Глава 8

ОПИСАНИЕ НА КОЛИЧЕСТВЕНИ ПРОМЕНЛИВИ. ИЗМЕРВАНЕ НА ВАРИРАНЕТО

1. Същност на варирането

Както бе посочено в предходния раздел, варирането е присъщо качество на всички живи организми. Дори в наблюдавани групи, които са привидно много сходни, трудно могат да се намерят, особено при малък брой случаи, абсолютно еднакви стойности на изучаваните количествени променливи.

Здравните професионалисти често трябва да решават дали даден индивид е болен или здрав, дали страда от конкретно заболяване или не, дали се нуждае от лечение или не и т.н. За решаване на такива задачи е необходимо определяне на т.нар. „нормални“ стойности на редица клинични, лабораторни, радиологични и други измервания. Понятието „нормална“ стойност обаче е статистическо понятие и зависи в голяма степен от разпределението на изучавания признак в извадката или популацията. Следователно, измерването на разсейването (варирането) на изучаваните променливи е изключително важно за осмислянето и интерпретирането на понятието „нормални“ стойности и за пълното описание на даден масив от здравни данни. Така че, освен обобщаващите характеристики на централната тенденция трябва да се определят и измерителите на варирането.

Пример: Нека разгледаме два вариационни реда, построени на основата на данни за възрастта на 10 първораждащи жени от различни извадки.

Първият вариационен ред включва следните стойности:

18 21 23 23 25 27 27 28 30 33

Изчислените средни величини са: средна аритметична – 25.5 години, медиана – 26 години и мода – 23 и 27 години.

Вторият вариационен ред е представен чрез:

23 23 24 25 26 26 27 27 27 27

Стойностите на средните величини са същите: средна аритметична – 25.5 години, медиана – 26 години и мода – 27 години.

Вижда се ясно, че двата вариационни реда, макар и да имат еднакви средни аритметични, медиани и моди, са доста различни по отношение на разсейването на индивидуалните стойности. Възниква въпросът „В коя от тези извадки средната аритметична по-добре описва типичното ниво?“ За да отговорим е необходима информация за степента на вариране на данните.

Необходимостта от измерване на числовите характеристики на вариабилността е свързана с:

- присъщото за биологичните обекти вариране, както и с варирането, предизвикано от други източници, които водят до систематично или неслучайно вариране в здравните измервания;
- идеята за обобщаване на варирането в едно единствено число с цел да се улесни сравняването на разсейването между различни групи;
- използването на понятията „нормални стойности“ и „нормален обхват“ в медицинската практика (например, нормални нива на систолното и диастолно налягане, на сърдечен ритъм, ръст, тегло, серумен холестерол, хемоглобин и др.);
- идеята за използване на вариабилността като индикатор за хомогенността или хетерогенността на данните.

2. Мерки за вариабилност (разсейване)

За измерване на варирането могат да се използват следните описателни числови характеристики: ***размах (обсег), интерквартилен обхват, дисперсия, стандартно отклонение, коефициент на вариация.***

2.1. Размах (обхват, обсег, лимит) на вариационния ред

Представлява ***разликата между екстремалните стойности (максималната и минималната) на променливата*** в дадено емпирично честотно разпределение. Означава се обикновено с латинската буква d (от difference – разлика). Следователно: $d = x_{max} - x_{min}$

В посочения по-горе пример обхватът на първия вариационен ред е 15 години, докато за втория – само 4 години. Следователно, вторият вариационен ред е по-компактен, неговите стойности по-плътно прилягат около средното ниво и изчислената средна аритметична е по-добра характеристика на централната тенденция за тази количествена променлива.

Размахът на вариационния ред като мярка за вариабилността се характеризира със следните ***особености:***

- изчислява се много бързо и е лесен за осмисляне и разбиране;
- крайните стойности на размаха са зависими от размера на извадката;
- не се опира на всички измервания, а само на две стойности и то най-нетипичните;
- достатъчно е наличие само на една рязко отклоняваща се стойност и размахът като мярка за варирането става абсолютно невалиден;
- трябва да се използва заедно с други числови характеристики на вариабилността или пък да се представят пълните честотни разпределения и средните аритметични.

Поради тези особености, размахът се използва сравнително рядко като характеристика на вариабилността, тъй като не е достатъчно информативен.

2.2. Интерквартилен обхват (IQR)

Интерквартилният обхват представлява ***разликата между третия и първия квартил*** ($Q_3 - Q_1$) в определен масив от данни. Q_3 е всъщност медианата за втората половина на вариационния ред, Q_1 е медианата за първата половина, а Q_2 е медианата на цялото честотно разпределение.

Следователно, интерквартилният обхват **IQR** дава представа за размаха от 25-я до 75-я персентил, т. е. за 50% от данните, разположени по средата на разпределението.

В посочените по-горе два вариационни реда за възрастта на 10 първораждащи жени:

- за първия вариационен ред $Q_1 = 23$ и $Q_3 = 28$, т.е. IQR = 5 години;
- за втория вариационен ред $Q_1 = 24$ и $Q_3 = 27$, т.е. IQR = 3 години.

Вижда се ясно, че варирането във второто разпределение е по-слабо. По такъв начин, средната аритметична за този ред е по-добра характеристика на централната тенденция, отколкото средната аритметична в първия ред, макар двете да имат еднакви стойности.

Подобно на размаха на вариационния ред, ***интерквартилният обхват е относително груба мярка на разсейването,*** но той все пак предоставя общ поглед за начина, по който се разпределят данните във вариационния ред. Предимството му е в това, че той е доста по-устойчив по отношение на рязко отклоняващи се стойности на разпределението.

2.3. Стандартно отклонение и дисперсия

Както бе посочено по-горе, точното измерване на варирането следва да се опира на всички отклонения на стойностите на променливата величина около средната аритметична, а не на избрани и непредставителни случаи. Такива мерки за вариабилност са ***стандартно отклонение и дисперсията,*** които са доста по-сложни за изчисление в сравнение с посочените по-горе.

Стандартното отклонение е най-често посочваната в научната литература мярка за разсейването. Обикновено средните аритметични величини винаги се посочват с придружаващите ги стандартни отклонения.

Стандартното отклонение представлява ***средното отклонение на резултатите от средната аритметична*** и се означава чрез символите:

s – стандартно отклонение в извадка

σ – стандартно отклонение за популация

При обсъждането на характеристиките на средната аритметична в предходния раздел, бе подчертано, че сумата от квадратите на отклоненията около средната аритметична е по-малка от сумата от квадратите на отклоненията около която и да е друга стойност. Следователно, ***стандартното отклонение е мярка за „най-малките квадрати“ на отклоненията около средната величина.***

Дисперсията и стандартното отклонение имат следните основни ***свойства:***

- тяхното изчисление се опира на всички наблюдения;
- изчисляват се по отношение на средната аритметична, т. е. стандартното отклонение характеризира отклонението от типичното ниво (централната тенденция). Следователно, то дава информация за това дали дадена средна аритметична характеризира добре типичното ниво на количествената променлива. При извадки с еднакви средни аритметични, по-добра характеристика на централната тенденция дава тази средна, която е придружена от по-малко стандартно отклонение.
- стандартното отклонение и дисперсията са най-широко използваните мерки за оценка на разсейването поради

свойствата на теоретичната нормална крива и използването им при оценка и сравняване на данни от репрезентативни проучвания, тъй като стандартното отклонение служи за основа на изчислението на стандартната грешка;

- дисперсията и стандартното отклонение не се променят, ако към индивидуалните стойности прибавим или извадим едно и също число, или ако ги умножим или разделим с едно и също число.
- дисперсията и стандартното отклонение се прилагат при условие, че разпределението на количествените променливи е нормално (симетрично, Гаус–Лапласово) или близко до него.

2.4. Коефициент на вариране

Тъй като стандартното отклонение и дисперсията са именувани величини (т.е. измерват се в същите мерни единици както средните аритметични), то при сравняване на варирането на различни променливи величини, се налага да се използва друг измерител, наречен ***коефициент на вариране***.

Коефициентът на вариране (V) представлява ***отношение на стандартното отклонение към съответната средна аритметична*** и се изразява в ***проценти***.

(8.1) $$V = \frac{s}{\bar{x}} \times 100$$

При стойност на V под 10% варирането е слабо; при 10% $<$ V $<$ 30% варирането е умерено и при $V >$ 30% се наблюдава силно разсейване около средната аритметична, което говори за значителна нееднородност на изучаваната съвкупност.

Коефициентът на вариация притежава следните свойства:

- не зависи от никакви мерни единици и се използва за сравняване на относителното вариране на две или повече разпределения от различни именувани променливи величини (напр., височина в см за едното разпределение и тегло в кг за другото разпределение; чрез сравняване на коефициентите на вариация можем да направим извод коя от двете променливи величини по-добре характеризира физическото развитие);

- измерва разсейването в данните по отношение на средната стойност;
- взема предвид всяка стойност на разпределението.

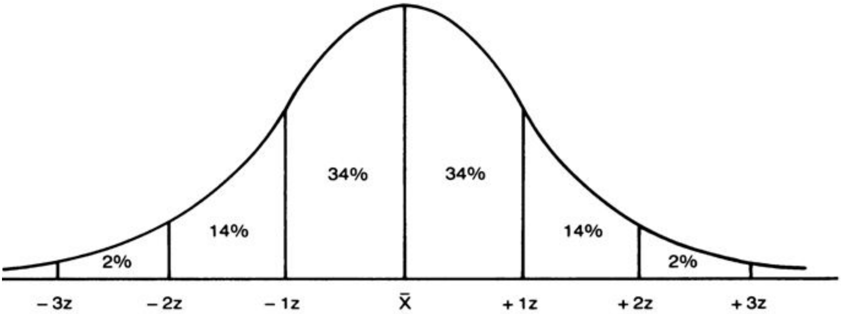
3. Тенденции на варирането. Нормално разпределение

Основната дилема, пред която се изправят често изследователите в медицината и здравната помощ, е фактът, че всеки масив данни неминуемо показва значително вариране на индивидуалните резултати от средното ниво. За наше облекчение повечето от данните, които срещаме в практиката, показват устойчиви, прости и лесноразбираеми модели на вариране.

Да вземем примера с измерване на диастолното кръвно налягане при 56 млади мъже, които са силни пушачи. Резултатите, групирани в интервали с ширина 5 мм, показват, че средната стойност е вероятно някъде между 85 и 95 мм Hg, а размахът на вариационния ред е от 60 до 120 мм Hg. Можем да отбележим също, че по-голямата част от индивидите се намират в пределите на много по-тесен обхват – от 75 до 105 мм Hg и че всички те са разположени почти симетрично около средната стойност.

Ако при повторни наблюдения върху значително по-голям брой случаи представим данните за диастолното налягане по същия начин, ще се убедим, че те още по-ясно ще се групират около средното ниво и с отдалечаване от средното ниво намалява броя на случаите с екстремни резултати. Тези данни добре илюстрират начина, по който се проявяват в практиката болшинството масиви от непрекъснати променливи.

За непрекъснати променливи, разпределени по този модел, се казва, че имат *нормално разпределение*. Графично то се представя с т.нар. *нормална крива,* която е *симетрична, камбановидна* и представлява *теоретично идеален честотен полигон*, *в който средната аритметична, медианата и модата напълно съвпадат и се разполагат в центъра на разпределението* ***(фиг. 8.1)***. Установено е, че много човешки черти (такива като интелигентност, поведение и личностови характеристики) се разпределят в популацията по такъв „нормален" начин.



Фиг. 8.1. Графично изображение на нормалната крива

Определящите характеристики на нормалното разпределение са $\overline{x}$ и ***s.***

Когато средната аритметична е равна на нула и стандартното отклонение е единица, то такова разпределение се нарича ***стандартно нормално разпределение***. Основата на нормалната крива се отмерва в ***единици стандартно отклонение***, които се означават с малка буква ***z.*** Резултат, който е едно стандартно отклонение над средната, се отбелязва чрез ***+1z*** и обратно – резултат, който е едно стандартно отклонение под средната с ***-1z***.

Коефициентът ***z*** може да се изчисли за всеки резултат в едно нормално разпределение по формулата:

$$z = \frac{x - \overline{x}}{s} \qquad (8.2)$$

Между двете определящи характеристики на нормалното разпределение ($\overline{x}$ и **s**) има точно определена връзка, която се изразява чрез ***закона за нормалното разпределение.***

Ако знаем стойностите на средната величина и стандартното отклонение, можем да предвидим точно как ще се разпределят отделните стойности около средното ниво ***(Табл. 8.1).*** Площта между нормалната крива и абсцисата се приема за единица или 100%. Резултатите в границите на $\overline{x} \pm z$ и тези извън $\overline{x} \pm z$, превърнати в %, винаги дават 100, но съотношението им се променя по точно определен начин.

Табл. 8.1. Връзка между средната аритметична и стандартното отклонение при нормално разпределение

Брой стандартни отклонения (*z*, *t*) около средната аритметична	Резултати, попадащи в границите $\overline{x} \pm s$ (в %)	Резултати лежащи извън $\overline{x} \pm s$ (в %)
0.5	38.2	61.4
1	68.2	31.8
1.96	95	5
2.58	99	1
3.00	99.7	0.3
3.29	99.9	0.1