

§1. Предмет на статистиката

1. Увод. Статистиката се занимава със събиране и анализ на **данни**. Данните се получават в резултат от наблюдения на определен брой статистически единици, които в психологията обикновено представляват отделни индивиди. Групата от индивиди, над които е извършено наблюдението, се нарича **извадка (sample)**, а броят на единиците в извадката се нарича **обем** на извадката (**sample size**). Извадката се разглежда като част от по-голяма група, наречена **генерална съвкупност** или още **популация (population)**. Популацията е съвкупността от статистически единици, които са еднородни относно наблюдаваните величини. Ролята на популацията в статистическите изследвания е твърде специфична както от гледна точка на предметната област на изследването, така и от гледна точка на статистическите процедури, които се прилагат за различните цели на анализа, и по тази причина детайлите в познавателното значение на понятието популация ще бъдат разкривани постепенно в рамките на този лекционен курс. Ще отбележим преди всичко, че популацията има винаги в някаква степен хипотетичен характер, а статистическите методи които ще използваме, я разглеждат като потенциално неограничена по брой (даже и в случаите, когато популацията практически съвпада със самата извадка).

Данните съдържат в себе си определен брой наблюдавани **променливи величини (variables)**, които се подреждат в таблици. Колоните на таблицата съответстват на различните променливи, а редовете съответстват на статистическите единици. Резултатите от наблюдението се отразяват непосредствено в таблицата като числа или други подходящи символи, които обикновено се заместват чрез избрани за целта числови кодове (кодирането на нечислови данни предлага технически удобства при преобразуване на данните от един формат в друг). Например да разгледаме следната таблица от данни, която включва разнообразна информация, извлечена от представително обзорно изследване за населението на САЩ, заложено в съпровождащите данни **USGSS93** на пакета SPSS.

Таблица 1.1.

Идентификато P (ID)	ПОЛ (SEX)	РАСА (RACE)	ВЪЗРАСТ T (AGE)	ЖИВО T (LIFE)	SEI	ДОХОД (INCOME)	SIB S
1	1	1	43	2	36. 1	18	3
2	1	2	44	3	63. 5	18	2
3	2	1	43	3	69. 2	18	2
4	2	1	45	1	31. 4	22	4
5	2	1	78	3	75. 2	21	1
6	1	1	83	2	53.	13	2

					9		
7	2	1	55	2	38. 4	19	2
8	1	1	75	2	45. 8	10	3
9	1	1	31	3	64. 1	21	1
10	2	1	54	3	80. 9	9	1
11	2	1	29	1	82. 7	16	1
12	2	1	23	2	26. 4	18	0
...

Данните от файла *USGSS93* включват много повече променливи, някои от които ще бъдат използвани по-нататък в различни примери. Първата колона всъщност не представя истинска променлива, а служи само за идентификация на статистическите единици. Тук за идентификация се използва условен номер. В качеството на идентификатор може да се използва всякаква информация, която определя индивида по единствен начин с цел резултатите от груповия анализ да могат да бъдат отнесени обратно върху индивида, когато процедурите на анализ позволяват това.

Втората колона показва данните за променливата "пол", като кодът "1" е за мъже, а кодът "2" за жени. Третата колона показва данните за променливата "раса", като кодът "1" означава "бели", кодът "2" означава "черни", а кодът "3" – друга раса. Четвъртата колона показва данните за променливата "възраст", измерена в години. Петата колона показва данните за променливата "живот", като кодът "1" означава, че анкетираното лице определя ежедневието си като "потискащо", кодът "2" означава оценка "еднообразно", а кодът "3" означава "въълнуващо". Шестата колона съдържа данни за променливата "социално-икономически индекс", именувана със "SEI", измерена по специална методика, която няма да коментираме. Седмата колона съдържа данни за променливата "годишен доход на семейството", кодирана по следния начин:

- "1" – по-малко от \$1000; -- "2" – между \$1000 и \$2999;
- "3" – между \$3000 и \$3999; -- "4" – между \$4000 и \$4999;
- "5" – между \$5000 и \$5999; -- "6" – между \$6000 и \$6999;
- "7" – между \$7000 и \$7999; -- "8" – между \$8000 и \$9999;
- "9" – между \$10000 и \$12499; -- "10" – между \$12500 и \$14499;
- "11" – между \$15000 и \$17499; -- "12" – между \$17500 и \$19999;
- "13" – между \$20000 и \$22499; -- "14" – между \$22500 и \$24999;

- "15" – между \$25000 и \$29999; -- "16" – между \$30000 и \$34999;
- "17" – между \$35000 и \$39999; -- "18" – между \$40000 и \$49999;
- "19" – между \$50000 и \$59999; -- "20" – между \$60000 и \$74999;
- "21" – над \$75000.

Осмата колона съдържа данни за променливата "брой на братя и сестри", като цифрите показват въпросния брой.

В някои случаи данните от наблюденията могат да бъдат допълнително обработени преди да бъдат отразени в таблицата. Такава вторична обработка винаги е свързана със загуба на известно количество информация. Удачната вторична обработка обаче може да опрости в добър смисъл и онагледи в по-висока степен статистическия анализ.

2. Видове променливи. Всяка променлива се явява резултат от някакво (психологическо) измерване. Измерването представлява детерминирано правило за приписване на числов (или някакъв друг знаков) резултат към измерваните статистически единици. Таблица 1.1 показва, че променливите от интерес за изследователя могат да имат разнообразна природа. Според количествената информация, заложена в различните променливи, е приета класификацията на Стивънс, при която условно различаваме четири типа променливи и съответно четири типа скали за измерване.

Номинални променливи. Такива са променливите, при които имаме определен брой категории, които се различават чрез техните имена, като всеки индивид се отнася точно към една от тези категории. В таблица 1.1, номинални са променливите "пол" и "раса". Номинални са например променливите, които показват религиозна или политическа идентификация към някоя обособена група (категория). В този случай за променливите се казва, че са измерени по номинална скала (скала на наименованията). Номиналните променливи участват по естествен начин като фактори в дисперсионния анализ.

Ординални променливи. При тях резултатите от наблюденията могат да се сравняват. В таблица 1.1, ординални е величината "life". Особеното при ординалните променливи е, че за тях се предполага единствено възможността за сравнение между резултатите на различните индивиди. Например, ако индивид *A* има резултат 1 (потискащ живот), индивид *B* има резултат 2 (еднообразен живот), а индивид *C* има резултат 3 (вълнуващ живот), то можем да установим, че качеството на живот в този смисъл на индивид *C* превъзхожда това на индивид *B*, което от своя страна превъзхожда това на индивид *A*, но не можем да интерпретираме количествено това различие въпреки, че разликите в кодовете на резултатите са равни: $C - B = 3 - 2 = 1$ и $B - A = 2 - 1 = 1$. В този случай за променливите се казва, че са измерени по ординална скала. Ординални са формално погледнато и всички променливи, свързани с постижения по отделни учебни дисциплини, когато тези постижения са отчетени в петобалната система (слаб 2, среден 3, добър 4, много добър 5, отличен 6). Статистическите процедури, които се отнасят за чисто ординални променливи работят

фактически с ранговете на отделните статистически единици вместо с техните натурални стойности.

Интервални променливи. Интервали са онези ординални променливи, за които интервалите между резултатите могат да се интерпретират. Редица автори приемат, че суровите балове от различни тестови измервания в експерименталната психология са интервални променливи. Въпросът дали дадена величина е интервална или ординална в много случаи е дискуссионен, като в този случай критерият е наличието на достатъчна **изменчивост (вариабелност)** на величината (например поне седем различни стойности), но всъщност основният критерий е достатъчният по големина обем на извадката и формата на разпределението, за което ще стане дума по-нататък. В този случай за променливите се казва, че са измерени по интервална скала.

Интервалните скали имат основно значение за психологическите изследвания, понеже според общо приетото схващане резултатите от психологическите тестове се разглеждат като интервални променливи.

Абсолютни променливи. Абсолютни са тези интервални променливи, при които има нулева точка на измерването. В таблица 1.1, абсолютни са например възрастта, дохода и (по замисъл) социално-икономическия индекс. Други типични абсолютни променливи са различните физиологически показатели като тегло, ръст, кръвно налягане, време за психомоторни реакции и т.н. В този случай за променливите се казва, че са измерени по абсолютна скала или още по **скала на отношенията**. Абсолютните променливи се срещат систематично в медицинските и спортните статистически изследвания.

Горната класификация отчита преди всичко в каква степен върху резултатите от даден тип величина могат да се извършват различни математически операции като броене, сравняване или събиране. Категориите на чисто номиналните променливи подлежат само на броене, можем да изброим индивидите, принадлежащи на определена категория, например брой мъже и брой жени в извадката за величината "пол". Резултатите от наблюдението по ординална величина позволява извадката да се подреди според степента на притежаване на измерваната характеристика. Когато величината е интервална или абсолютна, могат да се пресмятат средни стойности и дисперсии.

От процедурно-изчислителна гледна точка, която на практика се явява водеща, SPSS разделя променливите на три вида **номинални, ординални и метрични (скали)**. В този смисъл метрични са променливите, определени по-горе като интервални и абсолютни. От гледна точка на количественото представяне, променливите се делят условно на **непрекъснати** и **дискретни**. Дискретни са тези променливи, които могат да приемат предварително известни фиксирани стойности. Непрекъснатите променливи могат да приемат стойности от цял интервал. В много от случаите разполагаме с достатъчно основания да приемем дадена ординална величина за метрична и като правило това винаги се прави по отношение на тестовите резултати от различни психологически скали. Някои по природа метрични променливи след наблюдение се отразяват като дискретни, такава е например величината "доход" от таблица 1.1. Метрични променливи са всичките физиологически показатели и изобщо всички променливи, за които разполагаме с инструмент за измерване. Номиналните и ординалните променливи се наричат понякога **категорийни**.