

03-Упражнение

Тест на хипотеза. Т-тест.

ас.д-р Костадин Костадинов

Преговор

1. При стандартизация за да се премахне въздействието на потенциален “замъгляващ” фактор. Екстензивните показатели за разпределение (структурни) се ползват за стандарти по които умножаваме нестандартизираните интензивни показатели.¹
2. За да се характеризират променливите в една извадка е необходимо да се определи, тяхната централна тенденция и вариабилност.
3. Променливите са “парченца информация”, които събираме за да добием представа за изследвания обект или съвкупност. Те могат да бъдат както количествени (измерени в число), така и качествени (вид етикет, наименование).
4. Количествените променливи могат да бъдат измервани на силни скали като пропорционална или интервална, докато качествени променливи се отчитат на по-слаби скали като номинална и ординална.
5. Качествени признаци, които са подредим се измерват на ординална скала (стадии на онкологично заболяване).
6. Качествени признаци, които имат само две възможни стойности се наричат алтернативни и се измерват на дихотомна скала (мъж/жена)

¹ Нестандартизираните показатели не се сумират, за разлика от стандартизираните.

7. Показатели за централна тенденция са мода, средна аритметична, медиана.
8. Показатели за разсейване са размах, дисперсия, стандартното отклонение и интерквартилният размах.
9. За количествени променливи (измервани на пропорционална скала например), можем да използваме средна аритметична \bar{x} и стандартно отклонение SD . При асиметрично (ляво или дясно изтеглено) разпределение по-подходящи са медианата Me и интерквартилният размах IQR .
10. Стандартното отклонение показва, каква е вариационността на наблюденията. С други думи, колко добре средната аритметична представя извадката.²
11. Стандартното отклонение може да бъде използвано за да се прецени колко % от наблюденията имат над или под определена числена стойност. Това може да се случи, **само** ако изследваната променлива имат нормално разпределение.
12. **Boxplot** е визуална презентация на количествените данни. Представява “кутия с мустаци”. Двете страни на кутията съответстват на $Q1$ и $Q3$, а средната линия разделяща кутията представлява медианата. С него можем да определим редица показатели³
13. За качествените променливи (измервани на номинална скала например), можем да използваме пропорция, процент %. Те нямат показател за разсейване, както количествени променливи.
14. В основа на **извадката** е невъзможно да определим стойността на средната аритметична за **генералната съвкупност** с пълна (100%) сигурност и точност. Възможно е само да определим интервал от числа, в който с определена степен на сигурност можем да твърдим, че се намира средната аритметична за генералната съвкупност. Този интервал се нарича интервал на доверителност.⁴

² Извадки с едни и същи стойности (например 10 пациента с кръвна захар равна на 4.1 ммол/л) ще имат стандартно отклонение 0.

³ Медианата. Първият $Q1$ и третият квартил $Q3$. С тяхна помощ можем да изчислим и IQR . Вида на разпределението**, като визуално преценим на коя посока опашката е по-дълга. (ляво изтеглено-опашка на ляво, дясно изтеглено-опашка на дясно, симетрично- двете опашки са равни). Наличието на стойности бегачи (outliers), като използваме формулата на Тъки - долни стойности бегачи са тези наблюдения със стойност по-малка от $Q1 - 1.5 \cdot IQR$, а “горни” стойности бегачи са всички наблюдения със стойност на признака по-голяма от $Q3 + 1.5 \cdot IQR$

⁴ Доверителният интервал може да се построи, както за количествените променливи (средна аритметична), така и за качествените променливи (отношения, проценти). За да го изчислим е необходимо първо да преценим колко точна е нашата стойност, която сме получили за извадката.

15. Средната аритметична за извадката се нарича точкова оценка (или статистика). Нейната точност се определя, чрез стандартна грешка на средната аритметична SEM . За пропорции използваме стандартна грешка за относителен дял SE_p .
16. Интервалът на доверителност зависи от тези стандартни грешки, но също така от нивото на доверителност. В медицината най-често се използва $95\%CI$. Въпреки това, популярни са и 90% и 99% интервали. При равни други условия, ако увеличим степента на сигурност за генералната съвкупност: от 90 до 99% например, е необходимо да увеличим и гаранционният множител - тоест широчината на този интервал също ще расте.
17. Доверителният интервал зависи и от стандартната грешка: ако имаме много голяма стойност за стандартна грешка, то ще имаме и широк интервал на доверителност (при равни други условия)
18. Стандартната грешка от своя страна зависи от броя на наблюденията в извадката (голям брой наблюдения \sim малка грешка), както и от тяхната вариабилност около централната тенденция (стандартното отклонение, малко стандартно отклонение \sim малка грешка)

Тест на хипотези

Освен да опишем една извадка и да направим предположение за това каква е стойността на показателя в генералната съвкупност, ползваме статистиката и за тестване на хипотези. Хипотеза е предположението за наличието на връзка между дадени променливи (инфаркт на миокарда и тютюнопушенето например). Хипотезите се проверяват с помощта на статистически методи по данни от представителни извадки. В този случай резултатите, направените изводи и заключения са валидни за популацията. В статистиката използваме понятията “Нулева” и “Алтернативна хипотеза”.

Нулева хипотеза

Определение

Нулевата хипотеза H_0 е винаги **отричаща** наличието на връзка между променливите. Например: новият медикамент за хипертония **не** намалява стойностите на кръвното налягане, ако има подобно намаляване, то е случайно събитие или тестваният диетичен режим **не** води до намаляване на теглото, ако е наблюдавано такова, то е в резултат на случайност.

Алтернативна хипотеза

Определение

Алтернативната хипотеза H_1 представлява противоположност на нулевата. Тя гласи, че в действителност има значимо различие или значима асоциация между действащият фактор и интересуващият ни резултат. Например: новият медикамент за хипертония намалява стойностите на кръвното налягане, това намаление е **много малко вероятно да бъде случайно**.

Основна концепция - пример

При тестването на хипотези се определя, колко е вероятно едно събитие (например разлика в кръвното налягане при лечение) или една взаимовръзка (като белодробен рак и тютюнопушене) да бъде случайно. След като определи, каква е вероятността едно събитие да бъде случайно, можем да отхвърлим или не нулевата хипотеза като сравним това ниво на случайност с предварително зададена граница.

Представете си, че тестваме ново лекарство за хипертония. Избираме две напълно случайни рандомизирани групи пациенти с хипертония. Започваме да лекуваме едната група с новия медикамент, а другата с плацебо (таблетка, която не съдържа лечебно вещество, но прилича на новия медикамент). В края на изследването можем да сравним колко средно се е намалило кръвното налягане в експерименталната група (тези с истинския медикамент) спрямо това на плацебо групата.

Нека речем, че експерименталната група има средно 10 mmHg по-ниски стойности на кръвното налягане спрямо плацебо групата.⁵

Ние тестваме, *само нулевата* хипотеза. Определяме, **колко е възможно** (колко е вероятно) да получим подобна разлика в резултат на случайност. Да речем, сме изчислили, че е тази вероятност е 0.09 (интерпретира се като 9%). Това означава, че възможността тази разликата в стойностите на кръвното налягане да е случайна е равна 9%. С други думи, ако направим този експеримент по същият начин, с други случайни извадки 100 пъти, е възможно в 9 от тези опити да получим същата или дори по-голяма разлика напълно абсолютно случайно.

Можем ли да отхвърлим нулевата хипотеза? Ако шансът тази разлика да е случайна е само 9%, то в 91% тази разлика няма е случайна. В 91 от 100 изследвания тази разлика ще се дължи на новото лекарство. Колко обаче е нужно да бъде тази вероятност, за да отхвърлим нулевата хипотеза и индиректно да приемем алтернативната? Ако не можем да отхвърлим нулевата хипотеза, а се окаже, че лекарството наистина работи сме допуснали *грешка*. Също ще допуснем *грешка*, ако отхвърлим нулевата хипотеза, приемем оставащата

⁵ Работната (нулева) хипотеза би гласяла “Тази разлика между кръвното налягане на експерименталната и контролната група е получена случайно.” докато Алтернативната хипотеза е противоположна, тя би гласяла “Разликата между стойностите на кръвното налягане на плацебо и контролната група НЕ е в резултат на случайност.

възможност- алтернативната, а всъщност нулевата е вярна и лекарството наистина не работи.

Тези две грешки можем да онагледим в таблица ⁶:

Видове грешки

таблица 1: Видове грешки

	Лекарството не работи	Лекарството работи
Приемаме H_0	Правилно решение	Грешка от втори род β
Отхвърляме H_0	Грешка от първи род α	Правилно решение

⁶ В нашият пример преди да започнем с това проучване, ние сами трябва да определим нивото си на грешка α . Още преди началото на експеримента, трябва да знаем, с какво ниво на грешка ще работим. В медицинските изследвания, най-популярни са нивата 0.1; 0.05; 0.01; 0.001.

Да приемем, че сме избрали ниво на грешка 0.05 . Това означава, че ние си “позволяваме” да отхвърлим нулевата хипотеза и да приемем лекарството за ефективно (алтернативната), само когато вероятността да получим подобна разлика в кръвното налягане случайно е под 5% . В нашият пример, тази вероятност е 0.9 (9%). Крайният ни извод не ни позволява да отхвърлим нулевата хипотеза

Освен грешки от първи род съществуват и такива от втори: това са грешки, в които не можем да отхвърлим нулевата хипотеза, когато тя действително е грешна. Тоест не можем да докажем, че медикамента действа. Рискът за тези грешки са в зависимост от риска за грешка от 1-ви род. Ако решим да сме по-либерални и да увеличим риска от грешка от първи род, тоест по-лесно да отхвърляме нулевите хипотези, намаляваме риска за грешка за от втори и обратно. Тази зависимост обаче не е пропорционална. Рискът за грешка от втори род е свързана и с понятието мощност.

Мощност

Един статистически тест, е **мощен**, когато успява да установи дори и малки разлики между двете групи. Обратно, един

тест се счита за тест с ниска мощност, ако не успяваме да отхвърлим нулевата хипотеза с него, дори и разликата между групите е голяма.

Мощността на теста зависи от:

- Риска за грешка от 1-ви род α . При избирането на много високи нива на алфа (например $0.1 = 10\%$) е по-вероятно да се отхвърлят нулевите хипотези.⁷
- Размер на “терапевтичния ефект”. В ситуация в която “новият” медикамент действително е много силен и намаля драстично кръвното налягане, ефектът върху лекуваната група ще се отрази в голяма степен и на разликата в средните аритметични между двете групи. Колкото е по-голяма тази разлика по абсолютна стойност, толкова по-малко вероятно е да е в резултат на случайност.⁸
- Хомогенност на групата. Да приемем, че средното систолното кръвното налягане в експерименталната група е с 10 mmHg по-ниско от контролната група. Тази разлика ще се установи статистически, по-лесно ако всички участници в експерименталната група имат близки стойности на кръвното налягане. Това ще доведе до по-малка вариабилност вътре в групата, а от там по-малка стойност на стандартното отклонение. Малко стандартно отклонение от своя страна е причината за по-точна оценка и малка стандартна грешка.⁹
- Обем на извадките. Точността на оценка (в случая средното систолно артериално налягане) зависи и от обема на извадката. По-голям обем на всяка от групите води по-малка грешка (при равни други условия). Това от своя страна е причина за по-точна оценка и намаляване на случайността. Отново по-същата логика, мощността на теста расте.

⁷ Това намаля и риска за грешка от 2-ри род, което увеличава мощността на теста. При равни други условия, тест при ниво $\alpha = 0.1$ ще е по-мощен от тест при ниво на $\alpha = 0.05$.

⁸ Намалявайки вероятността разликата да е случайна, увеличаваме вероятността да отхвърлим нулевата хипотеза. Това води до намаляване и на риска за грешка от втори род и следователно до по-мощен тест.

⁹ Колкото е по-малка стандартната грешка, толкова по-тесен ще е интервала на доверителност (по-прецизен) и толкова по-малко ще е възможно разликата между групите да е случайна.

Видове тестове на хипотези.

Преди да изберем теста трябва да имаме отговор на следните 3 въпроса:

1. Какво е разпределението на данните?¹⁰
2. Какъв е типа на резултата от изследването?¹¹
3. Какъв е дизайна на изследването? ¹²

Статистическа и клинична значимост

Както установихме, статистиката понякога е обект на “атака”. В случай, в които изберем експерименталните групи малко по-различно от “напълно случайно”, ако увеличим броя на участниците с “още” и “още,” ако изберем по-либерално ниво за грешката от 1-ви род и използваме мощен тест то разбира се ще получим и *статистически значим резултат*. Това обаче, не означава, да приемаме получените резултати като “чиста монета”. Както установявате е възможно “статистиката” да бъде манипулирана в една или друга посока с цел да се получи желаният статистически значим. Затова трябва внимателно да преценяваме разликата между статистическа и клинична значимост

¹⁰ Тук възможностите са като цяло две - **нормално разпределени** (форма на камбанка, симетричен, като двете страни са на еднакво разстояние от центъра. Медианата е равна на средната аритметична) или **асиметрично** (ляво и дясно изтеглено)

¹¹ Дали става въпрос за количествена променлива (стойност на кръвното налягане, кръвна захар, тегло) или качествена такава (пропорция на оздравели, пропорция на пациенти в по-тежка степен на заболяване).

¹² Дали имаме само една група, която сравняваме с някаква известна стойност или две групи, които са зависими например преди и след лечение. Възможно е да изследваме две независими групи- контрола и експериментална или повече от две групи- една контролна с две експериментални на различни дозировки например. Възможно да имаме повече от две зависими групи (преди и след лечение).

Статистическата значимост

Представява оценка на вероятността да получим подобен или по-висок резултат в следствие на случайност. “Висока значимост” означава, че е много малко вероятно полученият резултат да е случаен.

Клинична значимост

Представява оценка на ефектът на новия медикамент или интервенция върху пациента. Нов медикамент, може да е “статистически значим” като редуцира кръвното с 2 mmHg, но това не е клинично значимо - нито пациента, нито близките, нито вие като лекар ще установите тази разлика субективно. Съществува и ситуация в която медикаментът не е статистически значим, но е клинично значим.

Когато отговорим на тези въпроси, можем да изберем тест, който използваме в следната таблица

таблица 2: Избор на статистически тест

	Номинална	Ординална	Пропорционална интервална
Вид	Непараметричен	Непараметричен	Параметричен
Разпределение	Асиметрично	Асиметрично	Нормално/Симетрично
Една извадка	χ^2 Тест	Тест на Wilcoxon	One sample t-test
Две свързани извадки	Тест на McNemar	Тест на Wilcoxon	Repeated measures t-test
Две независими извадки	χ^2 Тест	U тест Mann Whitney	Independent t-test
Повече от две свързани извадки	Тест на McNemar	Friedman	ANCOVA, RM-ANOVA
Повече от две независими извадки	χ^2 тест	Kruskal Wallis	ANOVA

Т-тест

В рамките на тези упражнения, ще се запознаем с Т теста (установяващ дали е налична статистически значима **разлика** в нормално разпределена количествена величина в две независими една от други групи)

Преди да изясним същността на Т теста, нека зададем някои условия към нас самите, за да го използваме:

- Признакът, за който ще подлагаме на Т тест (величината, която ще сравняваме) трябва да бъде **нормално** разпределена величин.
- Двете изследвани извадки следва да имат приблизително еднаква вариабилност (стандартно отклонение)
- Наблюдавали сме поне 30 единици (пациенти, изследвани лица) във всяка от групите.

След като изяснихме тези изисквания, можем да пристъпим към реализирането на теста. Той има един основен вариант (за количествени признаци) и един модифициран (когато искаме да установим дали разликата между две пропорции е статистически значима).

Т-тест за количествени признаци.

Нека разгледаме следния *пример*. Провели сме проучване, за да установим дали нов медикамент - бета блокер е ефективна терапия за намаляване на пулса.¹³

Целта на изследването е да докаже, че медикамента действително е ефективен. Тоест пациентите, които използват медикамента имат по-нисък пулс спрямо тези, които не приемат бета блокер. За да постигне тази цел, компанията е извършила първоначален експеримент с 80 случайно избрани пациенти - 40 са получили новия бета блокер, а останалите 40 са контролна група, те са приемали плацебо. За да докажем дали бета блокера е ефективен, измерваме сърдечната честота на всеки един участник в изследването и записваме стойността. Също така записваме

¹³ При пациенти със сърдечно-съдови заболявания, пулс около 60 е оптимален, защото позволява сърцето да има по-дълга диастола. Това води по-продължително време, в което сърцето да се пълни от коронарните артерии.

дали този участник приема бета блокер или е от контролната група. Стойностите са както следва:

1. За лекуваните пациенти:

[1] 58 58 58 59 60 60 60 60 61 61 62 62 63 64 64 64 64 65 65 65 65 66 66 66 66
[26] 67 67 67 68 68 68 69 69 70 70 70 71 71 71 72

2. За пациентите с плацебо

[1] 68 68 70 71 71 72 72 74 74 75 75 76 76 78 78 79 80 80 81 82 82 82 83 84 84
[26] 85 85 86 88 89 89 90 91 92 93 93 93 94 94 96

Нека да започнем със стъпките на Т теста:

Дефиниране на нулева хипотеза H_0 и алтернативна хипотеза H_1

Нулева хипотеза Няма разлика между средната сърдечна честота, сравнявайки двете групи пациенти. В случай, че такава се наблюдава, това е в резултат на случайност.¹⁴

Алтернативна хипотеза Има разлика между средната сърдечна честота, сравнявайки двете групи пациенти.¹⁵

Обърнете внимание, самият Т тест не може да потвърди дали бета блокера е ефективен за намаляване на пулса. За доказване на такава връзка t теста е необходим статистически тест, но не достатъчен.

Определяне на риска за грешка (алфа).

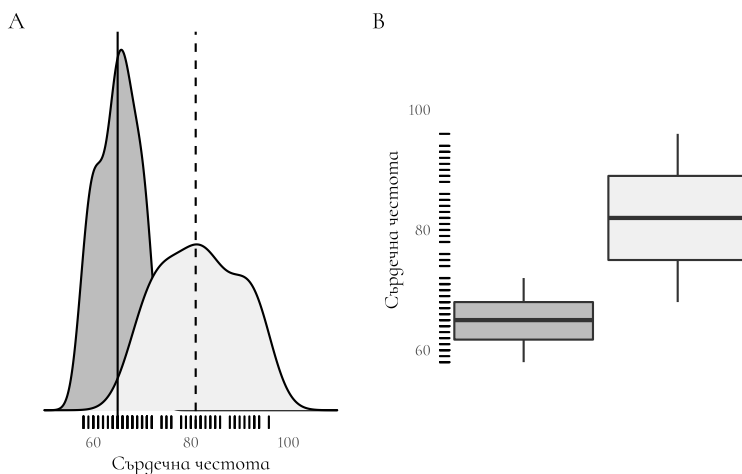
Както посочихме по-рано това ниво в медицинските изследвания обикновено е 0.05 (5%). Припомням, че можем да “отхвърлим” нулевата хипотеза, само ако вероятността да наблюдаваме явлението в резултат на случайност е под риска за грешка. След като имаме дефинирани нашите хипотези и риска за грешка от 1-ви род, преминаваме към следващата стъпка.

¹⁴ В случая това е хипотезата, която ще тестваме. Тази хипотеза предполага, че независимо в коя група е пациента (тази с бета блокер или контролната) разлика не се наблюдава. Дори и да има такава, то това ще е в резултат на случайността (а не на медикамента)

¹⁵ Алтернативната хипотеза е всичко онова, което остава, ако отхвърлим нулевата. Понеже нулевата хипотеза е ясно дефинирана “няма”, то логично алтернативната гласи, че такава разлика има и тя не е случайна.

Запознаване с данните.

От фигура 1 по-долу, както и от таблица 3 можем да добием представа за това как е разпределена величина пулс в двете групи. В групата с бета блокер установяваме среден пулс от 65уд/мин, докато в контролната група пулсът е близо 82 уд/мин. Разликата е почти 17уд/мин.



фигура 1: В групата с бета блокер установяваме среден пулс от 65уд/мин (непрекъсната линия), докато в контролната група пулсът е близо 81 уд/мин (прекъсната линия). Разликата е почти 17уд/мин. Също така добиваме и представа за това колко хомогенни са групите. Пулса на пациентите с бета блокер варира близо около средната аритметична, докато вариацията в стойностите на контролната групата изглежда доста по-голяма. Това се потвърждава и от стандартното отклонение

характеристика	Бета блокер	Плацебо
Средна аритметична	65.00	81.83
SD	4.08	8.16
Медиана	65.00	82.00
IQR	6.25	14.00

таблица 3: Описателни статистики в двете групи

Изчисляване на t-статистиката

За да достигнем до извод колко е вероятно тази разлика да е случайна, трябва да изчислим т.н стойност t. Стойността на t се изчислява, като се използва формула ¹⁶

Първо трябва да определим колко е стандартната грешка:

$$SE_{m1} = \frac{SD_1}{\sqrt{n_1}} = \frac{4,08}{\sqrt{40}} = \frac{4,08}{6,32} = 0,75$$

¹⁶ В нашият случай, вероятността е не просто по-малка от 0,05. р е по-малко от 0,01 !

След като сме определили точността с която сме оценили първата средна аритметична, следва да направим това и за пациентите в плацебо групата:

$$SE_{m2} = \frac{SD_2}{\sqrt{n_2}} = 1,29$$

Тук е момента да припомним, че втората група има по-голяма грешка поради факта, че групата е по-нехомогенна, тоест участниците в нея средно се различават повече от средната аритметична на групата, за разлика от пациентите лекувани с бета блокер.

$$t = \frac{|65,00 - 81,83|}{\sqrt{0,75^2 + 1,29^2}}$$

Заместваме всичко във формулата за t- теста.

$$t = \frac{|-16,83|}{\sqrt{0,56 + 1,66}} = \frac{16,83}{\sqrt{2,22}} = \frac{16,83}{1,49} = 11,3$$

Изчисляване на р стойността.

Сега вече имаме изчислена стойността на t критерия. Тази стойност ни показва каква е разликата между средните аритметични, като ги разделя на точността на оценките им (стандартните грешки). След като имаме тази t стойност. Можем да използваме t разпределението, за да разберем колко е вероятно да получим такава t стойност или по-голяма абсолютно случайно.

t разпределението е вероятностно разпределение. Това означава, че можем да определим, каква е вероятността да получим определена стойност на t критерия или по-висока в резултат на случайност. Тази вероятност се бележи с p . Наричаме я още р-стойност или ниво на значимост. За да определим p стойността, спрямо получената стойност на t критерия използваме листа с формули. В таблицата с t разпределението, се поглежда само последният ред, той изглежда така:

таблица 4: t-разпределение

df	p = 0.1	p = 0.05	p = 0.01
30	t = 1.645	t = 1.960	t = 2.576

Как обаче да интерпретираме тази таблица?

Първо, трябва да знаем, какво представлява df . Наричат се още **степенни на свобода**. Можем да посочим, че тези степенни на свобода зависят от единиците на наблюдения и за да улесним още повече работата си, винаги можете да наблюдавате само последния ред, в случай на повече от 30 пациента.

Всяка една колона представя каква е вероятността (p) за всяка стойност на t . Например, ако получим $t = 1,645$ вероятността за случайност е 10% (0.1). Колкото е по-висока стойността на t толкова по-малко вероятна е тя.

Както помним от формулата, стойността на t критерия зависи от разликата в средните аритметични и стандартната грешка (тоест зависи още от вариационността и броя на единиците за наблюдение). Това означава, че можем да направим следните заключения:

- Колкото е по-голяма t стойността, толкова по-малко вероятно е да се наблюдава тя. Тоест по-малко вероятно е разликата в средните аритметични да е случайна.
- Колкото абсолютната разлика в средните аритметични е по-голяма, толкова по-голяма е стойността на t критерия.
- Колкото са по-хомогенни групите (малко стандартно отклонение), толкова грешката е по-малка. Малка грешка води до висок t критерий.
- Колкото повече единици на наблюдение имаме, толкова по-малка ще е стандартната грешка. Малка стандартна грешка води до висок t критерий.

Сега е момента да направим финалното заключение за нашето проучване. Разликата между двете групи е 16,83. Тази разлика при отчитането на стандартните грешки резултира в t критерий 11,3. Поглеждаме таблицата. Там тази стойност

липсва. Най-близката стойност е 2,57, която съответства на вероятност 0,01 (1%). Това означава, че нашият t критерий е по-голям от тази стойност, следователно и вероятността да наблюдаваме подобно число е по-малка от 0,01. Тоест вероятността p да наблюдаваме подобна разлика в пулса е по-малка от 1%. Сега е момента да се върнем на тестването на хипотеза. Още в началото на изследването бяхме заложили риска за грешка 5% (0,05). Всяка вероятност под риска за грешка, означава че отхвърляме нулевата хипотеза. Това трябва да направим и в случая

Отхвърляме H_0 и приемаме H_1 , защото при $t=11,3$ $p < \alpha < 0.05$. Следователно, съществува статистически значима разлика в средните нива на пулса между двете извадки

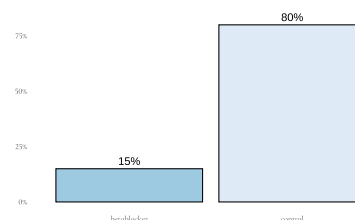
Т-тест за качествени признаци.

Нека си представим, че това изследване е продължено още поне 5 години. Пациентите са проследени и сме записали, кои от тях са починали. Все още пазим и старите ни записи, в които знаем кои са били на бета блокер (експериментална група) и кои са били в плацебо групата (контроли). Нека разгледаме данните. Починали са общо 38 пациенти. В групата с бета блокер от 40 са починали 6 (15%), а в групата с плацебо от 40 са починали 32 (80%). От фигура 2, ясно прави впечатление, че пропорцията на починалите сред експерименталната група е в пъти по-малка от тази на контролната група. Но дали тази разлика е значима статистически? Дали не е просто случайност?

За да установим това, отново можем да използваме t -теста. В този пример, обаче не използваме средна аритметична, а съотношения (процент). Въпреки това знаем, че имаме възможност да изчислим стандартна грешка за процент. Нека си припомним формулата:

$$SE_p = \sqrt{\frac{\hat{p} \cdot (100 - \hat{p})}{n}}$$

В нашият пример, трябва да установим стандартната грешка и за двете извадки. Нека започнем с пациентите приемачи



фигура 2: Пропорция на починалите в двете групи пациенти

бета блокер. При тях 15% са починали, каква е стандартната грешка на тези 15% получаваме като приложим формулата.

$$SE_{p1} = \sqrt{\frac{\hat{p}_1 \cdot (100 - \hat{p}_1)}{n}} = \sqrt{\frac{15 \cdot (100 - 15)}{40}} = \sqrt{\frac{15 \cdot 85}{40}} = \sqrt{\frac{1275}{40}} = \sqrt{31,88} = 5,6$$

След това правим същото за пациентите в контролната група, при тях починали са 80%. Стандартната грешка за тези 80% получаваме по формулата:

$$SE_{p2} = \sqrt{\frac{\hat{p}_2 \cdot (100 - \hat{p}_2)}{n}} = \sqrt{\frac{80 \cdot (100 - 80)}{40}} = \sqrt{\frac{80 \cdot 20}{40}} = \sqrt{\frac{1600}{40}} = \sqrt{40} = 6,32$$

Дефинираме нулева и алтернативна

Нулева хипотеза Няма разлика между съотношението на починалите в двете групи пациенти. Дори и да наблюдаваме такава, тя е резултат на случайност и не съществува в генералната съвкупност.

Алтернативна хипотеза Има разлика между относителните дялове на починалите пациенти в двете групи. Тази разлика не е в резултат на случайност.

Определяне на риска за грешка (ниво на значимост) (алфа).

Отново ще използваме нивото 0,05.

Изчисляване на t-статистиката

Отново имаме формула, която прилича много на тази с която вече работихме.

$$t = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{SE_{p1}^2 + SE_{p2}^2}}$$

Заместваме и извършваме изчислението.

$$t = \frac{|15 - 80|}{\sqrt{5,6^2 + 6,32^2}} = \frac{|-65|}{\sqrt{31,36 + 40,45}} = \frac{65}{\sqrt{31,36 + 40,45}} = \frac{65}{\sqrt{71,81}} = \frac{65}{8,47} = 7,6$$

След като получим стойността на t критерия проверяваме в коя посока се намира на таблицата: наляво към по-вероятните или надясно към числата с ниска вероятност

таблица 5: t-разпределение

df	p = 0.1	p = 0.05	p = 0.01
30	t = 1.645	t = 1.960	t = 2.576

Установяваме, че t критерият е надясно. Стойността му е по-висока от 2,576, тоест вероятността да получим това число е по-малка от 0,01.

Заклучение

Отхвърляме H_0 и приемаме H_1 , защото при $t=7,6 < \alpha < 0.05$ ¹⁷. Следователно, съществува статистически значима разлика в средните нива на пулса между двете извадки

¹⁷ В нашият случай, вероятността е не просто по-малка от 0,05. р е по-малко от 0,01 !