

ТЕРМИНОЛОГИЧЕН РЕЧНИК ПО СТАТИСТИКА

АЛФА-РИСК (Alpha-risk) – вероятността за отхвърляне на нулевата хипотеза, когато тя е вярна, т.е. вероятността за допускане на грешка от I-ви род

Z-РЕЗУЛТАТИ (Z-scores) – трансформирани резултати, които изразяват на колко стандартни отклонения под или над средната аритметична се намира даден конкретен резултат. Например, ако $z=-2$, това означава, че даден резултат е на две стандартни отклонения под средната аритметична.

АЛТЕРНАТИВНА ХИПОТЕЗА (Alternative hypothesis) – хипотеза, противоположна на нулевата. Означава се с H_1 . Известна е още като **експериментална хипотеза**, т.е. това е хипотезата, за която изследователят се стреми да получи подкрепа в статистическия анализ, отхвърляйки H_0 .

АТРИБУТ (attribute) – символ, асоцииран с нещо; свойство или качество, което дадено лице притежава или не притежава; свойство, с което се описва дадена характеристика на променлива величина чрез класифициране в категории, към които дадено лице принадлежи или не принадлежи. Оттук произлиза другото название на качествените променливи – **атрибутивни или категорийни променливи**.

БЕТА РАЗПРЕДЕЛЕНИЕ (Beta-distribution) - разпределение, което се използва за непрекъснати променливи величини, които са ограничени в рамките между 0 и 1.

БЕТА-РИСК (Beta-risk) – вероятността за приемане на нулевата хипотеза, когато тя е погрешна, т.е. вероятността за допускане на грешка от II-ри род.

БИНОМИНАЛНО РАЗПРЕДЕЛЕНИЕ (Binominal distribution) – разпределение, което представя вероятността за наблюдаване на X успешни опита при фиксиран брой (n) независими опити на Бернули; p представлява вероятността за успех при единичен опит

ВАЛИДНОСТ (Validity) – степента, в която даден тест измерва това, което е предназначен да измерва. Валидността може да бъде **външна и вътрешна**.

ВАЛИДНОСТ НА ОЦЕНКАТА (Validity of an estimate) – степента, до която дадена оценка съответства на параметъра, който тя оценява. Валидността на оценката зависи не от размера на извадката, а от репрезентативността на извадката. Следователно, тя зависи от вида или характера на извадката, от това как е подбрана извадката и от точността на информацията, от която е изчислена извадковата статистика и от самата изчислителна процедура.

ВАРИРАНЕ ИЛИ РАЗСЕЙВАНЕ (Variation or Dispersion) – свойство на данните в една извадка да приемат различни стойности. Когато разсейването е голямо, стойностите са силно разпръснати; когато то е малко, данните са компактно групирани. Най-често използваната мярка за варирането е **стандартното отклонение**.

ВЕРОЯТНОСТ (Probability) – продължителна относителна честота на дадено събитие при повторни опити при сходни условия; число между 0 и 1, което изразява доколко вероятно е да се случи дадено събитие.

ВЕРОЯТНОСТНО РАЗПРЕДЕЛЕНИЕ (Probability distribution) – математически модел, описващ вероятността за наблюдаване на различни стойности на дадена случайна променлива.

ВЗАИМНОИЗКЛЮЧАЩИ СЕ СЪБИТИЯ (Mutually exclusive events) - събития, които е невъзможно да се проявят едновременно или да бъдат налице по едно и също време.

ВЗАИМОДЕЙСТВИЕ (Interaction) – ситуация, в която ефектът на един фактор зависи от нивото на друг фактор. Взаимодействията се включват в статистически модели, когато факторите на действат по чисто допълващ се начин.

ВРЕМЕВИ СЕРИИ (time series) – поредица от данни, характеризиращи променливата X в еднакво разположени времеви точки (по години, месеци, дни, часове и т.н.)

ВЪНШНА ВАЛИДНОСТ (External validity) – степента, в която резултатите от дадено извадково проучване могат да бъдат генерализирани за популацията, от която е извлечена съответната извадка.

ВЪТРЕШНА ВАЛИДНОСТ (Internal validity) – в дадено проучване, вътрешната валидност се отнася до способността на изследователя да обясни различията между наблюдаваните групи или участници с влиянието на независимата променлива, т.е. с изучавания фактор.

ГОРЕН КВАРТИЛ (Upper quartile) – това е 75-я персентил, изчисляван чрез подреждане на данните от най-малкото към най-голямото и намиране на стойността, която заема 75-то място в поредицата от данни.

ГРЕШКА НА ПОДБОРА (Sampling error) – тъй като дадена извадка е по-малка от популацията, от която е извлечена, често има несъответствие между стойностите, получени за извадката и тези, които се отнасят за популацията. Това несъответствие се означава като грешка, свързана с подбора.

ГРЕШКА ОТ I РОД (Type I error) – неправилно отхвърляне на вярна нулева хипотеза. Вероятността за такава грешка се нарича **α -риск**. Прието е да се фиксира 5% вероятност за допускане на грешка от I род. Например, в клиничен опит върху ново лекарство H_0 може да гласи, че новото лекарство не е по-добро от сегашното или няма разлика между двете лекарства. Грешка от I род ще се допусне, ако се направи извод, че двете лекарства имат различен ефект, а фактически между тях няма разлика. Грешките от I род се считат за по-сериозни и важни за избягване, отколкото грешките от II род.

ГРЕШКА ОТ II РОД (Type II error) – неотхвърляне (приемане) на нулевата хипотеза, а тя фактически е невярна. Вероятността за такава грешка се нарича **β -риск**. Грешка от II род се допуска често поради малък размер на извадката. За всякаква серия от данни грешките от I и II род са обратно свързани: колкото в по-голям рискът за едната, толкова е по-малък рискът за другата.

ДАНИИ (Data) – информацията, събирана от даден изследовател.

ДВУСТРАНЕН ТЕСТ ЗА ЗНАЧИМОСТ (Two-sided or One-tailed test of significance) - тест за проверка на хипотези, при който не е определена посоката на различието между сравняваните статистики, т.е. не се интересуваме дали $X_1 < X_2$ или $X_1 > X_2$, а само от това дали има различие между тях. Тогава стойностите за отхвърляне на H_0 са разположени в двата края на вероятностното разпределение. Изборът между едностранен или двустранен тест се определя от целта на изследването.

ДИАГРАМА НА РАЗСЕЙВАНЕ (Scatter plot or Scattergram) – полезно средство за обобщаване на серия от данни за две свързани променливи, което обикновено се начертава преди изчисляване на коефициента на корелация или изглаждане на линията на регресия. Диаграмата на разсейване дава добра визуална представа за връзката между двете променливи и подпомага интерпретирането на корелационния коефициент и регресионния модел. Всяка единица на наблюдение отразява една точка на диаграмата. Начинът на разпределение на точките определя вида и силата на връзката между двете променливи.

ДИСКРЕТНА ПРОМЕНЛИВА (Discrete variable) – вид случайна променлива, която приема само определени стойности, изразени в цели числа.

ДИСПЕРСИЯ (Variance) – статистика за измерване на варирането (разсейването) на поредица от данни. Изчислява се като сумата от повдигнатите на квадрат отклонения на индивидуалните стойности от средната аритметична се разделя на броя на наблюденията. Означава се като s^2 и се измерва в същите мерни единици като оригиналните наблюдения – напр. cm^2 , kg^2 и т.н.

Колкото е по-голямо разсейването на данните, толкова е по-висока стойността на дисперсията. Ако всички стойности на данните са еднакви, дисперсията е равна на нула. Квадратният корен от дисперсията се нарича **стандартно отклонение**.

ДИХОТОМНА ХАРАКТЕРИСТИКА (Dichotomous attribute) – характеристика, класифицирана само в две категории, обикновено наличие или отсъствие на определено състояние (напр., здрав и болен, с подобрение и без подобрение и т.н.). Някои характеристики са присъщо дихотомни по своята природа (напр., мъж/жена, жив/мъртъв), но всички характеристики, независимо дали са дихотомни по природа, могат да бъдат превръщани в дихотомни чрез определяне и идентифициране на една подгрупа и поставяне на всички други наблюдения във втора подгрупа (напр., лица с висше образование и всички останали).

ДОВЕРИТЕЛЕН ИНТЕРВАЛ (Confidence interval) – доверителният интервал на дадена извадкова статистика представлява очаквания обхват, в границите на който при определено ниво на доверителност (гаранционна вероятност) ще се намира истинската стойност на неизвестния параметър за популацията. Построява се на основата на получени стойности на показателите в извадката и гаранционна вероятност 95% или 99%. Ширината на интервала дава известна представа за степента на несигурност на статистическия извод. Възприемането на по-висока гаранционна вероятност на извода е свързано с разширяване на доверителния интервал. Доверителните интервали са по-информативни от тестовете за проверка на хипотези (където решаваме дали нулевата хипотеза се приема или отхвърля), тъй като те предоставят диапазон от възможни стойности за неизвестния параметър в популацията.

ДОВЕРИТЕЛЕН ИНТЕРВАЛ ЗА ПРОПОРЦИЯ (Confidence interval for a proportion) – дава представа за обхвата от стойности, които има вероятност да приеме неизвестен параметър в популацията (означаван с π) на основата на изчислената пропорция (p) за извадка, извлечена от същата популация. Представя се като $\pi = p \pm t s_p$

ДОВЕРИТЕЛЕН ИНТЕРВАЛ НА СРЕДНА АРИТМЕТИЧНА (Confidence interval for a mean) – определя обхвата от стойности, в рамките на които се намира средната аритметична за популацията (μ) на базата на изчислената средна за извадката (\bar{x}) и възприетото ниво на доверителност. Представя се като $\mu = \bar{x} \pm t s_{\bar{x}}$. Изчислява се за различни нива на доверителност в зависимост от точността, с която се прави статистическото заключение. Например, при ниво на доверителност 95%, ние казваме, че сме 95% сигурни, че интервалът съдържа истинската средна аритметична за популацията.

ДОВЕРИТЕЛНИ ГРАНИЦИ (Confidence limits) – долната и горната граници на интервала при интервалното оценяване, т.е. стойностите, които определят обхвата на даден доверителен интервал. Наричат се доверителни граници, защото те се определят съгласно установено или общоприето ниво на доверителност или вероятност, че тези граници фактически ще включват оценявания популационен параметър. По такъв начин, 95% доверителни граници са стойности, между които с 95% сигурност ще се намира оценявания популационен параметър. В научните изследвания се работи с ниво на доверителност по-голямо от 95%. С повишаване на гаранционната вероятност доверителните граници се разширяват.

ДОЛЕН КВАРТИЛ (lower quartile) – 25-ят персентил, изчислен чрез подреждане на данните от най-малката към най-голямата стойност и намиране на стойността, която лежи на 25% от началото на вариационния ред.

ДЯСНО ИЗТЕГЛЕНО ИЛИ ПОЛОЖИТЕЛНО НАКЛОНЕНО РАЗПРЕДЕЛЕНИЕ (Positively skewed distribution) - когато графиката на разпределението показва издължена опашка надясно към по-високите стойности на абсцисата X .

ЕДИНИЦА НА ИЗСЛЕДВАНЕ (Unit of inquiry) – най-малката единица, върху която се събират данни.

ЕДНОСТРАНЕН ТЕСТ ЗА ЗНАЧИМОСТ (One-sided or One-tailed test of significance) - тест за проверка на хипотези, при който предварително е указана посоката на изследваното различие между извадковите статистики – напр., интересуваме се конкретно дали $X_1 > X_2$. Тогава стойностите за отхвърляне на H_0 са разположени изцяло в единия край на вероятностното разпределение. Изборът между едностранен или двустранен тест се определя от целта на изследването.

ЕКОЛОГИЧНА ВАЛИДНОСТ (Ecological validity) – степента, в която резултатите от дадено проучване могат да бъдат генерализирани за реалния заобикалящ свят.

ЕКСПЕРИМЕНТ (Experiment) – изследователски дизайн, включващ случайно разпределение на изследваните лица в групи и прилагане на различни интервенции към тези групи. Целта на експеримента е да позволи да се направи валидно заключение, че различията в резултатите за групите се дължат на различните интервенции.

ЕКСТРАПОЛАЦИЯ (Extrapolation) – метод за прогнозиране, опиращ се на метода на най-малките квадрати. Изхожда се от предположението, че дадено явление ще продължи да се развива по същия начин, както през наблюдавания период. Към стойността на Y се прибавя толкова пъти стойността

на коефициента на регресия b , за колкото времеви интервала напред искаме да прогнозираме.

ЕПИДЕМИОЛОГИЯ (Epidemiology) – изучаване на разпространението и детерминантите на заболяванията и други здравни събития в рамките на дадена общност и използване на резултатите за предотвратяване на тези явления.

ЗАВИСИМА ПРОМЕНЛИВА (Dependent variable) – променливата, която трябва да бъде измерена като резултат от даден експеримент. В идеалния случай, зависимата променлива трябва да бъде надеждна, чувствителна, лесна за измерване и разпределена по начин, който съответства на предположенията на даден статистически модел. Общоприето е зависимата променлива да се поставя на вертикалната ос Y , а независимата променлива – на хоризонталната ос X .

ЗАСЛЕПЯВАНЕ (Blinding) – сравняването при медицинските експерименти може да бъде изопачено, ако пациентът, експериментаторът или оценителят знаят какво лечение е приложено. За да се избегне това се прибегва до **единичен сляп опит** (само пациентът не знае) **или двоен сляп опит** (и пациентът, и експериментаторът и оценителят не знаят). Двойният сляп опит е най-приемливият вариант от научна гледна точка, но не винаги е възможен.

ЗНАКОВ ТЕСТ (Sign test) – непараметричен тест за проверка на хипотези, при който се сравняват медиани в уеднаквени извадки (напр., преди и след третиране). Не изисква нормално разпределение. Може да се прилага също и при ординални измерителни скали.

ИЗВАДКА (Sample) – подгрупа от дадена популация, чиито свойства се обобщават или трябва да бъдат генерализирани за популацията.

ИЗВАДКОВА ЕДИНИЦА (Sampling unit) – всяка избрана единица в процеса на формиране на дадена извадка (напр., лице, домакинство, регион). Тя не е непременно единицата на наблюдение или проучване.

ИЗВАДКОВА РАМКА (Sampling frame) – наборът от извадкови единици, от които трябва да се подбере дадена извадка (напр., списък с имената или местата, или други пунктове, които се използват като извадкови единици).

ИЗГЛАЖДАНЕ (Smoothing) - техники, които се използват за намаляване на случайните отклонения във времеви серии. Тук се отнасят **методите на укрупнените периоди, верижните средни, свободно изглаждане, метод на най-малките квадрати.**

ИЗМЕРВАНЕ (Measurement) – в статистическия анализ нивото на измерване определя по какъв начин да бъде анализирана дадена променлива.

Променливите могат да се класифицират в различни групи според начина на измерване: номинални, ординални, интервали и пропорционални.

ИНТЕРВАЛНА СКАЛА (Interval scale) – вид измерителна скала, която има следните свойства: стойностите са различни, подредени във възходящ ред, интервалите между точките върху скалата са еднакви, нулевата точка не е абсолютна (не означава отсъствие на количество).

ИНТЕРВАЛНА ЧЕСТОТА (Class frequency) – броят на наблюденията, които попадат във всеки клас; другото название е **абсолютна интервална честота**.

ИНТЕРВАЛНИ ГРАНИЦИ (Class limits) – стойностите на променливата, които разграничават всеки клас; долната и горната граница на класа.

ИНТЕРВАЛНИ ПРОМЕНЛИВИ (Interval variables) – това са количествени измервания, където разстоянието между стойностите има определен смисъл, но нямат истинска нулева точка.

ИНТЕРВАЛНО ОЦЕНЯВАНЕ (Interval estimation) – предоставяне на оценка за даден параметър на популацията от гледна точка на определен интервал или обхват от стойности, в рамките на които той вероятно се намира.

ИНТЕРКВАРТИЛЕН ОБХВАТ (Interquartile range) – разстоянието между първия и третия квартил. Използва се като мярка за вариране (разсейване); влияе се по-слабо от отклоняващи се стойности в сравнение със стандартното отклонение или обхвата на вариационния ред.

КАМБАНОВИДНА КРИВА (Bell shaped curve) – характерната форма на нормалното разпределение.

КАТЕГОРИЙНИ ДАННИ (Categorical data) – серия от данни, които могат да се сортират по категории – напр., по цвят, по пол и т.н. Всяка стойност или наблюдение трябва да принадлежи само към една категория и да не допуска припокриване на категориите.

КАЧЕСТВЕНА ПРОМЕНЛИВА (Qualitative variable) - променлива, за която се измерва дадено свойство или класификация – напр., пол или здравно състояние. Наричат се още **категорийни** променливи. Могат да бъдат **номинални и ординални**. Качествените променливи могат да бъдат кодирани цифрово, но присъдените кодове нямат истинско цифрово значение, а само облекчават процеса на статистическа обработка на данните.

КВАНТИЛИ (Quantile) – серия от “пресечни точки”, които разделят дадена извадка на групи, съдържащи еднакъв брой наблюдения. Примери за кван-

тили са **квартили** (на 4 равни части), **квинтили** (на 5 равни части), **персентили** (на 100 равни части).

КВАРТИЛИ (Quartiles) – статистики, които разделят наблюденията в една извадка от числа в 4 интервала, всеки от които съдържа 25% от данните. Има 3 квартаила. Първият, средният и третият квартаили се изчисляват чрез подреждане на данните от най-малката към най-голямата стойности и след това се намират стойностите, под които попадат 25%, 50% и 75% от данните. Средният квартал обикновено се нарича **медиана**. Квартилите са специални случаи на **персентили**. Първият, средният и третият квартаили съответстват на 25-я, 50-я и 75-я персентил.

КВИНТИЛИ (Quintile) – стойности, които делят дадена извадка на 5 части, съдържащи еднакъв брой наблюдения.

КЛАС (Class) – един от интервалите, на които се разделя целият обхват на променливата величина (напр., всеки от интервалите 0-9, 10-19, 20-29 и т.н. представлява отделен клас).

КЛАСИЧЕСКИ МЕТОДИ (Classical methods) – към тях се отнасят оценката на доверителните интервали и използването на тестове за проверка на хипотези, които се прилагат широко в течение на десетилетия, в противоположност на по-новите непараметрични методи, които не изискват нормално разпределение на данните.

КЛИНИЧНА ЗНАЧИМОСТ (Clinical significance) – степента, в която находките в дадено проучване са значими за клиничната практика.

КОЕФИЦИЕНТ НА ВАРИАЦИЯ (Coefficient of variation (CV)) - съотношение на стандартното отклонение и средната величина, изразено в проценти; понякога се нарича **относително стандартно отклонение**. Величината му не зависи от скалата на измерване и мерните единици на променливата величина. Използва се често в естествените науки при сравнение на вариабилността на променливи, които имат различни измерителни скали.

КОЕФИЦИЕНТ НА ДЕТЕРМИНАЦИЯ (Coefficient of determination) – квадратът на коефициента на корелация (r^2). Показва каква част от промените в зависимата променлива се дължат на влиянието на изучаваната независима променлива.

КОЕФИЦИЕНТ НА КОНТИНГЕНЦИЯ (Contingency coefficient) – изчислява се при таблици за контингенция на основата на хи-квадрат и измерва степента на връзка между променливите величини в редовете и колоните на таблицата; стойността му варира между 0 и 1, но по принцип не може да достигне 1.

КОЕФИЦИЕНТ НА КОРЕЛАЦИЯ (Correlation coefficient) – статистика, предназначена да измери големината и посоката на линейната връзка между две променливи. Означава се с "r". Абсолютната стойност на "r" показва силата на връзката и варира между + 1 and - 1, като -1 or +1 означава пълна линейна зависимост, а $r = 0$ показва липса на връзка. Знакът на коефициента на корелация показва посоката на връзката. Ако r е с положителен знак – налице е права (положителна) връзка; ако r е с отрицателен знак – връзката е обратна (отрицателна). Съществуват различни коефициенти на корелация, изборът на които зависи от вида и начина на представяне на изучаваните променливи величини.

КОЕФИЦИЕНТ НА КОРЕЛАЦИЯ НА ПИРСОН (Pearson's correlation coefficient) – измерва линейната връзка между две количествени променливи с нормално разпределение, които са представени върху интервална или пропорционална скала (напр., между ръста в см и теглото в кг). Варира между -1 за пълна отрицателна корелация и +1 за пълна положителна корелация. Ако формата на разпределението е неизвестна или асиметрична, по-подходящ е **непараметричният рангов коефициент на корелация на Спирман**.

КОЕФИЦИЕНТ НА КОРЕЛАЦИЯ НА СПИРМАН (Spearman's Rank Correlation Coefficient) – Изчислява се обикновено в случаите, когато е неподходящо, неикономично или дори невъзможно да се представят истинските стойности на променливите, а са присвоени рангови номера на отделните случаи на всяка променлива. Той е по-добър индикатор за наличие на връзка между две променливи, когато връзката е нелинейна.

КОЕФИЦИЕНТ НА КРАМЕР (Cramer's V) – изчислява се на базата на хи-квадрат в таблици за контингенция; измерва степента на връзката между стойностите на променливите в редовете и колоните на таблицата върху скала от 0 до 1, като за разлика от коефициента на контингенция може да достигне стойност 1.

КОЕФИЦИЕНТ НА РЕГРЕСИЯ (Regression coefficient) – статистика от регресионния анализ, която служи за оценка на неизвестни параметри на регресионния модел и показват с колко се променя зависимата променлива Y при промяна на независимата променлива X с единица. В уравнението на линията на регресия $Y = A + bX$, коефициентът на регресия е **b**. Неговата числена стойност се определя от наклона на регресионната линия, а алгебричният знак описва връзката между X и Y. Ако X нараства и Y нараства, наклонът е положителен и коефициентът **b** е със знак плюс. Обратно - ако X нараства, а Y намалява наклонът е отрицателен и **b** е със знак минус.

КОЕФИЦИЕНТИ ЗА ЧЕСТОТА (интензивни показатели) – описателни статистики за качествени променливи; измерват честотата на дадено явление в определена среда – среда е населението, а най-честите явления са заболявания, раждания, умирация и др. Изразяват се на 1000, 10000 или 100000 души. За по-рядко срещани явления се използва по-голям множител и обратно.

КОЛИЧЕСТВЕНА ПРОМЕНЛИВА (Quantitative variable) – променлива, при която се измерва числена стойност, даваща представа за определено количество. Това са истински цифрови променливи. Например, изчисляването на средна стойност за измерено тегло в дадена извадка има определен смисъл и значение. Количествените променливи могат да бъдат **дискретни (прекъснати)** и **непрекъснати (продължителни)**.

КОНТРОЛНА ГРУПА (Control group) – в даден експеримент контролна група обикновено е група, която не е подложена на изучаваното въздействие и тя се сравнява с експерименталната група, за да се изучат ефектите от интервенцията.

КОРЕЛАЦИЯ (Correlation) – мярка за линейна връзка между две променливи величини. Измерва силата и посоката на връзката между променливите. Ако двете променливи се променят в една и съща посока, това е позитивна корелация и обратно – ако те се променят в различна посока, корелацията е негативна (отрицателна). Най-често използваният коефициент за измерване на връзката е коефициента на корелация на Пирсон.

КРИТЕРИЙ t (t test) – критерий за проверка на хипотези, опиращ се на t -разпределението на Стюдент. Използва се често за проверка на хипотези относно различията между средни аритметични или коефициенти в линеен регресионен модел.

КРИТЕРИЙ t ПРИ ЗАВИСИМИ ИЗВАДКИ (Paired t -test) – използва се за определяне на статистическата значимост на различията между средните стойности на едни и същи измервания, направени при две различни състояния – напр., преди и след прилагане на дадено лечение. Двете измервания се правят при всяка единица в извадката. Този вид сравняване се нарича още **сравняване от типа “преди-после” (before-after)**.

КРИТЕРИЙ НА КРУСКАЛ-УОЛИС (Kruskal-Wallis test) – критерий, който сравнява медианите на множествени извадки, използвайки непараметрични техники. Най-напред данните се обединяват и се ранжират от най-малката към най-голямата стойност. След това се изчислява средният ранг на стойностите във всяка група и се определя дали има съществени разлики между средните рангове чрез таблица и P стойността на критерия.

КРИТЕРИЙ НА ФИШЕР (Fisher's test) – критерий от дисперсионния анализ. Използва се при сравняване на повече от две извадки. Изчислението му се основа на сравняване на вътрегруповата и междугруповата дисперсии.

КРИТИЧНА ОБЛАСТ (Critical Region) – нарича се още област на отхвърляне; за даден тест за проверка на хипотези критична област е серия от стойности на критерия, при които нулевата хипотеза се отхвърля; т.е. извадковото пространство се разделя на две области; критична област, която води до отхвърляне на H_0 и друга област, в която H_0 не се отхвърля.

КРИТИЧНА СТОЙНОСТ(И) (Critical Value(s)) – това е стойността на дадена статистика (извлечена от съответни таблици), която изчислената стойност за даден резултат трябва да превишава, за да се достигне статистическа значимост, т.е. това е прагът, с който се сравнява стойността на изчислен критерий от извадката, за да се определи дали да се отхвърли или не H_0 . Критичната стойност за всеки критерий зависи от нивото на значимост, на което се провежда даден тест и от това дали тестът е **едностранен или двустранен**.

КРОС-ТАБУЛАЦИЯ (Cross-tabulation) – честотна таблица, включваща поне две променливи, които са крос-класифицирани (табулирани една срещу друга).

КРЪГОВА ДИАГРАМА (Pie chart) – графично представяне на категорични данни чрез сектори от кръг, всеки от които отразява конкретна категория; площите на секторите са пропорционални на броя на случаите в съответните категории.

КУМУЛАТИВНА ВЕРОЯТНОСТ (Cumulative probability) – вероятността, че дадена случайна променлива ще бъде по-малка или равна на определена стойност. Например, кумулативната вероятност за медианата е равна на 0.5.

КУМУЛАТИВНА ОТНОСИТЕЛНА ЧЕСТОТА (Cumulative relative frequency) – броят на наблюденията, попадащи в даден клас на честотната таблица плюс всички наблюдения, попадащи в предходните класове, разделен на общия брой на наблюденията.

КУМУЛАТИВНА ЧЕСТОТА (Cumulative frequency) – брой на наблюденията, попадащи в даден клас на честотната таблица плюс всички наблюдения, попадащи в предходните.

ЛИНЕЕН МОДЕЛ (Linear model) – модел, който при обикновена регресия приема формата $Y = a + bX$

ЛИНИЯ НА РЕГРЕСИЯ (Regression line) – това е линия, прекарана през точките върху диаграмата на разсейване, за да обобщи връзката между изучаваните променливи. Когато линията е наклонена надолу (от горния ляв край към долния десен), това означава негативна или обратна връзка между променливите; когато наклонът е нагоре (от долния ляв край към горния десен) – връзката е положителна или права.

ЛЪЖЛИВА КОРЕЛАЦИЯ (Spurious correlation) – връзка между две променливи, която може да бъде артефакт, случайна, непредвидена, лъжлива или дължаща се на всякакви видове непричинни връзки, в резултат на случайност или систематична грешка.

ЛЯВО ИЗТЕГЛЕНО ИЛИ ОТРИЦАТЕЛНО НАКЛОНЕНО РАЗПРЕДЕЛЕНИЕ (Negatively skewed distribution) - когато графиката на разпределението показва издължена опашка наляво към по-ниските стойности на абсцисата X.

МЕДИАНА (Median) – статистика, която измерва центъра на даден масив от данни; определя се чрез стойността, която разделя данните на две равни половини, т.е. 50% от стойностите се намират под медианата и 50% над нея. За да се изчисли медианата, данните трябва да се сортират от най-малката към най-голямата стойност. При нечетен брой наблюдения медианата е равна на стойността в средата на масива. При четен брой наблюдения тя е равна на полусумата от двете стойности в средата на вариационния ред. При симетрично разпределение, каквото е нормалното разпределение, медианата е равна на средната аритметична. При дясно изтеглено разпределение медианата е по-малка от средната, а при лявоизтеглено разпределение тя е по-голяма от средната.

МЕТОД НА НАЙ-МАЛКИТЕ КВАДРАТИ (Least squares) – метод за изравняване на динамични редове или за нагаждане на специфичен модел към наблюдавани данни. Това е най-често използваният метод за прекарване на права линия през серия от точки върху диаграмата на разсейване.

МЕТОД НА ПОДБОР (Sampling method) – метод, чрез който дадена извадка се извлича от дадена популация. Най-общо, има два подхода: **случаен (непреднамерен) подбор**, при който всеки случай в популацията има равен шанс за избор и **неслучаен (преднамерен) подбор**, при който случаите имат различен шанс да бъдат избрани.

МНОЖЕСТВЕНА РЕГРЕСИЯ (Multiple regression) – регресионен модел, в който стойността на резултативната променлива Y е функция от стойностите на повече от една независими променливи.

МОДА (Mode) – най-често срещаната стойност в дадена серия наблюдения; понякога се използва като алтернатива на средната аритметична и на ме-

дианата като мярка за централна тенденция, особено когато данните са измерват върху номинална скала. При групирани данни модата се определя като срединна точка на интервала, съдържащ най-високата честота. При някои разпределения може да се наблюдава повече от една мода (**бимодални и полимодални разпределения**).

НАБЛЮДАВАНИ ЧЕСТОТИ (Observed frequencies) - това са честотите във всяка клетка на таблиците за контингенция, които действително са получени от наблюдаваната случайна извадка. Най-често се използват при изчисляване на хи-квадрат, когато наблюдаваните честоти се сравняват с очакваните, за да се установи има ли съществено различие между тях.

НАБЛЮДЕНИЕ (Observation) – наблюденията се отнасят до повтарящите се стойности на данните на променливата величина. Редовете на дадена колона представляват наблюдения.

НАДЕЖДНОСТ (Reliability) – отнася се до присъщото изпълнение на дадена процедура и изразява степента, в която резултатът от даден тест или измерване може да се възпроизведе. Надеждна процедура е тази, която дава съвместими резултати при прилагане повече от един път по отношение на същата единица на наблюдение при сходни условия. Една променлива величина или статистика са надеждни, ако при повторни експерименти стойностите им могат да бъдат измерени по същия начин.

НАСОЧЕНА ХИПОТЕЗА (Directional hypothesis) – хипотеза, която твърди, че различията между сравняваните групи ще се проявят в конкретно направление (посока) – напр. пушачите умират по-млади от непушачите. За проверка на такава хипотеза се прилага **едностранен тест**.

НЕЗАВИСИМА ПРОМЕНЛИВА (Independent variable) – фактор, чийто ефект трябва да бъде изучен в даден експеримент. Наричат се независими, защото експериментаторът може свободно да избира техните нива. Съществуват 2 вида независими променливи, които се третират често по различен начин в статистическите анализи: количествени и качествени променливи.

НЕЗАВИСИМИ ИЗВАДКИ (Independent samples) – извадки, които са избрани от една и съща или от различни популации и които нямат влияние една на друга.

НЕЗАВИСИМИ СЪБИТИЯ (Independent events) – две събития са независими, когато наличието или отсъствието на едно от тях не променя шансовете за проява на другото; т.е. събитията нямат влияние едно върху друго. Ако две събития са независими, те не могат да бъдат взаимно изключващи се и обратно.

НЕЗАВИСИМО НАБЛЮДЕНИЕ (Independent observation) – когато стойността на едно наблюдение не влияе или не променя стойността на друго. Болшинството статистически процедури приемат, че наличните данни представляват случайна извадка от независими наблюдения.

НЕЛИНЕЙНА ЗАВИСИМОСТ (Non-linear relationship) – когато формата на модела, описващ Y по отношение на X , не е права линия.

НЕНАСОЧЕНА ХИПОТЕЗА (Non-directional hypothesis) – хипотеза, която поддържа, че има различията между сравняваните групи, но не определя посоката на тези различия - напр. пушачите и непушачите имат различна средна продължителност на предстоящия живот. За проверка на такава хипотеза се прилага **двустранен тест**.

НЕПАРАМЕТРИЧНИ МЕТОДИ (Non-parametric methods) – статистически тестове, които са подходящи за анализ на ординални или номинални данни. Това са методи за проверка на хипотези, които не изискват предположения за вида на разпределението на данните в извадките и в популациите, докато болшинството класически методи изискват нормално разпределение. Много непараметрични методи се опират повече на рангово подреждане на данните, а не техните истински стойности.

НЕПРЕКЪСНАТА ПРОМЕНЛИВА (Continuous variable) – вид случайна променлива величина, която потенциално може да приеме неограничен брой стойности във всеки интервал.

НЕСИМЕТРИЧНОСТ, НАКЛОН (Skewness) – характеристика на формата на честотните разпределения, която измерва липсата на симетрия. Измерва се обикновено чрез коефициент за несиметричност, който е равен на нула за симетричните разпределения, по-голям от нула за положително наклонените разпределения и по-малък от нула за отрицателно наклонените разпределения.

НИВО НА ДОВЕРИТЕЛНОСТ (Confidence level) – това е стойността на вероятността $(1-\alpha)$, свързана с доверителния интервал. Изразява се често в проценти – например, $(1-0.05) = 0.95$, т.е. 95% ниво на доверителност. Нарича се още **гаранционна вероятност** на статистическите заключения.

НИВО НА ЗНАЧИМОСТ (Significance level) – вероятността за напълно случайно различие. При даден статистически тест за проверка на хипотези нивото на значимост представлява фиксираната вероятност за погрешно отхвърляне на H_0 , когато тя фактически е вярна; това е вероятността за грешка от I род, т.е. най-малкото α -ниво, при което H_0 би била отхвърлена. Обикновено се избира малка стойност на P - по-малка от 0.05 (5%).

НОМИНАЛНА СКАЛА (Nominal scale) – вид измерителна скала, в която величините представляват отделни категории.

НОМИНАЛНИ ПРОМЕНЛИВИ (Nominal variables) – това са категорични променливи (напр. пол), при която няма никакво естествено подреждане на стойностите, които те могат да приемат.

НОРМАЛНО РАЗПРЕДЕЛЕНИЕ (Normal distribution) – непрекъснато вероятностно разпределение, което се използва за характеристика на разнообразни количествени променливи. То е симетрично, камбановидно и се определя напълно от неговата средна аритметична и стандартното отклонение. Известно е още като **Гаусово разпределение** по името на своя откривател Фредерик Гаус.

НУЛЕВА ХИПОТЕЗА (Null hypothesis) – означава се с H_0 и гласи, че не съществува различие между сравняваните характеристики или ако има някакво различие, то е резултат на случайност, а не на влияние на закономерни, значими фактори. В процеса на статистическия анализ се проверява именно нулевата хипотеза и крайният извод се прави винаги от гледна точка на H_0 – или се отхвърля H_0 в полза на алтернативната хипотеза H_1 , или се приема H_0 . Ако не отхвърлим H_0 , това не означава че тя е вярна, а само предполага, че няма достатъчно доказателства срещу H_0 в полза на H_1 ; отхвърлянето на H_0 пък предполага, че H_1 може да е вярна.

ОБИКНОВЕНА (ПРОСТА) ЛИНЕЙНА РЕГРЕСИЯ (Simple linear regression) - статистически модел, свързващ единична зависима променлива с единична независима променлива чрез метода на най-малките квадрати с цел да се открие линейна връзка между тях.

ОБХВАТ ИЛИ РАЗМАХ (Range) – мярка за разсейването на данните, разлика между максималната и минималната измерена стойност на някаква количествена променлива. Влияе се силно от наличие на рязко отклоняващи се стойности.

ОПИСАТЕЛНА СТАТИСТИКА (descriptive statistics) – раздел от статистиката, предназначен за описание на характеристиките на дадена извадка; включва числови и графични обобщения на данните от извадката.

ОПИТ (Trial) – експеримент, в който резултатите не могат да бъдат предказани предварително.

ОПИТ НА БЕРНУЛИ (Bernuli trial) – експеримент с два възможни резултата (успех или провал, добро или лошо, ези или тура). Вероятността за получаване на определен резултат се означава с **P**.

ОРДИНАЛНА СКАЛА (Ordinal scale) – вид измерителна скала, в която стойностите са разграничени и подредени по определена логика, най-често по рангове.

ОРДИНАЛНИ ПРОМЕНЛИВИ (Ordinal variables) – те се състоят от категории, които могат да бъдат подредени по ранг – от най-ниско към най-високо, от най-лошо към най-добро и т.н.; т.е. при този вид променлива има определено естествено подреждане на стойностите, които тя може да приеме. Например, отговорите на даден въпрос могат да бъдат “силно неодобрявам”, “неодобрявам”, “одобрявам”, “силно одобрявам”.

ОСТАТЪК (Residual) – грешката, която остава след прилагане на даден статистически модел към данните от извадка. При регресионния анализ остатъци са разликите между наблюдаваните стойности и прогнозираните стойности чрез регресионния модел, т.е. това са отклоненията от линията на регресия.

ОТНОСИТЕЛНА ЧЕСТОТА (Relative frequency) – брой случаи в дадена извадка или брой стойности, попадащи в рамките на даден интервал, изразен като пропорция от общия брой наблюдения. Графично се представя чрез стълбова диаграма или хистограма.

ОТНОСИТЕЛНО СТАНДАРТНО ОТКЛОНЕНИЕ (Relative standard deviation) – другото име на **коефициента на вариация**.

ОТРИЦАТЕЛНА КОРЕЛАЦИЯ (Negative correlation) - корелационна връзка, при която промените в двете променливи (независима и зависима) имат различна насоченост; нарастването на стойностите на независимата променлива е съпроводено с намаляване на стойностите на зависимата променлива.

ОЦЕНКА (Estimation) – процес на използване на дадена извадка за оценка на характеристиките на дадена популация. Например, използвайки статистиката от извадката (например, средната величина) да се извлече най-добра оценка за параметъра на популацията (средната в популацията).

ОЧАКВАНИ СТОЙНОСТИ (Predicted values) – стойностите, прогнозирани на основата на даден статистически модел, който съответства на данните. Например, при изравняване на една линия по отношение на стойностите на X и Y, прогнозираните стойности се отнасят до разположението на линията при дадена стойност на X.

ОЧАКВАНИ ЧЕСТОТИ (Expected Frequencies) – това са честотите в таблиците за контингенция, които биха се получили, ако е вярна нулевата хи-

потеза, т.е. ако няма връзка между изучаваните променливи. Най-често се използват при изчисляване на хи-квадрат.

ПАРАМЕТРИЧНИ МЕТОДИ (Parametric tests) – статистически методи, които са подходящи за интервални или пропорционални скали.

ПАРАМЕТЪР (Parameter) – обобщаваща характеристика на дадена популация, чиято оценка се опира на извадковите статистики.

ПЕРСЕНТИЛИ (Percentiles) – стойности, които разделят вариационни ред на 100 равни части. Първият (долният) квартил, медианата и третият (горният) квартил съответстват на 25-я, 50-я и 75-я перцентил.

ПЛАЦЕБО (Placebo) – неактивно лечение или процедура (буквално означава “правя нищо”); феномен, при който някаква безсмислена интервенция все пак предизвиква подобрене в състоянието на пациента или във възприемането на неговото състояние. Плацебо-ефектът се приписва на очакванията на пациента, че лечението ще има ефект.

ПОЛОЖИТЕЛНА КОРЕЛАЦИЯ (Positive correlation) – корелационна връзка, при която промените в двете променливи (независима и зависима) имат еднаква насоченост; нарастването на стойностите на независимата променлива е съпроводено с нарастване на стойностите на зависимата променлива.

ПОПУЛАЦИОННА ВАЛИДНОСТ (Population validity) – степента, в която дадена извадка отразява характеристиките на популацията, от която е извлечена.

ПОПУЛАЦИЯ (Population) – всички възможни лица, институции, предмети, наблюдения, за които искаме да направим заключение. Обикновено е невъзможно да бъдат обхванати всички елементи на популацията и проучванията се опират на наблюдение на извадки. Ако извадката е репрезентативна, резултатите от проучването могат да бъдат генерализирани (обобщени) за популацията, от която тази извадка е подбрана.

ПРАВИЛО НА НОРМАЛНОТО РАЗПРЕДЕЛЕНИЕ (Normal distribution rule) - за данни, които имат нормално разпределение, приблизително 68% от всички стойности попадат в пределите на 1 стандартно отклонение от средната аритметична, 95% - в рамките на 2 стандартни отклонения и 99,7% - в рамките на три стандартни отклонения от средната за извадката.

ПРАВИЛО НА СУМИРАНЕ НА ВЕРОЯТНОСТИТЕ (Addition Rule) – определя вероятността на явление А или явление В, или на двете явления заедно. Отнася се за взаимноизключващи се явления. $P(A \cup B) = P(A) + P(B)$

ПРАВИЛО НА УМНОЖЕНИЕ НА ВЕРОЯТНОСТИТЕ (Multiplication Rule) – определя вероятността за проява и на двете явления А и В едновременно. За независими явления правилото има следния вид: $P(A \cap B) = P(A) \cdot P(B)$

ПРАВИЛО НА ЧЕБИШЕВ (Chebishev rule) - за данни, които не следват нормално разпределение, най-малко 75% от всички стойности попадат в пределите на 2 стандартни отклонения и 89% - в рамките на три стандартни отклонения от средната за извадката.

ПРЕСЕЧНА ТОЧКА (Intercept) – постоянен термин в уравнението на линията на регресия. Ако уравнението на линията се представя като $Y = A + bX$, то пресечната точка е А. Това е точката върху линията на регресия, където независимата променлива X е равна на 0.

ПРИЧИННОСТ (Causality) – за едно събитие или фактор (А) се доказва, че е причина за друго събитие (В), ако А предхожда по време проявлението на В.

ПРОМЕНЛИВА ВЕЛИЧИНА (Variable) – всяко свойство, качество, феномен или явление, които могат да имат различни стойности. Променливите биват **качествени и количествени**. Характерът на променливите величини и скалата на измерването им диктуват какви статистически процедури могат да бъдат приложени за обработка и анализ на данните.

ПРОПОРЦИИ (Proportions) – отношение на една величина към друга, изразено като част от цяло (най-често в проценти). Наричат се още екстензивни показатели, структурни показатели, относителни дялове.

ПРОПОРЦИОНАЛНИ ПРОМЕНЛИВИ (Ratio variables) – вид променлива, за която разстоянието между стойностите има определен смисъл и която има истинска нулева точка. Например, теглото се измерва върху такава скала. За такава променлива може да се каже, че една стойност е два пъти по-голяма от друга.

ПРОСТ СЛУЧАЕН ПОДБОР или проста случайна извадка (Simple random sampling) – основен метод за подбор на извадки от популации, при което всеки индивид се избира напълно случайно и има еднакъв шанс да бъде включен в извадката. Подходящ е при хомогенни популации.

ПРОСТО СЪБИТИЕ (Simple event) – събитие, което не може да бъде раздробено на каквито и да е други компоненти.

P СТОЙНОСТ (P value) - р-стойността показва доколко изчислената статистика за даден тест се различава от това, което се очаква, ако е вярна нулевата хипотеза. Малка стойност на Р показва, че нулевата хипотеза следва да бъде отхвърлена при всяко ниво на значимост, което е над изчислената

стойност. Например, ако $P = 0.02$, бихме отхвърлили нулевата хипотеза при ниво на значимост 5%, но не бихме я отхвърлили при ниво на значимост 1%.

РАЗМЕР НА ЕФЕКТА (Effect size) – количеството промяна, предизвикана от дадена интервенция, особено при експериментално проучване.

РАЗПРЕДЕЛЕНИЕ (Distribution) – вероятностна функция, която описва относителната честота на проявление на стойностите на данните в дадена извадка от популацията. Разпределенията могат да бъдат непрекъснати или дискретни в зависимост от характера на променливите величини.

РАЗПРЕДЕЛЕНИЕ НА СТЮДЕНТ (Student's t distribution) – вероятностно разпределение, което е много сходно по форма със **стандартното нормално разпределение**. Средната аритметична на t-разпределението на Стюdent е винаги разна на нула, докато стандартното отклонение е обикновено малко по-голямо от 1. За пълна характеристика на разпределението е необходим само един параметър, наречен **степени на свобода**. Стойностите на t-разпределението на Стюdent често се използват за формиране на доверителните интервали на средната аритметична за популацията, за проверка на значимостта на различието между две извадкови средни или за проверка на значимостта на коефициентите в регресионен модел.

РАНГОВА КОРЕЛАЦИЯ (Rank correlation) – число между -1 и +1, което измерва силата на връзката между две променливи. За разлика от обикновения коефициент на корелация, който се изчислява на основата на истинските стойности на данните, ранговата корелация се опира на ранговете на данните в рамките на всяка променлива величина.

РАНГОВЕ (Ranks) – стойности, получени след подреждане на наблюденията в дадена извадка от най-малката към най-голямата. На най-малката стойност се присвоява ранг 1, на следващата – ранг 2 и т.н. Ако има няколко еднакви стойности, техните рангове се сумират и се присвоява средният ранг. Ранговете се използват при много непараметрични методи, когато няма основание да се допусне, че извадката е от нормално разпределение.

РАНДОМИЗИРАНЕ (Randomisation) – процес, чрез който експерименталните единици се отнасят към различните третирани групи на принципа на случайност, а не на основата на някакъв субективен процес или преднамерен подход. Същността на рандомизацията е да се формират групи за проучване, които са сравними по отношение на всички фактори, освен действителното проучвано въздействие.

РЕГРЕСИОНЕН АНАЛИЗ (Regression analysis) – при дадени данни за зависима променлива Y и независима променлива X , регресионният анализ

включва намиране на най-добрият математически модел, който описва Y като функция от X или прогнозира Y на основата на X .

РЕГРЕСИОНЕН МОДЕЛ (regression model) – статистически модел, свързващ една зависима променлива с две или повече независими променливи.

РЕГРЕСИОННО УРАВНЕНИЕ (Regression equation) – алгебричен израз на връзката между две или повече променливи. Показва характера на връзката и степента, в която може да се предвидят някои променливи на основата на други известни такива. Уравнението на линейна регресия се изразява като $Y = a + bX + e$, където “ Y ” е зависимата променлива, “ a ” е пресечната точка (intercept), “ b ” е наклонът или коефициентът на регресия, “ X ” е независимата променлива и “ e ” е грешката. Уравнението представя средната величина на очакваната промяна в Y при определена промяна в X . Регресионното уравнение често се представя върху **диаграма на разсейване** чрез **линия на регресия**.

РЕЗОЛЮЦИЯ (Resolution) – броят на точките или “пикселите” върху екран или принтер. Качеството на графичните изображения зависи от резолюцията на изходното средство.

РЕПРЕЗЕНТАТИВНА ГРЕШКА ИЛИ ГРЕШКА НА ПОДБОРА (Sampling error) – разлика, която се проявява напълно случайно, между стойността на извадкова статистика и съответния параметър в популацията. Репрезентативната грешка не може да бъде избягната или напълно елиминирана и трябва винаги да се отчита при извличане на заключения за популацията на базата на извадкови статистики. Тя може да бъде намалена чрез увеличаване размера на извадката и използвана на по-подходящ метод за подбор на извадки.

РЕПРЕЗЕНТАТИВНА ИЗВАДКА (Representative sample) – извадка, която точно отразява характеристиките на популацията, от която е извлечена.

РЯЗКО ОТКЛОНЯВАЩА СЕ СТОЙНОСТ (Outlier) - стойност, която е необичайна по отношение на групата данни, в които се намира. Това може да бъде единична изолирана стойност, силно отдалечена от останалите стойности на променливата или стойност, която не се придържа към общия модел на останалите. Болшинството класически статистически методи се влияят чувствително от такива отклоняващи се стойности, докато непараметричните методи са по-нечувствителни към такива стойности.

СИЛА НА СТАТИСТИЧЕСКИ ТЕСТ (Power) – вероятността за отхвърляне на H_0 , когато тя наистина е невярна; силата на теста измерва вероятността да бъде взето правилно решение, т.е. правилно определяне на даден ефект, когато наистина има такъв. Добрите тестове (критерии) имат голяма

сила. С други думи, силата на теста е вероятността за недопускане на грешка от II род. Изчислява се като $1-\beta$. Колкото е по-близко до 1, толкова тестът е по-силен, по-мощен.

СИМЕТРИЧНО РАЗПРЕДЕЛЕНИЕ (Symmetric distribution) – разпределение, което изглежда идентично отляво и отдясно по отношение на неговата средна аритметична. За такова разпределение средната аритметична, модата и медианата имат еднакви стойности. Нормалното разпределение е най-чест пример за симетрично разпределение.

СИСТЕМАТИЧНА ИЗВАДКА (Systematic sample) – вид извадка, при която се подбира всеки 5-ти, 10-ти и т.н. случай от популацията. Нужен е списък на всички лица в популацията, Определя се начална стъпка на подбора по лотариен метод. Нарича се още **механична или пропорционална извадка**.

СКАЛА ЗА ИЗМЕРВАНЕ НА ВЕРОЯТНОСТТА (Scale of probability measurement) – вероятността може да бъде измерена върху непрекъснатата скала от стойности между 0 и 1 (включително). За събитие, което е невъзможно, се казва, че има вероятност 0, то никога не се случва. Събитие, което винаги се проявява, има вероятност равна на 1. Събитие с вероятност по-голяма от 0.5 е по-вероятно да се случи, отколкото да не се прояви. Обозначението $P(A)$ представя вероятността за проява на събитие A .

СЛОЖНО СЪБИТИЕ (Compound event) – състои се поне от две събития

СЛУЧАЕН ПОДБОР (Random sampling) – метод за подбор на извадки, при който всички елементи на популацията имат равен шанс да попаднат в извадката. По такъв начин намалява вероятността за систематична грешка.

СЛУЧАЙ (Case) – един случай се състои от измервания на променливи величини при отделен индивид или експериментална единица. Например, може да се извършат множество измервания при едно и също лице за променливи като тегло, ръст, пол и т.н. Измерванията на всички променливи величини при дадено лице представляват **случай**. Измерването на единична променлива се нарича **наблюдение**. Понякога двата термина се използват като синоними, но в точния смисъл на думата случай представлява един компютърен запис или един ред от данни в електронна таблица за една индивидуална експериментална единица. Файлът с данни ще има n редове, съответстващи на броя на случаите.

СПЕЦИФИЧНОСТ (Specificity) - използва се за оценка на скринингови тестове и измерва вероятността за правилно идентифициране на здравите лица чрез даден скриниращ тест. Представлява пропорцията на лицата без за-

боляване в скринираната популация, които са определени като такива чрез скриниращия тест.

СРЕДНА АРИТМЕТИЧНА (Mean) – сумата от всички стойности в дадена група наблюдения, разделена на броя на наблюденията; статистика, която измерва центъра (средното ниво) на дадена извадка данни. За разлика от медианата, средната аритметична може да бъде силно повлияна от рязко отклоняващи се стойности.

СТАНДАРТНА ГРЕШКА НА СРЕДНАТА АРИТМЕТИЧНА (Standard error of the mean) – това е стандартното отклонение на разпределението на средните аритметични за извадки, извлечени от една и съща популация. То измерва точността на средната аритметична като оценка за параметъра в дадена популация. Изчислява се като стандартното отклонение се разделя на квадратен корен от броя на случаите в извадката.

СТАНДАРТНО НОРМАЛНО РАЗПРЕДЕЛЕНИЕ (Standard normal distribution) – нарича се още стандартна нормална крива и представлява идеално нормално честотно разпределение със средна аритметична, равна на нула и стандартно отклонение, равна на 1.

СТАНДАРТНО ОТКЛОНЕНИЕ (Standard deviation) – квадратен корен от дисперсията; статистика, която измерва варирането или разсейването на серия от данни. Означава се с “**s**” и се изразява в същите мерни единици, в които са представени оригиналните наблюдения и средната аритметична. Колкото по-разпръснати са данните в честотното разпределение, толкова по-голямо е стандартното отклонение.

СТАТИСТИКА (Statistic) – всяко нещо, което може да бъде изчислено от една извадка данни. Най-честото използване на тази дума е за обобщаващи мерки, такива като средна аритметична на извадката и стандартно отклонение, но дори и графични изображения като хистограмите също са статистики. Има два вида статистики: **описателни статистики** (които описват характеристиките на извадката) и **статистики за изводи и заключения** (които се използват за прилагане на статистически тестове). Друго предназначение на термина “статистика” е за означаване на широки категории на статистическия анализ на данни: **описателна статистика (descriptive statistics)** и **статистика за изводи и заключения (inferential statistics)**.

СТАТИСТИКА ЗА ИЗВОДИ И ЗАКЛЮЧЕНИЯ (inferential statistics) – раздел от статистиката като наука, в който на базата на получени резултати от наблюдения на извадки се направят изводи и заключения за параметрите в популациите и за статистическата значимост на различията в групите.

СТАТИСТИЧЕСКА ЗНАЧИМОСТ (Statistical significance) – когато даден изследовател демонстрира статистическа значимост чрез прилагането на даден статистически тест, това означава, че полученият резултат вероятно не се дължи на случайност, а е истински, реален. Това обаче не означава, че резултатът е важен или интересен. Той би трябвало да се сравни с клиничната значимост.

СТАТИСТИЧЕСКИ МОДЕЛ (Statistical model) – модел, описващ връзката между зависима променлива Y и една или повече независими променливи X . Приема общата форма: $Y = f(X_1, X_2, \dots) + \text{грешката}$, където $f(X_1, X_2, \dots)$ е математическа функция, а грешката представлява случайните отклонения от модела.

СТАТИСТИЧЕСКИ ТЕСТ ИЛИ КРИТЕРИЙ (Test statistic) – величина, изчислена от данните в наблюдавана извадка, чиято стойност се използва за определяне на това дали нулевата хипотеза трябва да се отхвърли или не. Изборът на критерий зависи от възприетия вероятностен модел и изследваната хипотеза.

СТЕПЕНИ НА СВОБОДА (Degrees of freedom) – термин, използван за характеристика на броя на независимите части на информацията, съдържаща се в даден показател. Например, ако имаме n наблюдения в случайна извадка и сме изчислили средната аритметична, то за оценка на дисперсията или стандартното отклонение може да имаме само $n-1$ независими измервания.

СТРАТИФИЦИРАНА ИЗВАДКА (Stratified sample) – извадка, в която се запазва пропорционално представителство на важни подгрупи от популацията. Следователно, към такава извадка следва да се прибъгва, когато популацията е хетерогенна, т.е. има фактори, които я разделят на подгрупи (слоеве). Извадката трябва да се подбира така че пропорциите на всяка подгрупа в извадката и в популацията да бъдат еднакви.

СТЪЛБОВА ДИАГРАМА (Bar chart) – начин за обобщаване на серия от категорийни данни (номинални или ординални). Използва се за илюстриране на главните особености на разпределението на данните. Състои се от правоъгълници с една и съща ширина, всеки от които отразява конкретна категория. Дължината (височината) на всеки правоъгълник е пропорционална на броя на случаите в съответната категория. Стълбовите диаграми могат да бъдат вертикални или хоризонтални като между стълбовете се оставя пространство, за разлика от хистограмите, където стълбовете са слети един до друг.

СТЪПКОВА РЕГРЕСИЯ (Stepwise regression) – най-добър регресионен модел понякога се разработва на етапи. Налице е списък от няколко възможни факторни променливи и този списък се изследва повторно, за да се определи кои променливи да се включат в модела. Най-напред се използва променливата, която най-добре обяснява промените в резултата, след това втората най-добра и т.н.

СУБЕКТИВНА ВЕРОЯТНОСТ (Subjective probability) – описва личните съждения на индивида за това доколко вероятно е да се прояви дадено събитие. Тя отразява степента на вярване в събитието и, както останалите вероятности, се изразява върху скала от 0 до 1.

СУМА ОТ КВАДРАТИТЕ (Sum of squares) – математическа величина, използвана в различни статистически процедури за измерване на вариабилността. Например, при изчисляване на дисперсията се сумират квадратите на отклоненията на индивидуалните стойности от средната аритметична и получената сума се разделя на $n-1$.

СЪБИТИЕ (Event) – един от резултатите в даден опит или експеримент.

ТАБЛИЦА НА КОНТИНГЕНЦИЯ (Contingency table) – начин за представяне на връзката между категорийни променливи в таблична форма. Когато изучаваната извадка или популация е разпределена според две променливи, всяка от които има по две разновидности, таблицата се нарича **четирикратна (2x2)**. Когато поне една от изучаваните променливи има повече от две категории, таблицата е **многократна**.

ТЕНДЕНЦИЯ В РАЗВИТИЕТО (Trend component) – продължителна промяна във времеви серии. Може да бъде възходяща и низходяща. Изучаването на тенденцията в развитието на дадено явление позволява да се прогнозират бъдещите стойности на явлението.

ТЕОРЕМА ЗА ЦЕНТРАЛНИТЕ ГРАНИЦИ (Central limit theorem) – тази теорема гласи, че за дадена случайна извадка от наблюдения от което и да е разпределение с определена средна величина и дисперсия, средната за големи извадки ще се стреми към нормално разпределение. Тази теорема обосновава широкото използване на доверителните интервали, опиращи се на нормалното разпределение и на t -критерия при оценка на средната и сравняване на две средни величини.

ТЕСТ НА КОЛМОГОРОВ-СМИРНОВ (Kolmogorov-Smirnov test) – непараметричен тест, който се прилага при единични извадки за проверка на съвместимостта с определена функция на разпределение или при две извадки за проверка на това дали те произлизат от едно и също разпределение.

ТЕСТ НА КРУСКАЛ-УОЛИС (Kruskal-Wallis test) – непараметричен тест за сравняване на три или повече извадки. Представлява логично разширение на теста на Уилкоксон и Ман Уитни и е аналог на критерия на Фишер от дисперсионния анализ, но не изисква нормално разпределение.

ТЕСТ НА УИЛКОКСОН И МАН-УИТНИ (Wilcoxon Mann-Whitney test) – един от най-мощните непараметрични тестове за сравняване на две популации. Не изисква нормално разпределение. Използва се при рангово подредени данни, т.е. при ординални скали.

ТЕСТОВЕ ЗА ПРОВЕРКА НА ХИПОТЕЗИ (Hypothesis tests) – тестове, опиращи се на извадка от данни, чрез които се определя коя от двете хипотези е вярна – нулевата или алтернативната.

ТОЧКОВО ОЦЕНЯВАНЕ (Point estimation) – предоставяне на оценка за даден параметър на популацията от гледна точка на единична стойност, която параметърът най-вероятно може да има. Точковата оценка обикновено се предоставя чрез статистика на извадката. От само себе си, точковото оценяване игнорира репрезентативната грешка.

ТОЧНОСТ НА ОЦЕНКАТА (Precision of an estimate) – обратното на стандартната грешка на оценката; мярка за това колко близка е дадена оценка до истинската стойност на даден параметър. Колкото е по-малка стандартната грешка, толкова по-голяма е точността. Това означава, че колкото е по-малък интервалът на доверителност, толкова по-голяма е точността. Следователно, точността може да бъде определена чрез доверителния интервал или чрез стандартната грешка.

ТРЕТИРАНЕ (Treatment) – въздействие, манипулации, които изследователите прилагат по отношение на експерименталните единици (напр., различни методи на лечение, различни лекарствени средства и т.н.).

ФАЙЛ С ДАННИ (Data file) – сбор от наблюдения върху няколко характеристики, подредени във вид на електронна таблица. Всеки ред от таблицата представлява единичен случай на наблюдение, а всяка колона – единична характеристика или променлива величина.

ФАКТОР (Factor) – променлива величина в даден статистически модел, чийто ефект върху зависимата променлива трябва да бъде изучен. Фактор е контролираната независима променлива; променлива, чиито нива се определят от експериментатора.

ХИ-КВАДРАТ (Chi-square test) - непараметричен статистически тест за проверка на хипотези, използван често при категорийни данни. Опира се на сравняване на наблюдаваните честоти и очакваните честоти в различните

категории. Прилага се при четирикратни и многократни таблици за контингенция. Изчислява се на основата на абсолютни числа. Ако стойността на хи-квадрат е по-голяма от съответната таблична критична стойност, H_0 е подкрепена с ниска вероятност (по-малка от 0.05) и се отхвърля. Приема се алтернативната хипотеза H_1 и се прави извод за наличие на връзка между изучаваните променливи.

ХИПОТЕЗА (Hypothesis) – предположение, издигнато от изследователя, което се оценява с помощта на събраните данни; предположение за същността на даден факт или явление или за връзката между две или повече променливи величини. Проверката на хипотези е съществена част от статистическите изводи и заключение. Използват се два вида хипотези: **нулева и алтернативна**.

ХИСТОГРАМА (Histogram) – графично изображение на честотно разпределение на количествена променлива чрез сливащи се правоъгълници с еднаква ширина и височина, пропорционална на честотата в съответния клас (интервал).

ЦЕЛЕВА ПОПУЛАЦИЯ (Target population) - цялата група, от която се интересува даден изследовател и за която той иска да направи изводи.

ЦЕНТРАЛНА ТЕНДЕНЦИЯ (Central tendency) – обобщаващо понятие за “средната” стойност на дадено разпределение. Типични измерители на централна тенденция са средната аритметична, модата и медианата.

ЧАСТИЧНА КОРЕЛАЦИЯ (Partial correlation) – мярка за силата на връзката между две или повече количествени променливи, която обяснява тяхната взаимовръзка с две или повече допълнителни променливи. Върху скала от -1 до +1 тя измерва степента на чистата корелация между двете променливи, в която не участват другите променливи.

ЧЕСТОТА (Frequency) – число, което показва колко пъти се среща дадена стойност в извадката или броят на стойностите, попадащи в рамките на определен интервал. Табличното обобщение на данните се нарича честотна таблица, а нейното графично представяне може да бъде стълбова диаграма или хистограма.

ЧЕСТОТЕН ПОЛИГОН (Frequency polygon) – графично представяне на честотното разпределение на количествена променлива, с интервални честоти, които са нанесени като срединни точки за съответните интервали и след това точките са съединени чрез прави линии.

ЧЕСТОТНА ТАБЛИЦА ИЛИ РАЗПРЕДЕЛЕНИЕ (Frequency table or distribution) – начин за обобщаване на серия от данни чрез таблично представя-

не, показващо колко пъти се проявява всяка стойност, конкретна характеристика на изучаваната променлива в рамките на серия от данни. Може да бъде усъвършенстван чрез добавяне на проценти, които попадат във всяка категория. Използва се предимно за обобщаване на категорийни, номинални и ординални данни. Може да се използва и при непрекъснати данни, след като те се разделят в подходящи класове. При повече от една категорийна променлива честотната таблица придобива вид на **таблица за контингенция**.

ЧУВСТВИТЕЛНОСТ (Sensitivity) – използва се за оценка на скринингови тестове и измерва вероятността за правилно диагностициране на болните лица. Представява пропорцията на истински положителните за заболяване лица в скринираната популация, които са определени като такива чрез скриниращия тест.