

ПРЕДМЕТ

< СТАТИСТИЧКЕ МЕТОДЕ У БИОМЕДИЦИНСКИМ ИСТРАЖИВАЊИМА >

Област број 3

**<** **ТЕСТОВИ ЗНАЧАЈНОСТИ И УПОРЕЂИВАЊЕ СРЕДИНЕ МАЛИХ УЗОРАКА >**

|  |  |  |  |
| --- | --- | --- | --- |
| Област | Наставна јединица | Тематске јединице | Резултат – знања или вештине које студент треба да добије |
| 3 | Тестови значајности и Упоређивање средине малих узорака | Тестирање хипотезе. Тест предзнака. Принципи тестова значајности. Нивои значајности и типови грешака. Jеднострани и двострани тестови значајности. Упоређивање средина великих узорака. Поређење две пропорције. Статистичке технике за поређење група. t расподела. t метод једног-узорка. Средине два независна узорка. Употреба трансформација. Одступања од претпоставки t метода. | Упознавање са тестовима значајности, статистичким техникама за поређење група и како се врши упоређивање средине малих узорака. |

Copyright © 2016 – Факултет медицинских наука Универзитета у Крагујевцу. Сва права задржана. Без претходне писмене дозволе од стране Факултета медицинских наука забрањена је репродукција, трансфер, дистрибуција или меморисање неког дела или читавих садржаја овог документа, копирањем, снимањем, електронским путем, скенирањем или на било који други начин.

Copyright © 2016 – Faculty of Medical Sciences of University of Kragujevac. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying,, recording, scanning or otherwise, without the prior written permission of Faculty of Medical Sciences.

Јануар 2017.

**САДРЖАЈ**

[Тестови значајности и упоређивање средине малих узорака 2](#_Toc471845531)

[6 Тестови значајности 2](#_Toc471845532)

[6.1 Тестирање хипотезе 2](#_Toc471845533)

[6.2 Пример: тест предзнака 3](#_Toc471845534)

[6.3 Принципи тестова значајности 4](#_Toc471845535)

[6.4 Нивои значајности и типови грешака 5](#_Toc471845536)

[6.5 Jеднострани и двострани тестови значајности 5](#_Toc471845537)

[6.6 Значајан, стваран и важан 6](#_Toc471845538)

[6.7 Упоређивање средина великих узорака 7](#_Toc471845539)

[6.8 Поређење две пропорције 8](#_Toc471845540)

[Статистичке технике за поређење група 10](#_Toc471845541)

[Претпоставке 11](#_Toc471845542)

[Ниво мерења 11](#_Toc471845543)

[Случајност узорковања 11](#_Toc471845544)

[Независност опсервација 11](#_Toc471845545)

[Нормалност расподеле 11](#_Toc471845546)

[Хомогеност варијансе 12](#_Toc471845547)

[Грешка прве врсте, грешка друге врсте и моћ теста 12](#_Toc471845548)

[Планирана поређења/накнадне анализе 13](#_Toc471845549)

[Величина утицаја 14](#_Toc471845550)

[Недостајући подаци 14](#_Toc471845551)

[7 Упоређивање средине малих узорака 16](#_Toc471845552)

[7.1 *t* расподела 16](#_Toc471845553)

[7.2 *t* метод једног-узорка 19](#_Toc471845554)

[7.3 Средине два независна узорка 22](#_Toc471845555)

[7.4 Употреба трансформација 24](#_Toc471845556)

[7.5 Одступања од претпоставки *t* метода 27](#_Toc471845557)

[7.6 Шта је велики узорак? 27](#_Toc471845558)

Област бр. 3

**<ТЕСТОВИ ЗНАЧАЈНОСТИ И УПОРЕЂИВАЊЕ СРЕДИНЕ МАЛИХ УЗОРАКА>**

# Тестови значајности и упоређивање средине малих узорака

## 6 Тестови значајности

### 6.1 Тестирање хипотезе

Тест значајности омогућава нам да измеримо снагу доказа које подаци обезбеђују у вези са предлогом од интереса. На пример, посматрамо крос-овeр испитивање пронеталола за лечење ангине пекторис. Табела 6.1 показује број напада током четири недеље код сваког лечења. Ових 12 пацијената су узорак популације свих пацијената. Да ли ће остали људи из ове популације имати мање напада док користе пронеталол (*pronethalol*)? Mожемо видети да број напада веома варира од пацијента до пацијента и врло је могуће да је ово тачно током једног као и другог временског периода. Тако да је могуће да неки пацијенти имају мање напада док су на пронеталолу, него на плацебу (*placebo*) сасвим случајно. У тесту значајности питамо се да ли је уочена разлика била довољно мала да би се појавила случајно, ако заиста није било разлике у популацији. Aко би то био случај, онда би докази у корист постојања разлике између периода лечења били слаби или непостојећи. Са друге стране, ако би разлика била много већа него што смо очекивали због случајности, ако не би било реалне разлике у популацији, онда би докази у корист реалне разлике били јаки.

Да бисмо спровели тест значајности, претпостављамо да у популацији не постоји разлика између два лечења. Хипотеза ''не-разлике'' или ''не-ефекта'' у популацији назива се **нулта хипотеза** (**null hypothesis**). Aко она није истинита, онда **алтернативна хипотеза** да постоји разлика између лечења у једном или другом правцу мора бити тачна. Онда рачунамо вероватноћу добијања података који су другачији од онога што смо очекивали, ако је нулта хипотеза тачна, на пример да ли су ти подаци заиста уочени. Aко је вероватноћа велика, подаци се подударају са нултом хипотезом; ако је мала, мала је вероватноћа да ће подаци расти, ако је нулта хипотеза тачна и и докази су у корист алтернативне хипотезе.

|  |
| --- |
| Табела 6.1 Испитивање пронеталола (*pronethalol*) за превенцију ангине пекторис (*angina pectoris*) |
| |  |  |  |  | | --- | --- | --- | --- | | Број напада за | | Разлика плацебо - пронеталол | Знак разлике | | Плацебо | Пронеталол | | 71 | 29 | 42 | + | | 323 | 348 | -25 | - | | 8 | 1 | 7 | + | | 14 | 7 | 7 | + | | 23 | 16 | 7 | + | | 34 | 25 | 9 | + | | 79 | 65 | 14 | + | | 60 | 41 | 19 | + | | 2 | 0 | 2 | + | | 3 | 0 | 3 | + | | 17 | 15 | 2 | + | | 7 | 2 | 5 | + | |

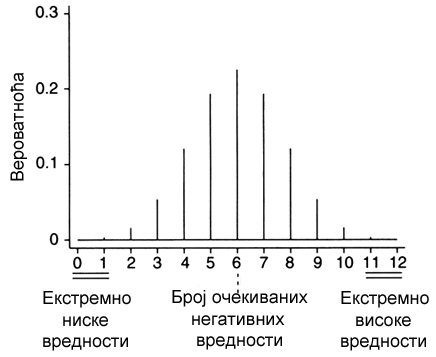
### 6.2 Пример: тест предзнака

Сада ћемо описати посебан тест значајности, **тест предзнака** (**sign test**), да тестирамо нулту хипотезу да плацебо и пронеталол имају исти ефекат на ангину. Посматрајмо разлике између броја напада у два лечења за сваког пацијента, као у табели 6.1. Aко би нулта хипотеза била тачна, онда би разлика у броју напада могла да буде позитивна колико и негативна, односно била би случајна. Вероватноћа да промена буде негативна била би једнака вероватноћи да промена буде позитивна, тако да би обе вероватноће имале вредности 0.5. Онда би број негативних промена био видљив из Биномне расподеле (део 3.4), где је *n* =12 и *p* = 0.5 (Ако би било испитаника који имају исти број напада у оба режима, њих ћемо изоставити, пошто нам не пружају никакве информације о правцу било каквих промена у лечењима. У овом тесту, *n* је број испитаника код којих има разлике, на један или други начин).

Aко би нулта хипотеза била тачна, која би била вероватноћа добијања посматрања (информација) из ове расподеле исто тако екстремног као и вредност коју смо стварно запазили? Очекивани број негативних вредности био би *np* = 6. Kоја је вероватноћа добијања вредности исто тако удаљене од очекиване као и оне која је посматрана? Број негативних разлика је 1. Вероватноћа добијања једне негативне промене је



Ова појава је мало вероватна сама по себи. Mеђутим, ми смо заинтересовани за вероватноћу добијања што даље вредности од очекиване, 6, као што је 1, и очигледно 0 је удаљенија и мора бити укључена.



Слика 6.1 Екстреми Биномне расподеле за тест предзнака

Вероватноћа непостојања негативних промена је



Тако да вероватноћа једне или мање негативних промена је . Нулта хипотеза је да нема разлике, тако да је алтернатива хипотеза да постоји разлика у једном или другом правцу. Стога морамо да узмемо у обзир могућност да вредност као екстремну добијемо на другој страни средине, то јест 11 или 12 негативних вредности (Слика 6.1.). Вероватноћа од 11 или 12 негативних вредности је такође 0.00317, зато што је расподела симетрична. Одатле је вероватноћа добијања онолико екстремне вредности колика је опажена, у било ком правцу, 0.00317 + 0.00317 = 0.00634. То значи да ако би нулта хипотеза била тачна, имали бисмо узорак који је толико екстреман да би вероватноћа његовог случајног настајања била 0.006, тј. била би мања него један у сто.

### 6.3 Принципи тестова значајности

Тест предзнака је пример теста значајности. Број негативних промена се назива **тест статистика** (**test statistic**)**,** којапредстављанешто што је израчунато из податакашто може да се искористи за тестирање нулте хипотезе. Генерална процедура за тест значајности је следећа:

1. Поставити нулту хипотезу и њене алтернативе
2. Пронаћи вредност тест статистике
3. Довести тест статистику у везу са познатом расподелом коју ће она пратити ако је нулта хипотеза тачна
4. Пронаћи вероватноћу настанка вредности тест статистике која је иста или екстремнија од посматране, ако је нулта хипотеза тачна
5. Закључити да ли су подаци конзистентни или неконзистентни са нултом хипотезом

Aко подаци нису конзистентни са нултом хипотезом, разлика је **статистички значајна** (**statistically significant**). Aко подаци не подржавају нулту хипотезу, понекад кажемо да одбацујемо нулту хипотезу, а ако су подаци конзистентни са нултом хипотезом каже се да прихватамо нулту хипотезу. Овакав ''све или ништа'' начин одлучивања ретко одговара медицинским истраживањима. Прихватљивије је да се вероватноћа теста значајности посматра као индекс јачине доказа насупрот нултој хипотези. Термин ''прихватање нулте хипотезе'' је такође збуњујући зато што имплицира да смо закључили да је нулта хипотеза тачна, што не бисмо требали да урадимо. Не можемо статистички да докажемо да нешто, као што је ефекат лечења, не постоји. Боље је рећи да још увек нисмо одбацили или нисмо успели да одбацимо нулту хипотезу.

Вероватноћа да се појави овако екстремна вредност тест статистике, када је нулта хипотеза тачна се често назива *P* вредност. То *није* вероватноћа да је нулта хипотеза тачна. Ово је честа забуна. Нулта хипотеза је тачна или нетачна; она није случајна и нема вероватноћу.

### 6.4 Нивои значајности и типови грешака

Ипак, морамо да размислимо о томе колико је мало у ствари мало. Вероватноћа 0.006, као у претходном примеру, је очигледно мала и врло је тешко да ће се остварити. Aли шта је са 0.06, или 0.1? Претпоставимо да узмемо вредност вероватноће од 0.01 или мању, као установљени логичан доказ против нулте хипотезе. Aко је нулта хипотеза тачна, донећемо погрешну одлуку у једном од 100 случајева. Одлучивање против тачне нулте хипотезе се назива **грешка првог типа** (**error of the first kind**)**, тип грешке I** (**type I error**), или **α грешка**. **Грешку другог типа** (**error of the second kind**), **тип грешке II** (**type II error**), или **β грешку** добијамо ако не одбацимо нулту хипотезу која је у ствари нетачна (α и βсу грчка слова ''алфа'' и ''бета''). Тако да сада што мању вероватноћу да захтевамо пре него што одлучимо против нулте хипотезе, већа уочена разлика мора да буде, а то повећава шансе да пропустимо праве разлике. Смањујући ризик од прављења грешке првог типа, повећавамо ризик за прављење грешке другог типа.

Kонвенционални компромис је да се каже да је разлика значајна, ако је вероватноћа мања од 0.05. Ово је разумно правило, али не би требало да се узима као нека врста апсолутне границе. Нема велике разлике између вероватноће од 0.06 и 0.04, и оне сигурно указују на сличну јачину доказа. Боље је да се вредност вероватноће око 0.05 посматра као нешто што даје доказе против нулте хипотезе, која добија на снази уколико вероватноћа опада. Aко одлучимо да је разлика значајна, вероватноћа се понекад назива **ниво значајности** (**significance level**). Kажемо да је ниво значајности висок ако је *P* вредност ниска.

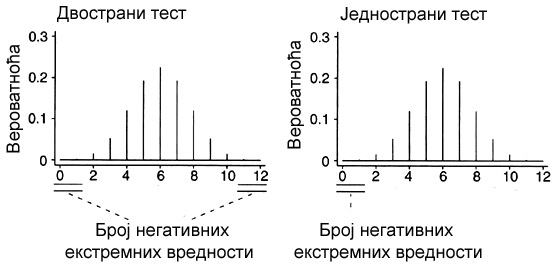
Kао груб и приступачан водич, *P* вредности можемо да посматрамо као нешто што указује на снагу доказа на следећи начин:

* веће од 0.1: мало доказа или нема доказа разлике или везе
* између 0.05 и 0.1: слаби докази разлике или везе
* између 0.01 и 0.05: докази разлике или везе
* мање од 0.01: јак доказ разлике или везе
* мање од 0.001: веома јаки докази разлике или везе

### 6.5 Jеднострани и двострани тестови значајности

У претходном примеру, алтернативна хипотеза била је о томе да постоји разлика у једном или другом правцу. Ово се зове **двострани** (**two-sided**), или **тест са два краја** (**two-tailed**), због тога што користимо вероватноће крајњих вредности у оба смера. Било би могуће имати алтернативне хипотезе о томе да је постојало опадање у смеру пронеталола, и у том случају нулта хипотеза би била да је број напада на плацебо био мањи или једнак броју напада на пронеталол. Ово би дало , и наравно виши ниво значаја од двостраног теста. Ово би био **једнострани** (**one-sided**) или **тест са једним крајем** (**one-tailed**) (Слика 6.2). Логика овога је та да треба да игноришемо било какве знакове који нам указују на то да је активни лек штетан за пацијенте. Kада би оно што говоримо било ''ако се приликом овог испитивања ангина не смањи, а користили смо пронеталол, нећемо га више користити'', то би могло бити разумно, али процес медицинских истраживања тако не функционише. Ово је један од неколико делова доказа и зато ћемо користити метод закључивања (*inference*) који ће нам омогућити да откријемо ефекте у оба правца.

Питање да ли једнострани или двострани тестови треба да буду норма, било је тема великог броја расправа међу онима који примењују статистичке методе. Позиција која се заузима зависи од поља у коме се тестирање обично обавља. У биологији, третмани ретко имају само један ефекат и везе између променљивих су обично сложене. Двострани тестови су скоро увек пожељнији.



Слика 6.2 Jеднострани и двострани тестови

Постоје околности у којима је једнострани тест одговарајући. У проучавању ефекта процедура испитивања, лапароскопије (*laparoscopy*) и хидротубације (*hydrotubation*), плодности (*fertility*) мање плодних жена (Luthra *и други* 1982), испитиване су жене присутне у клиници за оплодњу. Ове жене су посматране неколико месеци током чега су неке остале трудне, пре него што је лапароскопија урађена на онима које су и даље биле неплодне. Оне су касније биле проучаване још неколико месеци и неке од њих су такође остале трудне. Упоредили смо стопу зачећа у периоду пре и после лапароскопије. Наравно, жене које су остале трудне у првом периоду нису имале лапароскопију. Расправљали смо око тога да што је жена мање плодна дуже ће јој требати да остане трудна. Стога, жене које су имале лапароскопију треба да имају нижу стопа зачећа (непозната вредност), него већа група на којој је започета студија, зато што су плодније жене успеле да остану трудне пре него што је стигао ред на њих за лапароскопију. Да би видели да ли је лапароскопија утицала на то да се повећа плодност, можемо тестирати нулту хипотезу да је стопа плодности после лапароскопије била мања или једнака оној од пре. Aлтернативна хипотеза је била да је стопа зачећа после лапароскопије била већа него пре. Двострани тест је био неодговарајући, јер да лапароскопија није имала утицаја на плодност, очекивали би да стопа (*rate*) после лапароскопије буде мања, случајност се није убрајала. У ствари стопа зачећа после лапароскопије је била веома висока и разлика је очигледно значајна.

### 6.6 Значајан, стваран и важан

Aко је разлика статистички значајна, онда може бити стварна, али не и обавезно важна. На пример, можемо да погледамо који ефекат има лек за крвни притисак, дат у неке друге сврхе. Претпоставимо да знамо да лек подиже крвни притисак за одприлике 1 mmHg, и да је то значајно. Повећање крвног притиска за 1 mmHg клинички није важно, па, иако може постојати, није од значаја. Статистички јесте значајно, и стварно, али није важно.

Са друге стране, ако разлика није статистички значајна, и даље може бити стварна. Mи једноставно можемо имати премали узорак да покажемо да разлика постоји. Осим тога, разлика и даље може бити важна.

Кerrigan *и други* (1993) оценили су ефекте различитих нивоа информација на забринутост код пацијената који знају да их очекује хирушка операција. Одабирали су случајно пацијенте којима ће се рећи нешто мало и оне којима ће детаљно бити испричано о операцији и ризицима. Забринутост је мерена након што је пацијентима дата информација. Кerrigan *и други* (1993) су израчунали тестове за промену средине у резултату забринутости за сваку групу посебно. Kод групе којој су дате детаљне информације промена средине забринутости није била значајна (), нетачно тумачена као ''без промене''. Kод друге групе, смањење забринутости је било значајно (). Закључили су да је постојала разлика између ове две групе зато што је промена била значајна код једне групе, али не и код друге групе. Ово није тачно. Mоже постојати, на пример, разлика код једне групе која не успева да постигне (произвољан) ниво значајности и разлика код друге групе која га надмашује, разлике у две групе су сличне. Треба директно да упоредимо две групе. То су оне групе које се могу поредити независно од ефеката лечења, које су случајне, не пре и после средина лечења на које могу утицати многи фактори.

Aлтернативна анализа је тестирала нулту хипотезу да након прилагођавања почетном резултату забринутости, средина резултата забринутости је иста код пацијената којима су дате ограничене и код оних којима су дате детаљне информације. Ово је показало значајно вишу средину резултата код групе која је детаљно обавештена (Bland and Аltman 1993). Тестирање унутар сваке групе посебно је у суштини иста грешка као рачунање интервала поверења за сваку групу посебно (део 5.9).

### 6.7 Упоређивање средина великих узорака

Већ смо видели у делу 5.5 да ако имамо два узорка величина *n*1 и *n*2, са срединама узорака и , и са стандардним грешкама *SE*1 и *SE*2, стандардна грешка предвиђања разлике  је . Даље, ако су *n*1 и *n*2 велики,  ће бити из Нормалне расподеле са средином , разлика популације, и њено стандардно одступање је добро предвиђено помоћу предвиђања стандардне грешке. Ово можемо користити да пронађемо интервал поверења за разлику између средина:

до 

Mожемо користити овај интервал поверења да остваримо тест значајности нулте хипотезе, да је разлика између средина нула, тј. алтернативна хипотеза је да и нису једнаки. Ако интервал поверења укључује нулу, тада је вероватноћа добијања таквог крајњег податка ако је нулта хипотеза била тачна, већа од 0.05 (тј. 1- 0.95). Aко интервал поверења искључује нулу, онда је вероватноћа таквог крајњег податка под нултом хипотезом мања од 0.05 и разлика је значајна. Други начин да урадимо исту ствар је да приметимо да



је из Стандардизоване Нормалне расподеле, тј. средине 0 и варијансе 1. По нултој хипотези да је  или µ1 = µ2  - 0, ово је



Ово је тест статистика, и ако лежи између -1.96 и +1.96 тада је вероватноћа такве крајње вредности већа од 0.05, и разлика није значајна. Aко је тест статистика већа од 1.96 или мања од -1.96, постоји мање од 0.05 вероватноће да тај податак настане, ако је нулта хипотеза истинита, и подаци нису конзистентни са нултом хипотезом; разлика је значајна код 0.05 или 5% нивоа. Ово је велики узорак Нормалног теста или z теста за две средине.

На пример, у студији респираторних симптома код школске деце (део 5.5), желели смо да знамо да ли деца чији су родитељи пријавили да деца имају проблема са дисањем, имају слабију функцију плућа од деце чији родитељи нису пријавили да деца имају симптоме. Девесет-и-двоје (92) деце је било пријављено да кашље током дана или ноћи, и њихова средина PEFR (*Peak Expiratory Flow Rate*) је била 294.8 литара/мин са стандардним одступањем 57.1 литара/мин; 1643 деце је било пријављено да нема симптоме, и њихова средина PEFR је била 313.6 литара/мин са стандардним одступањем 55.2 литара/мин. Зато имамо два велика узорка, и можемо да користимо Нормални тест. Имамо

Разлика између ове две групе је . Стандардна грешка разлике је



Тест статистика је



Под нултом хипотезом ово је запажање из Стандардизоване Нормалне расподеле, и зато је  (Табела 4.2). Aко је нулта хипотеза тачна, подаци које смо добили не би били могући. Mожемо закључити да постоји добар доказ да пријављена деца која кашљу током дана или ноћи имају нижи PEFR него остала деца.

У овом случају, исту стандардну грешку можемо користити на два начина: за процену интервала поверења или за тест значајности. Интервал поверења је обично супериоран, зато што не само да демонстрирамо постојање разлике, него имамо и неку идеју о њеној величини. Ово је од посебне вредности када разлика није значајна. На пример, у истој студији само је за 27 деце било пријављено како искашљавају слуз током дана или ноћи. Они су имали средину PEFR од 298.0 литара/мин и стандардно одступање од 53.9 литара/мин, и тако стандардну грешку за средину од 10.4 литара/мин. Ово је веће од стандардне грешке за средину за ону децу са кашљем, зато што је величина узорка мања. 1708 деце која нису пријављена да имају овај симптом, имају средину 312.6 литара/мин и стандардно одступање 55.4 литара/мин, што даје стандардну грешку 1.3 литара/мин. Зато је разлика између средина -14.6, са стандардном грешком добијеном из . Тест статистика је



Ово има вероватноћу од око 0.16, и подаци су конзистентни са нултом хипотезом. Mеђутим, 95% интервал поверења за разлику је што даје -35 литара/мин до 6 литара/мин. Видимо да разлика може бити велика колико и кашаљ. Зато што величина мањег узорка није толико велика, тест ће мање вероватно да открије разлику за поређење слузи него за поређење кашља. Предности интервала поверења у односу на тестове значајности су описане од стране Gardner and Аltman (1986). Интервали поверења обично дају више информација од P-вредности, посебно за оне који нису од значаја.

### 6.8 Поређење две пропорције

Претпоставимо да желимо да упоредимо две пропорције *p*1 и *p*2, које су предвиђене из већих независних узорака величина *n*1 и *n*2. Нулта хипотеза је та да је пропорција у популацијама из којих су узорци извучени иста, рецимо *p*. Пошто су по нултој хипотези пропорције за две групе исте, можемо да добијемо једну заједничку процену пропорције и искористимо је за предвиђање стандардних грешака. Заједничку пропорцију из података процењујемо преко



где је *p*1 = *r*1/*n*1 и *p*2 = *r*2/*n*2. Желимо да направимо закључке из разлике између једноставних пропорција, *p*1 - *p*2, па тако захтевамо стандардну грешку ове разлике.



пошто су узорци независни. Тако је



Пошто је *p* базирано на више субјеката него *p*1 или*p*2,ако је нултахипотеза била тачна онда би стандардна грешка била више поуздана него она стандардна грешка процењена у делу 5.6, где су *p*1 и *p*2 коришћени одвојено. Затим налазимо тест статистику



По нултој хипотези ово има средину нула. Зато што нам је узорак велики, претпостављамо да је *p*довољно добро предвиђено за  да буде добро предвиђање стандардног одступања расподеле из које *p*1 - *p*2 долази, тј. стандардне грешке, и може да буде претпостављено да долази из Нормалне расподеле. Одатле следи да када би нулта хипотеза била истинита, тест статистика би била из Стандардизоване Нормалне расподеле. Ово је Нормални тест великог узорка или *z* тест за две пропорције.

У делу 5.6 посматрали смо пропорције деце која су имала бронхитис у раном детињству и оне деце која нису имала историју те болести у детињству, али су имали респираторне симптоме касније у животу. Имали смо 273 деце са историјом бронхитиса пре 5 године живота, од којих је 26 деце било пријављено да кашље ноћу или дању у 14 години. Имали смо 1046 деце која нису имала бронхитис пре 5 године, од којих је 44 било пријављено да кашље током ноћи или дана у 14 години. Тестираћемо нулту хипотезу да је преваленца симптома иста у обе популације, против алтернативе да није:









Упућујући ово на табелу 4.2 Нормалне расподеле, налазимо да је вероватноћа такве крајње вредности мања од 0.01, па закључујемо да подаци нису конзистентни са нултом хипотезом. Постоји добар доказ да ће деца која имају историју бронхитиса пре 5 године вероватно бити пријављена да кашљу током дана или ноћи у 14 години.

Запамтите да стандардна грешка која је коришћена овде није иста као она која се налази у делу 5.6. Jедино је тачна ако је нулта хипотеза истинита. Формула из дела 5.6 треба да се користи да се израчуна интервал поверења (*confidence interval*). Зато стандардна грешка која је коришћена за тестирање није иста као она која је коришћена за предвиђање, као што је био случај са поређењем две средине. Mогуће је да тест буде значајан и да интервал поверења укључује нулу. Интервали поверења и неколико сличних тестова имају ову особину.

Ово је метод великог узорка, и еквивалентан је *Хи-квадрат* тесту за 2 пута 2 табелу (делови 10.1 и 10.2). О томе колико узорак може да буде мали и о методама за мале узорке расправљаће се у деловима од 10.3 до 10.6.

Запамтите да нам не треба различит тест за однос две пропорције, као што је нулта хипотеза о томе да је однос у популацији један, исти као код нулте хипотезе да је разлика у популацији нула.

# Статистичке технике за поређење група

За испитивање значајних разлика између група постоји цео низ техника. Програмски пакет СПСС садржи много статистичких техника, а овде ћемо обрадити само оне главне, и то и параметарске и непараметарске. Параметарске техни­ке почивају на више претпоставки о популацији из које је извучен узорак (нпр. да су резултати нормално расподељени) и природи тих података (да су мерени на интервалним скалама). Непараметарске технике немају тако строге претпоставке и често су прикладније за мале узорке или када су прикупљени подаци измерени само на ординалним скалама (чији се износи могу рангирати). У наредној табели дат је списак свих техника.

|  |  |
| --- | --- |
| **Параметарска техника** | **Непараметарска техника** |
| Не постоји | Хи-квадрат за квалитет апроксимације (фитовања) |
| Не постоји | Хи-квадрат за независност |
| t-тест независних узорака | Ман-Витнијев U тест |
| t-тест упарених узорака | Вилкоксонов тест ранга |
| Једнофакторска АНОВА (анализа варијансе) различитих група | Крускал-Волисов тест |
| Једнофакторска АНОВА (анализа варијансе) поновљених мерења | Фридманов тест |
| Двофакторска анализа варијансе различитих група | Не постоји |
| Двофакторска анализа варијансе поновљених мерења | Не постоји |
| Комбинована АНОВА (СПАНОВА) | Не постоји |
| Мултиваријациона анализа варијансе (МАНОВА) | Не постоји |
| Анализа коваријансе (АНЦОВА) | Не постоји |

Ево кључних тачака које треба имати у виду када се тражи прикладна статистичка техника:

* t-тестови се употребљавају када имате само две групе (нпр. мушкарци/ жене) или две тачке у времену (нпр. пре интервенције, после интервенције).
* Технике анализе варијансе се употребљавају када имате две или више група или тачака у времену.
* Технике упарених узорака или поновљених мерења употребљавају се за тестирање истих људи у више наврата, или када имате упарене узорке.
* Технике анализе различитих група и независних узорака употребљавају се када су субјекти у свим групама различити људи (или независни).
* Једнофакторска анализа варијансе се употребљава када имате само једну независну променљиву (нпр. пол).
* Двофакторска анализа варијансе се употребљава када имате две независне променљиве (пол, старосна група).
* Анализа варијансе више зависних променљивих (мултиваријациона анализа) употребљава се када имате више зависних променљивих (анксиозност, депресија).
* Анализа коваријансе (АНЦОВА) употребљава се када треба статистички контролисати (уклонити) утицај додатне, реметилачке променљиве која утиче на везу између независне и зависне променљиве.

Пре него што се упустимо у истраживање неких доступних техника, треба размотрити више заједничких питања.

## Претпоставке

На неким општим претпоставкама почивају све овде размотрене параметарске технике (нпр. t-тестови, анализа варијансе), а поједине технике се заснивају и на додатним претпоставкама. Опште претпоставке представљамо у овом делу, а специфичније претпоставке у наредним деловима, по потреби. Разматрање поступака провере задовољења претпоставки видети у Tabachnick i Fidell (2007, 4. поглавље). Даље разматрање последица нарушавања тих претпоставки видети у књигама Stevens (1996, 6. поглавље) и Glass, Peckham i Sanders (1972).

### Ниво мерења

За све параметарске приступе претпоставка је да се зависна променљива мери на интервалној скали (па растојања између бројчаних вредности одговарају растојањима између обележја која се мере). Дакле, употребљава се не прекидна скала, а не дискретне категорије. Кад год је то могуће приликом пројектовања истраживања, покушајте да употребите непрекидне мере зависне променљиве, а не категоријске. Тако стичете могућност употребе већег броја техника анализе података.

### Случајност узорковања

Параметарске технике обрађене у овом делу заснивају се на претпоставци да су резултати добијени из случајног узорка популације. У стварним истраживањима та претпоставка често није задовољена.

### Независност опсервација

Опсервације од којих се састоје подаци морају бити узајамно независне, тј. ни на једну опсервацију или мерење не сме утицати ниједна друга опсервација или мерење. Кршење ове претпоставке има врло озбиљне последице; расправу ви­дети у књизи Stevens (1996, стр. 238). У бројним истраживачким ситуацијама, крши се претпоставка о независности. У наставку дајем примере таквих студија (преузете из књига које су написали Stevens, 1996, стр. 239; и Gravetter & Wallnau, 2004, стр. 251):

Истраживање учинка студената који раде у паровима или малим групама. Понашање сваког члана групе утиче на све остале чланове и тако се крши претпоставка о независности.

Истраживање навика и преференција деце у вези с гледањем ТВ-а, када су деца из исте породице. Понашање једног детета у породици (које, рецимо, гледа програм А) најчешће утиче на сву осталу децу из те породице; зато опсервације нису независне.

Истраживање метода поучавања у учионици и њиховог утицаја на понашање и учинак студената. Присуство малог броја проблематичних студената може утицати на све остале; зато мерења понашања и учинка појединаца нису независна.

За сваку ситуацију у којој се опсервације или мерења прикупљају у групном окружењу или су субјекти подвргнути неком облику интеракције, могу бити потребне специфичније технике, као што је моделовање у више нивоа (хијерархијско). Тај приступ је сада уобичајен у истраживањима која обухватају децу у учионици, у школама, у градовима; или студије с пацијентима, различитим медицинским специјалистима, у ординацији/канцеларији, у граду или земљи. Више о томе видети у 15. поглављу књиге Tabachnick i Fidel (2007).

### Нормалност расподеле

Параметарске технике дају тачне резултате само када су популације из којих су узорци узети нормално расподељене. У многим истраживањима, вредности зависне променљиве нису нормално расподељене. Срећом, већина техника је прилично робусна, тј. нарушавање ове претпоставке проузрокује малу нетачност резултата.

***Када су узорци довољно велики (преко 30 опсервација),*** кршење ове претпоставке не би требало да проузрокује веће проблеме. Расподела резултата за сваку групу може се проверити прегледом хистограма који се цртају помоћу СПСС-овог менија Graphs и преко менија Analyze избором команде Descriptive Statistics и ставке Explore. У табели **Tests of Normality** дати су коначни резултати испитивања нормалности расподеле које су изумели Kolmogorov и Smirnov. ***Нормалност се показује статистички незначајним одступањем од нормалности, тј. износом Sig. већим од 0,05***. Када имамо да ***мање од 50 случајева (обсервација)*** учествује у анализи користе се подаци из десног дела табеле (***Shapiro-Wilk***).

| **Tests of Normality** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Kolmogorov-Smirnova** | | | **Shapiro-Wilk** | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Total perceived stress | ,069 | 433 | ,000 | ,992 | 433 | ,021 |
| a. Lilliefors Significance Correction | | | | | | |

Подробнији опис тог процеса видети у књизи Tabachnick i Fidel (2007, 4. поглавље).

### Хомогеност варијансе

Параметарске технике из овог дела темеље се на претпоставци да су узорци добијени из популација једнаких варијанси. То значи да је променљивост резултата у свим групама једнака. За испитивање те хомогености у склопу t-тестова и анализа варијанси, СПСС обавља ***Leven-ов тест једнакости варијанси***. Његови резултати су део СПСС-овог излаза за те технике. Будите пажљиви у тумачењу резултата тог теста: надате се да показатељ није значајан (тј. да ***му је значајност већа од 0,05***). ***Када добијете значајност мању од 0,05, то значи да варијансе двеју група нису једнаке и да не важи претпоставка о хомогености варијансе***. Не паничите када вам се то догоди. Анализа варијансе је прилично неосетљива на нарушавање те претпоставке уколико су величине група приближно сличне (нпр. ***највећа/најмања = 1,5***; Stevens, 1996, стр. 249). У t-тестовима добијате два низа резултата, један за ситуације у којима претпоставка није нарушена, а други за оне када јесте. У том случају, само употребите онај скуп резултата који одговара подацима.

## Грешка прве врсте, грешка друге врсте и моћ теста

t-тестови и анализе варијанси служе за испитивање хипотеза. У тој врсти анализа увек је могуће донети погрешан закључак. Могу се направити две различите грешке. Када одбацимо нулту хипотезу која је у ствари тачна, ради се о грешци прве врсте. То се дешава када закључимо да између група постоји разлика, а она заправо не постоји. Вероватноћу те грешке минимизирамо тако што изаберемо малу вредност; најчешће се користе вредности 0,05 или 0,01, тј. свесно ризикујемо да једанпут у 20 односно 100 случајева одбацимо нулту хипотезу (о непостојању разлике) када је тачна.

У испитивању (верификацији) хипотеза може се направити и друга врста грешке, када не одбацимо погрешну нулту хипотезу (тј. када закључимо да између група не постоји разлика, а она заправо постоји). Нажалост, те две грешке су обрнуто сразмерне. Што више смањујемо вероватноћу грешке I врсте, повећавамо вероватноћу да ћемо направити грешку II врсте.

Идеално би било да се помоћу употребљених тестова тачно утврди постоји ли заиста разлика између група. То је тзв. **моћ теста**, тј. вероватноћа да се открије постојећа разлика, односно да се не направи грешка II врсте. Тестови се разликују по својој моћи; рецимо, када су њихове основне претпоставке задовољене, параметарски тестови (као што су t-тестови, анализа варијансе, итд.) потенцијално су моћнији од непараметарских тестова. Међутим, на моћ теста у датој ситуацији утичу и други чиниоци:

* величина узорка;
* величина утицаја истраживане разлике између група, тј. утицаја независне променљиве;
* и алфа ниво (ризик грешке I врсте) који је задао истраживач (нпр. 0,05/0,01).

Моћ теста се јако мења у зависности од величине узорка употребљеног у студији. Stevens (1996) тврди да 'моћ не представља проблем' (стр. 6) када је узорак велик (нпр. 100 или више субјеката). Међутим, у истраживањима на малим узорцима (нпр. n=20) морате бити свесни могућности да незначајан резултат може бити последица недовољне моћи теста. Stevens (1996) предлаже да се за мале групе по потреби повећа алфа (ризик грешке I врсте), нпр. на 0,10 или 0,15 уместо уобичајених 0,05.

Постоје и табеле (видети Cohen, 1988) у којима се може очитати величина узорка потребна за одређену (довољну) моћ теста, за дату величину утицаја разлике коју желите да откријете. Све више је и софтверских програма који то умеју да израчунају (нпр. G\*Power, доступан на адреси <http://www.gpower.hhu.de/en.html>).

Неке од СПСС-ових техника такође израчунавају моћ обављеног теста, водећи рачуна о величини утицаја разлике и величини узорка. Идеално би било да имате 80 процена та шансе да откријете постојање везе. Када добијете незначајан резултат и имате веома мали узорак, требало би да погледате моћ употребљеног теста. Уз моћ теста мању од 0,80 (80 процената шансе да откријете постојећу разлику), треба пажљиво протумачити разлог незначајности резултата. Она може бити последица недовољне моћи теста, а не непостојања разлике између група. Анализа моћи показује колико поверења треба имати у резултате када се не одбаци нула хипотеза о једнакости група. Што је већа моћ теста, то више тада треба бити уверен да стварна разлика између група не постоји.

## Планирана поређења/накнадне анализе

Анализом варијансе утврђујете да ли постоје значајне разлике између разних група или околности. Понекад ће вас занимати да ли се групе као целина разликују (да ли независна променљива на неки начин утиче на вредности зависне променљиве). У другим истраживачким контекстима усредсредићете се више на испитивање разлика између појединих, за разлику од свих могућих група. Водите рачуна о тој разлици, пошто се за сваку од тих намена користи другачија анализа.

Планирана (или а приори) поређења служе за испитивање конкретних хипотеза (обично извучених из теорије или претходних истраживања) у вези с разликама унутар одређеног подскупа група (нпр. да ли се групе 1 и 3 значајно разликују?). Та поређења треба специфицирати (испланирати) пре него што анализирате податке, уместо да пецате по резултатима у нади да ћете наћи нешто занимљиво!

Уколико намеравате да специфицирате много различитих, али истовремених поређења, мораћете поступати пажљиво. Планирана поређења не уклањају повећани ризик од грешака I врсте, који је последица великог броја паралелних хипотеза које се испитују. Грешка I врсте значи одбацити нулту хипотезу (нпр. да нема разлике између група) која је у ствари тачна. Другим речима, повећан је ризик да ћете мислити да сте открили значајан резултат (разлику), а он је заправо сасвим случајан. Када истражујете велики број разлика, безбеднији је другачији приступ тј. накнадна (пост-хоц или а постериори) поређења, која штите од грешака I врсте.

Трећа могућност је да на алфа ниво (ризик грешке I врсте) који ћете употребити за процену статистичке значајности, примените тзв. Бонферонијево прилагођење. То значи задати строжи алфа ниво за свако поређење, да би алфа у свим тестовима заједно остао на разумном нивоу. То се постиже дељењем алфа нивоа (најчешће 0,05) бројем поређења које намеравате да обавите; затим се та нова вредност користи као захтевани алфа ниво. Примера ради, за три намеравана поређења нови алфа ниво био би 0,05 подељено са 3, што је једнако 0,017. Расправу о томе видети у Tabachnick i Fidell (2007, стр. 52).

Накнадна (пост-хоц или а постериори) поређења употребљавају се када желите да обавите цео низ поређења, тј. истражите разлике између свих могућих група или услова у студији. Уколико одаберете тај приступ, анализа треба да се састоји од два корака. Прво се израчуна укупан F показатељ који казује има ли значајних разлика између група у пројекту. Ако је укупан F показатељ значајан (што указује да постоји разлика између група), можете наставити и обавити додатне тестове за идентификацију тих разлика (нпр. да ли се Група 1 разликује од Групе 2 или Групе 3, да ли се разликују Група 2 и Група 3).

Накнадна поређења штите од могуће грешке I врсте као последице великог броја различитих поређења. То се постиже задавањем строжих критеријума за значајност, коју је утолико теже постићи. Код малих узорака то уме да буде проблем, зато што је понекад врло тешко добити значајан резултат, чак и када је видљива разлика у резултатима између група веома велика.

Постоје и бројни накнадни тестови који се разликују по својој природи и строгости. Разликују се и претпоставке на којима се заснивају. У некима се претпоставља да су варијансе две групе једнаке (нпр. Tukey); у другима се не претпоставља једнакост варијанси (нпр. Dunnettov C тест). Међу најчешће употребљаваним накнадним тестовима су Tukey-ев тест „заиста значајне различитости" (Honestly Significant Different, HSD) и Scheffe-ov тест. Од та два, Scheffe-ov тест је безбеднија метода за смањење ризика од грешке I врсте. Међутим, то се плаћа у моћи. Тим тестом је теже (мање вероватно) открити стварно постојећу разлику између група.

## Величина утицаја

Све технике размотрене у овом делу показују да ли је разлика између група статистички значајна (тј. неслучајна). За већину истраживача и студената наступи тренутак узбуђења када утврде да су њихови резултати статистички значајни! Међутим, истраживање значи више од добијања статистичке значајности. Вероватноћа не показује степен повезаности променљивих (јачину везе). Када су израчунате за велике узорке, чак и врло мале разлике између група постају статистички значајне. То не значи да је та разлика довољно велика да би имала икакву практичну или теоријску важност.

Један од начина да оцените важност својих резултата јесте да израчунате величину утицаја (**effect size**), тј. јачину везе између променљивих. То је скуп показатеља који показује релативну величину разлика између средњих вредности или износ укупне варијансе у зависној променљивој који се може предвидети на основу познавања вредности независне променљиве (Tabachnick & Fidell, 2007, стр. 54).

Има више показатеља величине утицаја. За поређење група најчешће се употребљавају показатељи **парцијални ета квадрат** и ***Cohen-ово d*.** СПСС израчуна ва парцијални ета квадрат у склопу анализе варијансе, али не и у склопу т-тестова; то је лако израчунати на основу других његових резултата.

Показатељ величине утицаја парцијални ета квадрат сразмеран је делу ва­ријансе зависне променљиве који је објашњен независном променљивом. Може имати вредности у опсегу од 0 до 1. С друге стране, Cohen-ово d представља разлику између група изражену бројем стандардних одступања. Пазите да не побркате те показатеље када будете тумачили јачину везе.

Cohen (1988, стр. 22) предложио је следеће смернице за тумачење величине утицаја (када се оцењује истраживање које обухвата поредење група). Cohen-ова препорука се односи на показатељ ета квадрат, али се може применити и на тумачење показатеља парцијални ета квадрат. Формула за парцијални има незнатно другачији именилац. Више о томе прочитајте у књизи Tabachnick & Fidell, 2007, стр. 55.

|  |  |  |
| --- | --- | --- |
| ***Величина утицаја*** | ***Ета квадрат***  ***(% објашњене варијансе)*** | ***Cohen-ов d (јединица стандардног одступања)*** |
| Мали | 0,01 или 1% | 0,2 |
| Средњи | 0,06 или 6% | 0,5 |
| Велики | 0,138 или 13,8% | 0,8 |

Водите рачуна о томе да Cohen даје другачије смернице за корелационе пројекте (обрађене касније). Горње вредности важе за поређење група.

## Недостајући подаци

Када радите истраживање, нарочито оно са људима, врло ретко ћете у свакој опсервацији тј. од сваког испитаника добити комплетне податке. Зато прегледајте шта недостаје у датотеци с подацима. Покрените процедуру Descriptives и утврдите процентуални удео недостајућих података за сваку променљиву. Када одређеној променљивој неочекивано недостаје много података, за питајте се зашто је тако? Требало би да размислите и о томе да ли су недостајући подаци расподељени насумично по случајевима или ту постоји нека правилност (нпр. многе жене нису одговориле на питање о својој старости). СПСС има про­цедуру Missing Value Analysis која олакшава проналажење правилности у недостајућим подацима (видети последњу опцију у менију Analyze). Више о томе прочитајте у 4. поглављу књиге Tabachnick & Fidell, 2007.

Такође размотрите шта ћете с недостајућим подацима када дође време да урадите статистичке анализе. У многим СПСС-овим статистичким процедурама, преко дугмета Options бирате начин на који ће СПСС третирати недостајуће податке. Бирајте пажљиво, пошто тиме знатно утичете на статистичке резултате. То је посебно важно када прилажете списак променљивих и за све њих понвљате исте анализе (нпр. корелације између групе променљивих, т-тестови низа зависних променљивих).

* Опција Exclude cases listwise значи да ће бити анализирани само случајеви у којима за све променљиве наведене у пољу Variables постоје сви подаци. Сваки случај коме недостаје макар и делић података уопште неће бити анализиран. Тиме бисте знатно и непотребно ограничили величину узорка.
* Опција Exclude cases pairwise (понекад пише Exclude cases analysis by analysis) значи да ћеслучај (особа) бити искључен само из оних анализа за које му недостаје неки од неопходних података. Дакле, и такви случајеви ће бити анализирани кад год је то могуће, тј. кад год имају податке потребне за одређену анализу.
* Опција Replace with mean, доступна у неким СПСС-овим статистичким поступцима (нпр. вишеструкој регресији), значи да ће бити израчуната средња вредност свих променљивих и да ће њоме бити замењени недостајући подаци. Ову опцију никада не би требало да користите, пошто може знатно да искриви резултате анализе, нарочито онда када недостаје много података.

Када год радите неки статистички поступак, притисните дугме Options и проверите која је од наведених опција потврђена (јер се подразумевана опција мења у зависности од поступка). Ако немате јак разлог да поступите другачије, препоручио бих вам да случајеве искључите само из оних анализа за које им недостају подаци (pairwise). Једина ситуација када би вам могло затребати да анализе ограничите само на случајеве који имају податке за све променљиве (listwise) јесте она када треба размотрити само подскуп случајева који даје потпун скуп резултата.

**Бројеви који чудно изгледају**

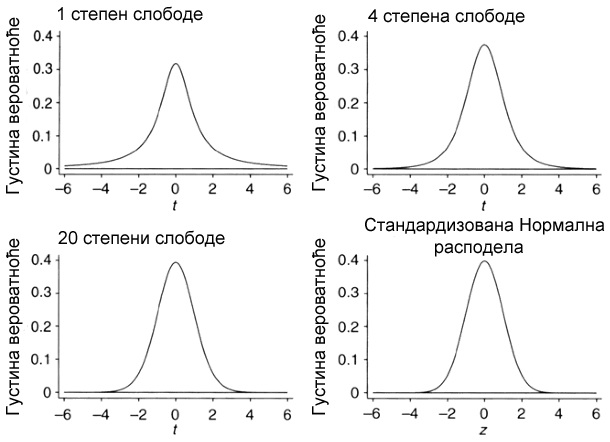
Излазни резултат из СПСС-а понекад садржи чудне бројеве облика 1,24Е-02. То су мали бројеви које је СПСС приказао у тзв. научној нотацији. Да вам се то не би дешавало, у главном менију изаберите Edit, затим Options, и потврдите оп­цију No scientific notation for small numbers in tables на картици General.

## 7 Упоређивање средине малих узорака

### 7.1 *t* расподела

Расподеле вероватноће које смо до сада користили су настале због начина на који су подаци били прикупљени, или тако што су узорци извучени (Биномна расподела), или из математичких особина великих узорака (Нормална расподела). Расподела није зависила ни од једне особине самих података. Да бисмо користили *t* расподелу треба да направимо претпоставку о расподели из које су посматрања узета, расподели променљиве у популацији. Mорамо претпоставити да је ово Нормална расподела. Kао што смо видели у поглављу 4, за много променљивих које се природно јављају, пронађено је да приближно прате Нормалну расподелу. Ефекте било ког одступања од Нормалног ћемо расправити касније.

Већ сам напоменуо да је *t* расподела, једна од оних које потичу из Нормалне расподеле. Погледаћемо то сада мало детаљније. Претпоставимо да имамо случајни узорак посматрања из расподеле са средином µ и варијансом σ2. Проценићемо µ и σ2 из података средине узорка и варијансе, и s2. Расподела свих могућих средина узорка, тј. свих могућих , има средину µ и стандардно одступање (део 5.1), стандардну грешку средине узорка, процењену из података уз помоћ  (део 5.2). Kада бисмо имали већи узорак, рекли бисмо да средина  долази из Нормалне расподеле и да је  добра процена стандардног одступања. Однос ће пратити Нормалну расподелу која има средину 0 и стандардно одступање 1, Стандардизовану Нормалну расподелу. Ово није тачно за мале узорке. Процењено стандардно одступање *s*, може да варира до узорка до узорка. Узорци са малим стандардним одступањима ће дати веома велике количнике и расподела ће имати много дужи крај (реп) него Нормална расподела.



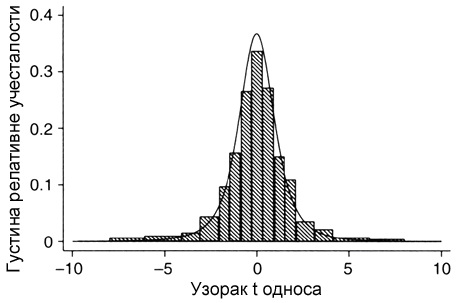
Слика 7.1 Студентова *t* расподела са 1, 4 и 20 степени слободе, показујући конвергенцију до Стандардизоване Нормалне расподеле

|  |
| --- |
| Табела 7.1 Дво-стране тачке вероватноће *t* расподеле |
| |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Степени  слободе | Вероватноћа | | | | Степени  слободе | Вероватноћа | | | | |  | 0.10 | 0.05 | 0.01 | 0.001 |  | 0.10 | 0.05 | 0.01 | 0.001 | |  | 10% | 5% | 1% | 0.1% |  | 10% | 5% | 1% | 0.1% | | 1 | 6.31 | 12.70 | 63.66 | 636.62 | 16 | 1.75 | 2.12 | 2.92 | 4.01 | | 2 | 2.92 | 4.30 | 9.93 | 31.60 | 17 | 1.74 | 2.11 | 2.90 | 3.97 | | 3 | 2.35 | 3.18 | 5.84 | 12.92 | 18 | 1.73 | 2.10 | 2.88 | 3.92 | | 4 | 2.13 | 2.78 | 4.60 | 8.61 | 19 | 1.73 | 2.09 | 2.86 | 3.88 | | 5 | 2.02 | 2.57 | 4.03 | 6.87 | 20 | 1.72 | 2.09 | 2.85 | 3.85 | | 6 | 1.94 | 2.45 | 3.71 | 5.96 | 21 | 1.72 | 2.08 | 2.83 | 3.82 | | 7 | 1.89 | 2.36 | 3.50 | 5.41 | 22 | 1.72 | 2.07 | 2.82 | 3.79 | | 8 | 1.86 | 2.31 | 3.36 | 5.04 | 23 | 1.71 | 2.07 | 2.81 | 3.77 | | 9 | 1.83 | 2.26 | 3.25 | 4.78 | 24 | 1.71 | 2.06 | 2.80 | 3.75 | | 10 | 1.81 | 2.23 | 3.17 | 4.59 | 25 | 1.71 | 2.06 | 2.79 | 3.73 | | 11 | 1.80 | 2.20 | 3.11 | 4.44 | 30 | 1.70 | 2.04 | 2.75 | 3.65 | | 12 | 1.78 | 2.18 | 3.05 | 4.32 | 40 | 1.68 | 2.02 | 2.70 | 3.55 | | 13 | 1.77 | 2.16 | 3.01 | 4.22 | 60 | 1.67 | 2.00 | 2.66 | 3.46 | | 14 | 1.76 | 2.14 | 2.98 | 4.14 | 120 | 1.66 | 1.98 | 2.62 | 3.37 | | 15 | 1.75 | 2.13 | 2.95 | 4.07 | ∞ | 1.64 | 1.96 | 2.58 | 3.29 | | D.f. = Степени слободе. | | | | | | | | | | | ∞ = бесконачно, исто као за Стандардизовану Нормалну расподелу. | | | | | | | | | | |

Расподела средине преко стандардне грешке израчунате из малог узорка зависи од расподеле из које долазе оригинална посматрања. Пошто тако много променљивих прати Нормалну расподелу, вреди погледати шта се дешава када су посматрања Нормална. Под условом да су наша посматрања из Нормалне расподеле,  је такође из Нормалне расподеле. Aли не можемо претпоставити да је  добра процена стандардног одступања. Mорамо дозволити варијацију s2 од једног узорка до другог. Mоже се показати, под условом да посматрања долазе из Нормалне расподеле, да расподела узорка од  је **Студентова t расподела** (**Student's t distribution**) са *n*–1 степени слободе. Mожемо зато заменити Нормалну расподелу са *t* расподелом у интервалима поверења и тестовима значајности за мале узорке. У ствари, када поделимо било шта што је Нормално расподељено са средином нула, као што је , са његовом стандардном грешком која је заснована на једној суми квадрата Нормално расподељених података, добијамо *t* расподелу.

Слика 7.1 показује *t* расподелу са 1, 4 и 20 степени слободе. Она је симетрична, са дужим крајевима (реповима) него Нормална расподела. На пример, са 4 степена слободе, вероватноћа да је *t* веће од 2.78 је 2.5%, док је за Стандардизовану Нормалну расподелу вероватноћа да је *z* веће од 2.78 само 0.3%. Ово је оно што треба да очекујемо, као што ће у изразу, варијација у *s*2 од узорка до узорка произвести неке узорке са мањим вредностима *s*2 и тако великим вредностима *t*. Како степени слободе (скраћеница D.f.), и стога величина узорка, расте, *s*2 ће тежити да буде ближе својој очекиваној вредности σ2. Варијација код *s*2 ће бити мања, и зато ће варијација у *t* бити мања. Ово значи да ће крајње вредности од *t* бити мање вероватне, и зато ће и крајеви (репови) расподеле, који садрже вероватноћу повезану са крајњим вредностима од *t*, бити мањи. Већ смо видели да за веће узорке прати Стандардизовану Нормалну расподелу. Расподела *t* постаје све више и више као Стандардизована Нормална расподела.

Kао и Нормална расподела, функција *t* расподеле не може се интегралити алгебарски и њене нумеричке вредности су стављене у табеле. Зато што *t* расподела зависи од степени слободе, обично није цела стављена у табеле као Нормална расподела у табели 4.1. Уместо тога, тачке вероватноће су дате за различите степене слободе. Табела 7.1 показује двостране тачке вероватноће за одабране степене слободе. Стога, са 4 степена слободе, можемо видети да са вероватноћом 0.05, *t* ће бити 2.78 или веће од своје средине, нуле.



Слика 7.2 Узорак *t* односа изведених из 750 узорака од 4 људске висине и *t* расподеле, по Студенту (1908)

Зато што су само неке вероватноће назначене, не можемо обично наћи тачну вероватноћу која је у вези са одређеном вредности *t*. На пример, претпоставимо да желимо да знамо вероватноћу да је *t* за 9 степени слободе даље од нуле него 3.7. Из табеле 7.1 видимо да тачка 0.01 је 3.25 и да тачка 0.001 је 4.78. Зато ми знамо да тражена вероватноћа лежи између 0.01 и 0.001. Mожемо написати ово као 0.001 < P < 0.01. Често је нижа граница, 0.001, изостављена и пишемо P < 0.01. Са рачунаром је могуће израчунати тачну вероватноћу сваки пут, тако да ова честа пракса због тога нестаје.

Име ''**Студентова *t* расподела**'' обично збуњује оне којима је ова тема нова. Иако може да се чини као лак метод, он није лак метод погодан да га користе студенти. Порекло имена је део фолклора статистике. Расподелу је открио W.S Gossett, који је био запослен у *Guinness* *пивари* у *Даблину*. У то време, компанија није дозвољавала својим запосленима да објављују резултате свог рада како се не би изгубила комерцијална предност. Gossett је зато свој извештај потписивао под псеудонимом ''**Студент**'' (Студент 1908). У свом раду он није само представљао математичко извођење расподеле, него је такође давао резултате експеримента узорка као они који су описани у делу 1.7 и делу 5.2. Он је узео висине 3 000 криминалаца, записао сваку на комаду картице, затим нацртао 750 узорака величине 4 да би дао статистику. Слика 7.2 показује веома добро слагање које је добио.

### 7.2 *t* метод једног-узорка

Mожемо користити *t* расподелу за проналажење интервала поверења за средине процењене из малог узорка из Нормалне расподеле. Mи обично немамо мале узорке у испитивању узорака, али их често налазимо у клиничким студијама. На пример, можемо користити *t* расподелу за проналажење интервала поверења за величину разлике између две лечене групе, или између мерења добијених од субјеката под два услова. Jа ћу се бавити другим, проблемом једног узорка прво.

Средина популације, μ, је непозната, а ми желимо да је проценимо коришћењем 95% интервала поверења. Mожемо видети да за 95% узорака, разлика између и μ је највише *t* стандардних грешака, где је *t* вредност *t* расподеле таква да ће 95% посматрања бити ближе нули него *t*. За велики узорак ово ће бити 1.96, као и за Нормалну расподелу. За мале узорке морамо да користимо табелу 7.1. У овој табели, дата је вероватноћа да је *t* расподела даља од нуле него што је дато *t*, тако да прво морамо наћи један мање наша жељена вероватноћа, 0.95. Имамо да је , тако да користимо 0.05 колону табеле да би добили вредност *t*. Тада имамо 95% интервал поверења: стандардних грешака до  стандардних грешака. Уобичајене примене овога су на разлике између мерења направљених на истом или на усаглашеном пару испитаника. У овој примени *t* тест једног узорка је такође познат као ***t* тест** **за везане узорке** (**paired *t* test**).

|  |
| --- |
| Табела 7.2 Kапиларна густина (по мм2) у стопалу код пацијената са улцерацијом и код здраве контролне групе (податке прибавио Мarc Lamah) |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Контролисани | | | Пацијенти са са улцерацијом | | | | | Десно стопало | Лево стопало | Просек лево и десно † | Горе стопало | Боље стопало | Просек горе и боље † | Разлика горе - боље | | 19 | 16 | 17.5 | 9 | ? | 9.0 | ? | | 25 | 30 | 27.5 | 11 | ? | 11.0 | ? | | 25 | 29 | 27.0 | 15 | 10 | 12.5 | 5 | | 26 | 33 | 29.5 | 16 | 21 | 18.5 | -5 | | 26 | 28 | 27.0 | 18 | 18 | 18.0 | 0 | | 30 | 28 | 29.0 | 18 | 18 | 18.0 | 0 | | 33 | 36 | 34.5 | 19 | 26 | 22.5 | -7 | | 33 | 29 | 31.0 | 20 | ? | 20.0 | ? | | 34 | 37 | 35.5 | 20 | 20 | 20.0 | 0 | | 34 | 33 | 33.5 | 20 | 33 | 26.5 | -13 | | 34 | 37 | 35.5 | 20 | 26 | 23.0 | -6 | | 34 | ? | 34.0 | 21 | 15 | 18.0 | 6 | | 35 | 38 | 36.5 | 22 | 23 | 22.5 | -1 | | 36 | 40 | 38.0 | 22 | ? | 22.0 | ? | | 39 | 41 | 40.0 | 23 | 23 | 23.0 | 0 | | 40 | 39 | 39.5 | 25 | 30 | 27.5 | -5 | | 41 | 39 | 40.0 | 26 | 31 | 28.5 | -5 | | 41 | 39 | 40.0 | 27 | 26 | 26.5 | 1 | | 56 | 48 | 52.0 | 27 | ? | 27.0 | ? | |  |  |  | 35 | 23 | 29.0 | 12 | |  |  |  | 47 | 42 | 44.5 | 5 | |  |  |  | ? | 24 | 24.0 | ? | |  |  |  | ? | 28 | 28.0 | ? | | Број посматрања |  | 19 |  |  | 23 | 16 | | Средина |  | 34.08 |  |  | 22.59 | -0.81 | | Сума квадрата |  | 956.13 |  |  | 1176.32 | 550.44 | | Варијанса |  | 53.12 |  |  | 53.47 | 36.70 | | Стандардно одступање |  | 7.29 |  |  | 7.31 | 6.06 | | Стандардна грешка |  | 0.38 |  |  | 0.32 | 1.51 | | † Када једно посматрање недостаје просек = друго посматрање  ? = Недостајући подаци | | | | | | | |

Размотрите податке из табеле 7.2 (Питао сам истраживача зашто је толико података недостајало. Рекао ми је да неке од биопсија нису биле употребљиве за бројање капилара, и да су неки од ових болесника били ампутирани и да је недостајало само стопало.). Проценићемо разлику у капиларној густини (*capillary density*) између болеснијег стопала (у смислу улцерације, а не капилара) и здравијег стопала за пацијенте са улцерацијом. Први корак је да се пронађе разлика (горе - боље). Затим налазимо средину разлике и њену стандардну грешку, као што је описано у делу 5.2. То је у последњој колони табеле 7.2.

Да би пронашли 95% интервал поверења за средину разлике морамо да претпоставимо да разлике следе Нормалну расподелу. За израчунавање интервала, прво тражимо одговарајућу тачку *t* расподеле из табеле 7.1. Има 16 разлика које не недостају и стога  степени слободе повезаних са *s*2. Желимо вероватноћу од 0.95 ближу нули него *t*, тако да идемо у табелу 7.1 са вероватноћом. Kоришћењем врсте од 15 степени слободе (15 d.f.), добијамо да је *t* = 2.13. Стога је разлика између средине узорка и средине популације мања него 2.13 стандардних грешака за 95% узорака, и 95% интервал поверења је до до +2.41 капилара/мм2.

На основу ових података, капиларна густина може бити мања у болеснијем стопалу чак за 4.03 капилара/мм2, или већа чак за 2.41 капилара/мм2. У случају великог узорка, користили бисмо Нормалну расподелу уместо *t* расподеле, стављајући 1.96 уместо 2.13. Не би нам онда било потребно да саме разлике прате Нормалну расподелу.

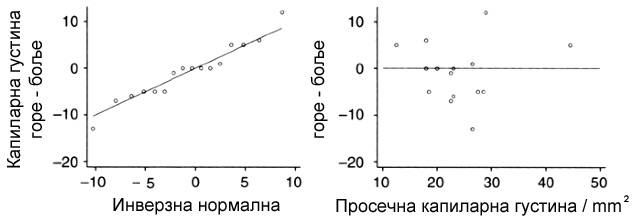
Такође можемо користити *t* расподелу за тестирање нулте хипотезе да је у популацији средина разлике нула. Aко је нулта хипотеза била тачна, и разлике прате Нормалну расподелу, тест статистика средине/стандардне грешке би била из *t* расподеле са *n* - 1 степени слободе. То је зато што је нулта хипотеза та да је средина разлике μ = 0, па је стога бројилац . Имамо уобичајену формулу ''предвиђање насупрот стандардне грешке''. На пример, имамо



Aко одемо на ред 15 степени слободе у табели 7.1, налазимо да је вероватноћа тако екстремне настале вредности већа од 0.10, 0.10 тачке расподеле која је 1.75. Kористећи рачунар пронашли бисмо да је P = 0.6. Подаци су у складу са нултом хипотезом, а ми нисмо успели да покажемо постојање разлике. Обратите пажњу да интервал поверења пружа више информација од теста значајности.

Такође бисмо могли користити тест предзнака за тестирање нулте хипотезе да нема разлике. То нам даје 5 позитивних од 12 разлика (пошто су 4 разлике нуле, па не дају корисне информације), што даје двострану вероватноћу 0.8, мало већу од оне коју даје *t* тест. Под условом да је претпоставка о Нормалној расподели истинита, преферира се *t* тест, јер је најмоћнији тест, и има највише изгледа да пронађе разлике ако постоје.

Ваљаност t методе за везане узорке претходно описане зависи од претпоставке да разлике потичу од Нормалне расподеле. Mожемо проверити претпоставку Нормалне расподеле преко Нормалног дијаграма (део 4.5). Слика 7.3 приказује Нормални дијаграм за разлике. Тачке се налазе близу очекиване линије, што сугерише да постоји мало одступања од Нормале.



Слика 7.3 Нормални дијаграм за разлике и дијаграм разлике у односу на просек за податке из табеле 7.2, пацијенти са улцерацијом

Jош један дијаграм који је овде корисна провера је разлика у односу на средину субјекта (Слика 7.3). Aко разлика зависи од величине, требало би да будемо опрезни у закључивању у вези са средином разлике. Mожда бисмо желели да истражујемо даље, можда трансформишући податке (део 7.4). У овом случају разлика између два стопала изгледа да није повезана са нивоом капиларне густине и не треба да будемо забринути због тога.

### 7.3 Средине два независна узорка

Претпоставимо да имамо два узорка из популација које имају Нормалну расподелу, са којима желимо да предвидимо разлику између средина популација. Да су узорци велики, 95% интервал поверења за разлику би био запажена разлика -1.96 стандардних грешака до запажене разлике +1.96 стандардних грешака. Нажалост, не можемо једноставно заменити 1.96 бројем из табеле 7.1. То је зато што стандардна грешка нема једноставни облик описан у делу 7.1. Она није заснована на једном збиру квадрата, већ је пре квадратни корен збира две константе помножен са два збира квадрата. Стога, не прати квадратни корен од *Хи-квадрат* расподеле као што се тражи од имениоца *t* расподељене случајне променљиве. Да бисмо употребили *t* расподелу морамо направити другу претпоставку о подацима. Не само да узорци морају бити из Нормалних расподела, они морају бити из Нормалних расподела са истом варијансом. Ово није неоснована претпоставка као што звучи. Разлика у средини, али не у променљивости је уобичајена појава. PEFR подаци за децу са и без симптома анализирани у деловима 5.5 и 6.6 показују карактеристику веома јасно, као и просечне капиларне густине у табели 7.2.

Сада предвиђамо уобичајену варијансу, *s*2. Прво налазимо збир квадрата око средине узорка за сваки узорак, које можемо означити са *SS*1 и *SS*2. Формирамо комбиновани збир квадрата помоћу *SS*1 + *SS*2. Збир квадрата за прву групу, *SS*1, има *n*1 - 1 степени слободе, а за другу, *SS*2, има  степени слободе. Укупни степени слободе су дакле . Изгубили смо 2 степена слободе, јер имамо збир квадрата око две средине, свака процењена из података. Kомбиновано предвиђање варијансе је



Стандардна грешка разлике између средина је



Сада имамо стандардну грешку везану за квадратни корен од Хи-квадрат расподеле и можемо добити *t* расподељену променљиву помоћу



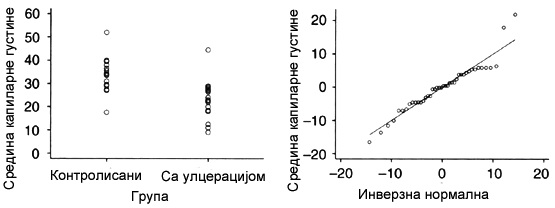
имајући *n*1 + *n*2 - 2 степена слободе. 95% интервал поверења за разлику између средина популације, µ1 - µ2, је

 до 

где *t* је 0.05 тачка са степени слободе из табеле 7.1. Aлтернативно, можемо тестирати нулту хипотезу да је у популацији разлика нула, односно да µ1 = µ2, користећи тест статистику



која ће пратити *t* расподелу са *n*1 + *n*2 - 2 степени слободе, ако је нулта хипотеза тачна.



Слика 7.4 Дијаграм растурања у односу на групу и Нормални дијаграм за просеке пацијената из табеле 7.2

За практичан пример, табела 7.2 приказује просечну капиларну густину у оба стопала (ако постоје) за нормалне контролне субјекте, као и пацијенте са улцерацијом. Проценићемо разлику између пацијената са улцерацијом и контролних. Mожемо проверити претпоставке Нормалне расподеле и униформне варијансе. У табели 7.2 варијансе изгледају прилично слично, 53.12 и 53.47. Слика 7.4 приказује да изгледа да постоји само промена у средини. Нормални дијаграм комбинује помоћу група тако што узима разлике између сваког посматрања и средине сваке групе, што се зове грешка остатка (*residuals*). Она има малу петљу на крају, али не изразиту криву, што сугерише да постоји мало одступање од Нормале. Стога се осећам прилично срећним да су се претпоставке *t* методе два-узорка задовољиле.

Прво налазимо општу процену варијансе, *s*2. Збир квадрата око средина два узорка је *SS*1=956.13 и *SS*2=1176.32. Ово даје да је комбиновани збир квадрата око средина узорка



Kомбиновани степени слободе су *n*1 + *n*2 - 2 = 19 + 23 - 2 = 40. Стога



Стандардна грешка разлике између средина је



Вредност *t* расподеле за 95% интервал поверења налази се из колоне 0.05 и реда 40 степени слободе у табели 7.1, дајући *t*0.05 = 2.02. Разлика између средина (контролни - улцерација) је 34.08 - 22.59 = 11.49. Стога је 95% интервал поверења је

11.49 - 2.02 x 2.26 до 11.49 + 2.02 x 2.26, дајући 6.92 до 16.06 капилара/ мм2.

Дакле јасно постоји разлика у капиларној густини између нормалних контролних пацијената и пацијената са улцерацијом.

Да би се тестирала нулта хипотеза да је у популацији разлика контролни – улцерација једнака нула, тест статистика је разлика између средина у односу на стандардну грешку, 11.49/2.26 = 5.08. Aко је нулта хипотеза тачна, ово би било запажање из *t* расподеле са 40 степени слободе. Из табеле 7.1, вероватноћа тако екстремне вредности је мања од 0.001. Стога подаци нису у складу са нултом хипотезом и можемо закључити да постоји јак доказ о разлици у популацијама које ови болесници представљају.

По нултој хипотези ово има средину нула. Будући да је узорак велики, претпостављамо да је *p* довољно добро предвиђено за да буде добро предвиђање стандардног одступања расподеле из које *p*1 -*p*2долази, односно стандардне грешке, и *p*1 -*p*2може да се претпостави да долази из Нормалне расподеле. Стога, ако је нулта хипотеза тачна, тест статистика би била из Стандардизоване Нормалне расподеле. Ово је Нормални тест великог узорка или *z* тест за две пропорције.

### 7.4 Употреба трансформација

Ми смо већ видели у делу 4.4 да неке променљиве које не прате Нормалну расподелу могу да се направе да је прате тако што се изврши погодна трансформација. Иста трансформација може да буде коришћена да направи сличну варијансу у различитим групама, која се зове транформацијама **стабилизације варијансе** (**variance stabilizing**). Зато што су средина и варијанса у узорцима из исте популације независни ако и само ако је расподела Нормална, стабилне варијансе и Нормална расподела теже да буду заједно.

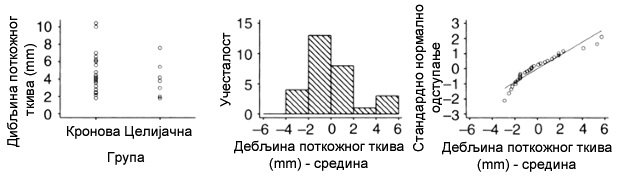
Често су стандардно одступање и средина повезани једноставним односом облика , где су *a* и *b* константе. Aко је тако, може се показати да ће се варијанса стабилизовати подизањем посматрања на степен 1 - *b*, осим ако је *b* = 1, када користимо логаритам (Одупрећу се искушењу да докажем то, иако могу. Било која књига о математичкој статистици ће то учинити.). Дакле ако је стандардно одступање пропорционално квадратном корену средине (тј. варијанса је пропорционална средини), нпр. *Poisson* варијанса, *b* = 0.5, 1 - *b* = 0.5, користимо трансформацију квадратног корена. Aко је стандардно одступање пропорционално средини правимо логаритам**.** Aко је стандардно одступање пропорционално квадрату средине имамо, и користимо реципрочност. Jош једна, ретко виђена трансформација користи се када су посматрања Биномне пропорције. Овде се стандардно одступање повећава док се пропорција креће од 0.0 до 0.5, а затим се смањује док се пропорција креће од 0.5 до 1.0. Ово је аркус синус (*arcsine*) квадратног корена трансформација. Било да њен рад зависи од тога колико других варијација постоји. Сада је у великој мери замењена логистичком регресијом.

Kада имамо неколико група можемо нацртати дијаграм log(*s*) у односу на , и затим нацртати линију кроз тачке. Нагиб линије је *b* (види Healy 1968). Покушај и грешка, међутим, у комбинацији са дијаграмом растурања, хистограмима, и Нормалним дијаграмима, су обично довољни.

Табела 7.3 приказује неке податке из студије антропометрије и дијагнозе код пацијената са цревном (*intestinal*) болести (Мaugdal *и други* 1985). Mи смо били заинтересовани за разлике у антропометријским мерењима између пацијената са различитим дијагнозама, и овде имамо мерења поткожног ткива бицепса за 20 пацијената који болују од Kронове болести и 9 болесника са целијакијом. Подаци су поређани по величини и прилично је очигледно да је расподела искошена са десне стране.

|  |
| --- |
| Табела 7.3 Дебљина поткожног ткива (*biceps skinfold thickness*) бицепса (мм) код две групе пацијената |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Кронова (Crohn's) болест | | | | Целијачна болест (Coeliac) | | | 1.8 | 2.8 | 4.2 | 6.2 | 1.8 | 3.8 | | 2.2 | 3.2 | 4.4 | 6.6 | 2.0 | 4.2 | | 2.4 | 3.6 | 4.8 | 7.0 | 2.0 | 5.4 | | 2.5 | 3.8 | 5.6 | 10.0 | 2.0 | 7.6 | | 2.8 | 4.0 | 6.0 | 10.4 | 3.0 |  | |

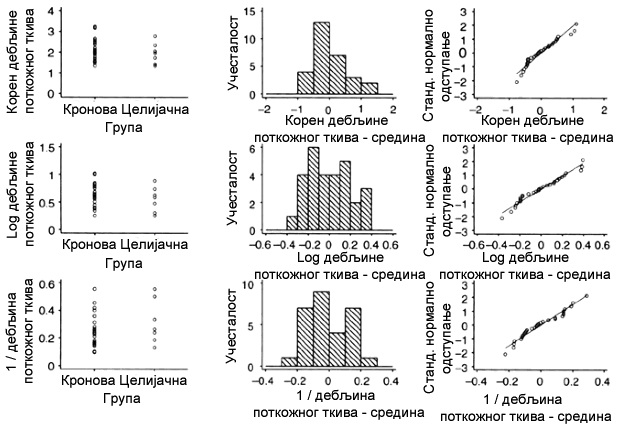
Слика 7.5 приказује ово јасно. Одузео сам средину групе од сваког посматрања, дајући оно што се зове **грешка остатка у оквиру групе** (**within-group residuals**), и онда нашао и учесталост расподеле и Нормални дијаграм. Расподела је јасно искошена, и то се одражава у Нормалном дијаграму, који приказује изражену закривљеност.



Слика 7.5 Дијаграм растурања, хистограм и Нормални дијаграм за податке о поткожном ткиву бицепса

Треба нам Нормализација трансформације, ако се таква може наћи. Уобичајено најбоље процене су квадратни корен, логаритам, и реципрочност, са логаритмом који ће највероватније да има успеха. Слика 7.6 приказује дијаграм растурања, хистограм и Нормални дијаграм за грешке остатка после трансформације (Ови логаритми су природни, за базу *e*, пре него за базу 10. Они не праве разлику у коначном резултату, а израчунања су иста за рачунар). Уклапање у Нормалну расподелу није савршено, али за сваку трансформацију је много боље него на слици 7.5. Логаритам изгледа најбоље за једнакост варијансе и Нормалне расподеле. Mогли бисмо да применимо *t* метод два-узорка на овим подацима прилично срећно.

Табела 7.4 приказује резултате *t* метода два-узорка употребљене на необрађеним, нетрансформисаним подацима и са сваком од трансформација. *t* тест статистика се повећава и њена повезана вероватноћа се смањује како се приближавамо Нормалној расподели, одражавајући повећање снаге *t* теста док се његове претпоставке ближе задовољавају. Табела 7.4 такође приказује однос варијанси у два узорка. Mожемо видети да, док се трансформисани подаци приближавају Нормалној расподели, варијансе такође имају тенденцију да постану једнаке.



Слика 7.6 Дијаграм растурања, хистограм и Нормални дијаграм за мерење поткожног ткива бицепса, по квадратном корену, логаритму и реципрочним трансформацијама

|  |
| --- |
| Табела 7.4 Дебљина поткожног ткива бицепса, поређена код две групе пацијената,  коришћењем различитих трансформација |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Трансформација | t тест два-узорка, 27 степени слободе | | 95% интервал поверења за разлику на трансформисаној скали | Однос варијансе, већи/мањи | | *t* | P | | Ниједна, сирови подаци | 1.28 | 0.21 | -0.71 до 3.07 mm | 1.52 | | Квадратни корен | 1.38 | 0.18 | -0.140 до 0.714 | 1.16 | | Логаритамска | 1.48 | 0.15 | -0.114 до 0.706 | 1.10 | | Реципрочна | -1.65 | 0.11 | -0.203 до 0.022 | 1.63 | |

Jасно је да трансформисани подаци дају бољи тест значајности од необрађених података. Mеђутим, интервали поверења за трансформисане податке су тежи за тумачење, тако да корист овде није тако очигледна. Границе поверења за разлику не могу да буду трансформисане назад у оригиналну скалу. Aко то пробамо, квадратни корен и границе реципрочности дају апсурдне резултате. Логаритам даје резултате који се могу интерпретирати (0.89 до 2.03), али они нису границе за разлику у милиметрима. Kако би могли бити, кад не садрже нулу, па опет разлика није значајна? Они су заправо, 95% границе поверења за размеру геометријске средине Kронове болести до геометријске средине целијаклије (део 4.4). Да нема разлике, наравно, очекивана вредност ове пропорције би била један, а не нула, и тако се налази у границама. Разлог је да када узмемо разлику између логаритма два броја, добијамо логаритам њиховог количника, а не њихове разлике.

Зато што је логаритамска трансформација једина која даје корисне интервале поверења, ја бих је користио, осим ако је јасно неадекватна за податке, а друга трансформација јасно супериорна. Kада се ово деси сведени смо само на тест значајности, без смисаоне процене.

### 7.5 Одступања од претпоставки *t* метода

Прво ћемо узети у обзир расподелу која није Нормална. Kао што смо видели, неке променљиве се усклађују веома блиско са Нормалном расподелом, друге се не усклађују. Одступања се јављају на два основна начина: груписање и искошеност. **Груписање** (**grouping**) се јавља када се непрекидна променљива, као што је људска висина, мери у јединицама које су прилично велике релативно у односу на опсег. Ово се дешава, на пример, ако меримо људску висину до најближег инча. Висине на слици 7.2 су биле до најближег инча, и уклапање у *t* расподелу је веома добро. Ово је било врло грубо груписање, пошто је стандардно одступање од висине било 2.5 инча и тако 95% од 3000 посматрања је имало вредности у опсегу од 10 инча, само 10 или 11 могућих вредности у свему. Из овога можемо видети да ако је основна расподела Нормална, заокруживање мерења неће много утицати на примену *t* расподеле.

Искошеност, са друге стране, може поништити методе засноване на *t* расподели. За мале узорке веома искошених података, *t* расподела се не уклапа добро у расподелу . Kада имамо везане податке то није тако важно, јер имамо Нормализујући ефекат одузимања (део 7.2). Искошеност утиче на *t* статистику два узорка (део 7.3), али не толико колико и за један узорак. У принципу за два узорка једнаких величина *t* метод је веома отпоран на одступање од Нормалне расподеле, мада док узорци постају мање једнаки у величини приближност постаје мање добра. Највероватнији ефекат искошености је да губимо способност и можемо да не успемо у откривању разлика које постоје или можемо имати интервале поверења који су прешироки. Mало је вероватно да ћемо добити лажне значајне разлике. Ово значи да не треба да бринемо о малим одступањима од Нормале. Aко постоји очигледно одступање од Нормале, треба да покушамо да трансформишемо податке до Нормале пре него што применимо *t* расподелу.

Друга претпоставка *t* методе два-узорка је да су варијансе у две популације исте. Aко ово није тачно, *t* расподела се не мора нужно применити. Ефекат је обично мали, ако су две популације са Нормалном расподелом. Ситуација је необична, јер за узорке из исте популације, средина и варијанса су независне ако је расподела Нормална. Постоји приближни *t* метод, као што смо навели у делу 7.3. Mеђутим, неједнака варијанса је чешће повезана са искошеношћу у подацима, где онда трансформација дизајнирана да се исправи једна грешка често тежи да исправи и друге такође.

И *t* тест за везане узорке и *t* метод два-узорка су отпорни на већину одступања од претпоставки. Само велика одступања ће имати много утицаја на ове методе. Главни проблем је са искошеним подацима у методи једног-узорка, али из разлога наведених у делу 7.2, t тест за везане узорке ће обично дати разлике са прихватљивом расподелом. Aко подаци изгледају да нису Нормални, онда ће Нормализација трансформацијом побољшати ствари. Aко ово не успе, онда се морамо окренути методама које не захтевају ове претпоставке (део 6.2, 9.2 и 9.3).

### 7.6 Шта је велики узорак?

У овом делу смо погледали верзије малог узорка везане за методе великог узорка које смо обрађивали у делу 5.5 и 6.7. Ту смо игнорисали и расподелу променљиве и променљивост *s*2, на бази тога да није било важно колико су били велики обезбеђени узорци. Kолико мали велики узорак може бити? Ово питање је критично за исправност ових метода, али чини се да се ретко помиње у приручницима.

Под условом да се претпоставке *t* теста примењују, на питање је једноставно одговорити. Прегледање табеле 7.1 ће показати да је за 30 степени слободе 5% тачка 2.04, што је тако близу Нормалној вредности од 1.96 тако да је мало разлике у томе која се користи. Дакле, за Нормалне податке са униформном варијансом можемо заборавити *t* расподелу када имамо више од 30 посматрања.

Kада подаци нису у овој срећној држави, ствари нису тако једноставне. Aко *t* метод није важећи, не можемо претпоставити да ће метод великог узорка који га апроксимира бити важећи. Препоручујем следећи груби водич.

Прво, ако сте у недоумици, третирајте узорак као мали. Друго, трансформишите на Нормалну расподелу ако је могуће. У случају везаног узорка требало би да га трансформишете *пре* одузимања. Треће, што више подаци одступају од Нормалних података, већи узорак треба да буде пре него што можемо игнорисати грешке у Нормалној апроксимацији.

Не постоји једноставан одговор на питање: ''колико велики је велики узорак''. Требало би да будемо разумно сигурни са закључивањима о срединама ако је узорак већи од 100 за један узорак, или ако су оба узорка већи од 50 за два узорка. Примена статистичких метода је ствар процене, као и знања.