próby statystycznej jest bardzo istotne przed rozpoczęciem badania. Od wielkości próby będzie bezpośrednio zależeć moc statystyczna naszej analizy.

Prawdopodobieństwo

Każde zdarzenie można dać kilka przewidywalnych wyników. Szansa na wystąpienie danego wyniku oscyluje pomiędzy 0 (nigdy) a 1 (zawsze). W badaniach klinicznych prawdopodobieństwo (lub wartość p) jest niezbędnym elementem prezentacji każdego rodzaju danych. Stanowi potwierdzenie, że prezentowany wynik to efekt działania badanych zmiennych, a nie dzieło przypadku. Przeanalizujemy następujący przykład: zbadaliśmy różnicę w zapotrzebowaniu na przetoczenie krwi w grupie kontrolnej oraz grupie osób, które wcześniej otrzymały kwas traneksamowy. Przypuśćmy, że wartość p naszego wyniku wyniosła 0,01. Oznacza to, że zaobserwowana między grupami rozbieżność w zapotrzebowaniu na krew z bardzo małym prawdopodobieństwem może wynikać z błędu losowego (choć nie można go całkowicie wykluczyć). Inne terminy używane w statystyce medycznej do opisu prawdopodobieństwa wystąpienia danego zdarzenia to iloraz szans (ang. *odds ratio*, OR) i ryzyko względne (ang. *risk ratio*, RR).

Iloraz szans

Szansa to stosunek ilości wystąpień dwóch wzajemnie wykluczających się wyników zdarzenia. Przykładowo w pewnej grupie 160 ciężarnych 110 urodziło naturalnie, a 50 poprzez cięcie cesarskie. Stąd szansa porodu poprzez cesarskie cięcie wyniosła $50 : 110$, czyli 0,45. W badaniach klinicznych ważniejszy i częściej używany jest jednak iloraz szans. Służy on do pomiaru wpływu konkretnej interwencji na prawdopodobieństwo wystąpienia jakiegoś zdarzenia. Przykładem będzie iloraz szans dla naturalnego porodu u rodzących w znieczuleniu zewnątrzoponowym w po-

równaniu z grupą kontrolną. W grupie kobiet rodzących w znieczuleniu 90 urodziło naturalnie, a 60 miało cesarskie cięcie. Szansa porodu przez cesarskie cięcie wyniosła więc 60 : 90, czyli 0,66. Iloraz szans dla cesarskiego cięcia między grupą ze znieczuleniem zewnątrzoponowym, a grupą kontrolną ma się jak 0,66 : 0,45, czyli 1,46.

Iloraz szans równy jedności oznacza, że nie zaobserwowano istotnych różnic pomiędzy obydwoma grupami. W rozważanym przypadku iloraz szans wskazuje na większe prawdopodobieństwo porodu poprzez cesarskie cięcie, gdy kobieta została wcześniej znieczulona zewnątrzoponowo.

Ryzyko względne

Ocena ryzyka względnego jest powszechnie używana w badaniach epidemiologicznych. Koncepcyjnie jest zbliżona do ilorazu szans. Ważne, aby zrozumieć, że obliczając ryzyko względne, mianownik ułamka tworzymy uwzględniając liczebność całej badanej populacji. Na przykład ryzyko względne cięcia cesarskiego z poprzedniego przykładu dla grupy kontrolnej wynosiło 50 : 160, czyli 0,312 (w porównaniu z szansą równą 0,45).

Tab. 6. Różnica pomiędzy bezwzględną, względną redukcją ryzyka i liczbą wymagających leczenia

Bezwzględna redukcja ryzyka (ARR)	<ul style="list-style-type: none"> • różnica pomiędzy ryzykiem danego zdarzenia w grupie eksperymentalnej (ang. Experimental event rate, EER) i w grupie kontrolnej (ang. Control event rate, CER) • wyliczane jako $EER - CER$
Względna redukcja ryzyka (RRR)	<ul style="list-style-type: none"> • część ryzyka wystąpienia danego zdarzenia usunięta dzięki zastosowanej interwencji • wyliczane jako iloraz $(EER - CER) / CER$
Liczba wymagających leczenia (NNT)	<ul style="list-style-type: none"> • liczba pacjentów, którzy powinni zostać poddani rozważanej interwencji (leczeniu), aby można było zapobiec wystąpieniu jednemu niekorzystnemu zdarzeniu (np. zgonowi) • odwrotność bezwzględnej redukcji ryzyka

W dużych badaniach epidemiologicznych pomiar bezwzględnej redukcji ryzyka, względnej redukcji ryzyka oraz liczby wymagających leczenia są wykorzystywane do oceny efektywności danej interwencji (tabela 6). Pomagają również w przewidzeniu konsekwencji finansowych wynikających z jej wprowadzenia. Ryzyko względne nie może być wyliczone dla badań kliniczno-kontrolnych, aczkolwiek iloraz szans stanowi jego przybliżenie.

Znaczenie powyższych wskaźników powinno być oceniane w kontekście konkretnej analizy. Znanym faktem jest, iż w badaniach na dużej populacji względna redukcja ryzyka może być złudnie wysoka w porównaniu do realnej korzyści klinicznej. Przykład: ryzyko pooperacyjnych nudności i wymiotów (ang. *postoperative nausea and vomiting*, PONV) po zastosowaniu nowego leku przeciwwymiotnego w porównaniu z lekiem pierwszego wyboru (złotym standardem).

- Grupa A (kontrolna): wymiotowało 10 na 1000 pacjentów
- Grupa B (nowy lek): wymiotowało 5 na 1000 pacjentów

Bezwzględna redukcja ryzyka równa się $0,5 - 1\% = -0,5\%$, podczas gdy względna redukcja ryzyka = $(0,5 - 1\%) / 1\% = -0,5$, czyli -50% .

Niektórzy autorzy użyliby względnej redukcji ryzyka, aby podkreślić zmniejszenie o połowę pooperacyjnych nudności i wymiotów. Jednakże, gdy wyliczysz liczbę wymagających leczenia (*number needed to treat*, NNT) to wyjdzie $100 : 0,5 = 200$. Tylu pacjentów powinno otrzymać leki, aby uniknąć PONV u jednego dodatkowego chorego, w porównaniu ze złotym standardem. Może to podważyć sens zastosowania takiego leczenia, jeśli przy tym lek powoduje częste efekty uboczne lub jest istotnie droższy.

Hipoteza zerowa i hipoteza alternatywna

Metody wnioskowania statystycznego pozwalają na przyjęcie lub odrzucenie danej hipotezy za pomocą analizy danych doświadczalnych. W statystyce medycznej zakłada się, że dane dwie grupy nie różnią się istotnie pod badanym względem, dopóki nie udowodni się, że jest inaczej. Takie rozumowanie nazywa się to hipotezą zerową. Hipoteza alternatywna jest jej zaprzeczeniem, przez co dla badacza jest zwykle bardziej interesująca. Zakłada ona istnienie pewnej różnicy, którą w wypadku prób medycznych przypisujemy zastosowanej interwencji klinicznej. Głównym

celem większości testów statystycznych jest umożliwienie odrzucenia hipotezy zerowej oraz potwierdzenia hipotezy alternatywnej.

Rodzaje błędów

Błąd pierwszego rodzaju (wynik fałszywie dodatni), zwany także błędem α , polega na odrzuceniu hipotezy zerowej, mimo iż jest ona prawdziwa. Oznacza to, że stwierdza się statystycznie istotną różnicę tam, gdzie realnie jej nie ma. W statystyce medycznej istotna jest wartość p , która odzwierciedla najmniejszą wartość α , dla której jesteśmy skłonni odrzucić hipotezę zerową. Zwykle w badaniach medycznych ustalamy jej wartość progową na poziomie 0,05 (innymi słowy, pozwalamy na 1 fałszywie dodatni wniosek na 20 serii badań).

Wartość p powinna być wyliczona i uwzględniona w publikacji naukowej, lecz niewystarczające jest samo określenie jej jako większej lub mniejszej niż 0,05.

Przykładowo, jeśli dwa różne badania analizujące taką samą interwencję miały odpowiednio $p = 0,051$ oraz $p = 0,049$, w rzeczywistości wiarygodność ich wyników jest zbliżona. Podanie wartości p w formie odpowiednio: $<0,05$ lub $>0,05$ zaciemniłoby stan faktyczny.

Błąd drugiego rodzaju (wynik fałszywie ujemny), zwany także błędem β , pojawia się gdy hipoteza zerowa nie zostanie odrzucona, choć w rzeczywistości jest ona fałszywa. Oznacza to, że pomimo istnienia statystycznie istotnej różnicy nie została ona znaleziona. Najczęstszą przyczyną tego błędu jest nieodpowiednia wielkość próby statystycznej. Stąd przed rozpoczęciem pracy bardzo ważne jest wyliczenie minimalnej wielkości próby, przy której można zarówny wykryć różnicę, a zatem uniknąć popełnienia błędu drugiego rodzaju. Nazywa się to ustalaniem mocy testu.

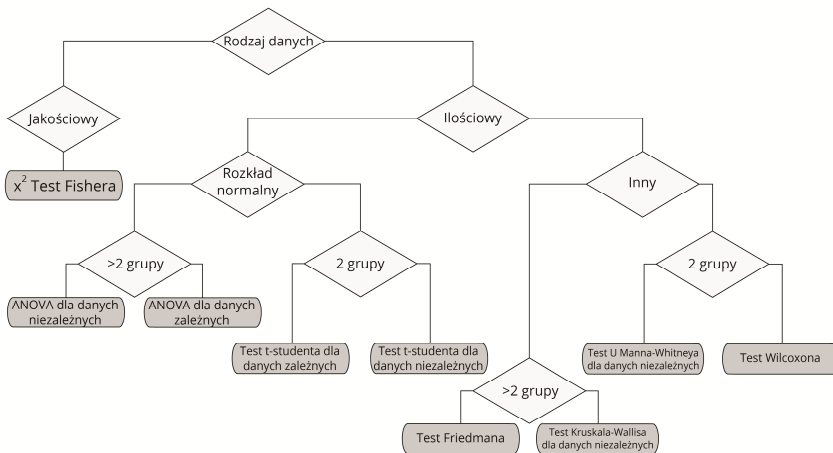
Moc testu jest prawdopodobieństwem wykrycia statystycznie istotnej różnicy, jeśli takowa istnieje. Jest obliczana jako $(1 - \beta)$. Większość badaczy medycznych akceptuje wartość β na poziomie 0,2. Jest mało prawdopodobne, że projekt badania zostanie zaakceptowany bez jej uprzedniego wyliczenia.

Rodzaje testów statystycznych

Bardzo ważny jest wybór odpowiedniego testu statystycznego do danych, które analizujemy. Jest kilka zasad, do których warto się stosować, by wybrać odpowiedni test dla Twoich danych:

1. Określ czy analizowane dane są jakościowe, czy ilościowe.
2. Przy danych ilościowych określ ich rozkład.
3. Określ ile grup będzie analizowanych.
4. Określ czy dane są zależne (sparowane), czy niezależne (niesparowane).

Dokładny charakter tych badań nie będzie omówiony w tym artykule, jednak algorytm wyboru testu przedstawiono na rycinie 9.



Ryc. 9. Algorytm wyboru odpowiedniego testu statystycznego

Próba statystyczna a populacja generalna

Do dokładnego wyznaczenia, w jakim stopniu badana próba odpowiada populacji, z której została wyodrębniona, można użyć kilku narzędzi. Obejmują one błąd standardowy średniej (ang. *standard error of mean*, SEM) oraz przedziały ufności (ang. *confidence intervals*, CI).

Błąd standardowy średniej pokazuje, jak blisko średniej z całej populacji znajduje się wyliczona średnia z próby statystycznej. Uzyskamy go z równania:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

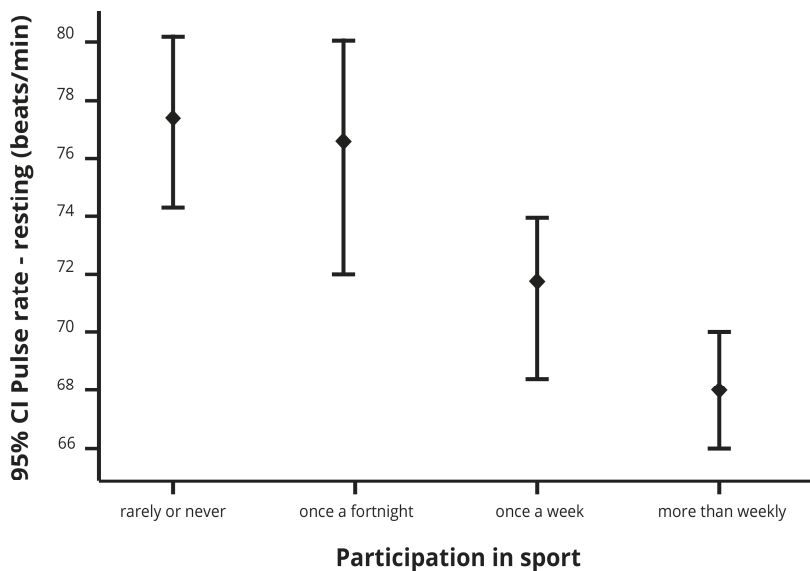
gdzie SE to błąd standardowy średniej, s odchylenie standardowe próby, a n to ilość obserwacji (wielkość próby statystycznej).

Z równania tego wynika, że im większa jest próba statystyczna, tym mniejsza wartość SEM i tym samym lepsze odzwierciedlenie populacji generalnej. Jest to oczywiste, gdyż będzie ona wówczas zawierać więcej informacji o całości, a w związku z tym rezultat będzie bardziej precyzyjny.

Dla wybranego parametru przedział ufności (CI) to szacowany zakres liczbowy, w którym z określonym prawdopodobieństwem leży jego prawdziwa wartość. Zakres ten jest obliczany na podstawie danych z analizowanej próby statystycznej. Czynniki, które wpływają na wielkość przedziałów ufności to wielkość próby statystycznej oraz jej odchylenie standardowe. Im większa próba statystyczna, tym bardziej reprezentatywna dla danej populacji i tym węższe przedziały ufności da się wyznaczyć.

Warto wspomnieć, iż przedział ufności dla mierzonego efektu świadczy przeciwko jego istnieniu, kiedy obejmuje swoim zasięgiem przeciwną tendencję. Na przykład: pewne badanie przeprowadzone wśród pacjentów we wstrząsie septycznym wykazało wzrost średniego ciśnienia tętniczego o 20 mmHg po zastosowaniu wazopresyny. Przedział ufności CI wyniósł jednak od -5 do +45 mmHg, więc zaobserwowany efekt może być bez znaczenia w odniesieniu do całej populacji. Jest tak, ponieważ zmiany w ciśnieniu tętniczym mogą w rzeczywistości równie dobrze wynosić zero, a nawet może być to spadek o 5 mmHg.

Przykład na rycinie 10 pokazuje przedziały ufności w badaniu nad wpływem aktywności fizycznej na tętno spoczynkowe.



Ryc. 10. Przedziały uwidaczniają zakres zmienności obserwowanego tętna spoczynkowego w grupach deklarujących różny stopień zaangażowania w aktywność fizyczną

Bibliografia

1. McCluskey A., Lalkhen A. *Statistics A publication series from Continuous Education in Anaesthesia, Critical Care and Pain Management*, Oxford University Press in 2007.
2. Harris M., Taylor G. *Medical Statistics Made Easy*, 2nd Edition, Scion Publishing Ltd, 2008.
3. Bowers D., *Medical Statistics from Scratch an Introduction for Health Professionals, 2nd Edition*, John Wiley & Sons Ltd, UK 2008.
4. Spoons C., Kiff K. *Oxford Specialty Training: Training in Anaesthesia, 1st Edition*, Oxford University Press in 2010.

Polecana literatura

- <http://www.cyclismo.org/tutorial/R/index.html>
- <http://www.stats.gla.ac.uk/steps/glossary/index.html>
- http://www.statsoft.pl/textbook/stathome_stat.html