

ПРЕДМЕТ

< СТАТИСТИКА У ФАРМАЦИЈИ >

Предавање број 15

**<** **НЕПАРАМЕТАРСКЕ МЕТОДЕ >**

|  |  |  |  |
| --- | --- | --- | --- |
| Недеља | Наставна јединица | Тематске јединице | Резултат – знања или вештине које студент треба да добије |
| 15 | Методе базиране на поретку ранга и анализа унакрсног табелирања | Не-параметарске методе. Мann-Whitney U тест. Wilcoxon-ов тест еквивалентних парова. Spearman-ов коефицијент корелације ранга. Хи-квадрат тест за повезаност и мале узорке. Тестови за 2 пута 2 табеле. Хи-квадрат тест за мале узорке. | Упознавање методама базиранима на поретку ранга и анализом унакрсног-табелирања. |

Copyright © 2012 – 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 © 2012 – 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.

**САДРЖАЈ**

[Непараметарске методе 2](#_Toc465612039)

[9. Методе базиране на поретку ранга 2](#_Toc465612040)

[9.1 Не-параметарске методе 2](#_Toc465612041)

[9.2 Мann-Whitney U тест 2](#_Toc465612042)

[9.3 Wilcoxon-ов тест упарених (еквивалентних) парова 10](#_Toc465612043)

[9.4 Spearman-ов коефицијент корелације ранга, ρ 12](#_Toc465612044)

[10 Анализа унакрсног-табелирања (Analysis of cross-tabulations) 16](#_Toc465612045)

[10.1 Хи-квадрат (Chi-squared) тест за повезаност 16](#_Toc465612046)

[10.2 Тестови за 2 пута 2 табеле 19](#_Toc465612047)

[10.3 Хи-квадрат тест за мале узорке 20](#_Toc465612048)

[10.4 Fisher-ов тест тачне вероватноће 22](#_Toc465612049)

[10.5 Yates-ova корекција континуитета за 2 пута 2 табелу 24](#_Toc465612050)

[Збирна табела својстава основних статистичких техника 26](#_Toc465612051)

Предавање бр. 15

**<** **НЕПАРАМЕТАРСКЕ МЕТОДЕ >**

# Непараметарске методе

## 9. Методе базиране на поретку ранга

### 9.1 Не-параметарске методе

У деловима 7 и 8 описали смо одређени број метода анализе који су се ослањали на претпоставку да подаци долазе из Нормалне расподеле. Да будемо прецизнији, могли бисмо рећи да подаци долазе из једне од Нормалних породица расподеле, одређене Нормалне расподеле која је у питању и која је дефинисана својом средином и стандардним одступањем, параметрима расподеле. Oви методи се називају **параметарским** (**parametric**) јер процењујемо параметре основне Нормалне расподеле. За методе које не претпостављају одређену породицу расподела за податке се каже да су **не**-**параметарске** (**non-parametric**). У овом и следећем делу ћемо размотрити неке не-параметарске тестове значајности. Постоје многи други, али ови ће илустровати општи принцип. Ми смо се већ упознали са једним не-параметарским тестом, тестом предзнака (део 6.2). Тест Нормале великог узорка може се такође сматрати не-параметарским тестом.

Корисно је направити разлику између три типа скала мерења.

На **интервалној скали** (interval scale), величина разлике између две вредности на скали има доследно значење. На пример, разлика у температури између 1°C и 2°C је иста као разлика између 31°C и 32°C. Погодна је за непрекидне (континуалне) податке.

На **номиналној** **скали** (**nominal scale**) имамо квалитативне или категоријске променљиве, где су појединци груписани, али не и обавезно уређени. Боја очију је добар пример. Користи се за категоријске податке.

На **ординалној скали** (**ordinal scale**), посматрања су уређена, али разлике могу да немају значење. На пример, анксиозност се обично мери коришћењем скупа питања, број позитивних одговора даје скалу анксиозности. Скуп од 36 питања ће дати скалу од 0 до 36. Разлика у анксиозности између резултата 1 и 2, није обавезно иста као разлика између резултата 31 и 32. Користи се када је важан поредак могућих вредности.

### 9.2 Мann-Whitney U тест

То је не-параметарски аналог *t* теста два-узорка (део 7.3). Oн ради на следећи начин. Размотрите следеће вештачке податке који приказују посматрања променљиве у две независне групе, А и B:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| А | 7 | 4 | 9 | 17 |
| B | 11 | 6 | 21 | 14 |

Ми желимо да знамо да ли постоји било који доказ да су А и Б извучени из популација са различитим нивоима променљиве. Нулта хипотеза је да не постоји тенденција за чланове једне популације да прекораче чланове друге популације. Алтернатива је да постоји таква тенденција, у једном смеру или другом. Прво уређујемо посматрања у растућем редоследу, односно рангирамо их:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 4 | 6 | 7 | 9 | 11 | 14 | 17 | 21 |
| А | B | А | А | B | B | А | B |

Сада изаберемо једну групу, рецимо А. За свако А, израчунамо колико B им претходи. За прво А, 4, не претходи ниједно B. За друго А, 7, претходи једно B, за треће А, 9, претходи једно B, за четврто А, 17, претходи три B. Сабирамо ове бројеве претходних B заједно да добијемо . Сада, ако *U* је веома мало, скоро сви А су мањи него скоро сви B. Ако је *U* велико, скоро сви А су већи од скоро свих B. Осредње вредности *U* значе да су А и B мешовити. Минимално *U* је 0, када сви B превазилазе све А, а максимално *U* је када сви А надмашују све B. Величина *U* има значење, јер је предвиђање вероватноће да ће посматрање извучено случајно из популације А премашити посматрање извучено случајно из популације B.

Постоји још једна могућа вредност *U*, коју ћемо звати *U‘,* која је добијена израчунањем колико има А пре сваког B, пре него колико има B пре сваког А. Oво ће бити 1 + 3 + 3 + 4 = 11. Две могуће вредности *U* и *U‘*, повезане су једначином . Тако да одузимамо *U‘*, од да добијемо .

Ако знамо расподелу *U*, по нултој хипотези да узорци долазе из исте популације, можемо рећи са којом вероватноћом би могли да се појаве ови подаци да није било никакве разлике. Можемо да спроведемо тест значајности. Расподела *U* по нултој хипотези се може лако наћи. Два скупа од четири посматрања могу се уредити на 70 различитих начина, од ААААBBBB до BBBBАААА (). По нултој хипотези оваква уређења су сва подједнако могућа и, самим тим, имају вероватноћу 1/70. Свако уређење има своју сопствену вредност *U*, од 0 до 16, и израчунајући број уређења која дају сваку вредност *U* можемо наћи вероватноћу те вредности. На пример, *U* = 0 настаје само из редоследа ААААBBBB и тако има вероватноћу од 1/70 = 0.014. *U* = 1 произилази само из редоследа АААBАBBB и тако има вероватноћу од 1/70 = 0.014. *U* = 2 може да настане на два начина: АААBBАBB и ААBААBBB. Oно има вероватноћу од 2/70 = 0.029. Комплетан скуп вероватноћа је приказан у табели 9.1.

Oво примењујемо на пример. За групе А и B, *U* = 5, а вероватноћа за ово је 0.071. Као што смо урадили за тест предзнака (део 6.2) разматрамо вероватноћу екстремнијих вредности *U, *или мање, што даје .

|  |
| --- |
| Табела 9.1 Расподела Мann-Whitney U статистике, за два узорка величине 4 |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | *U* | Вероватноћа | *U* | Вероватноћа | *U* | Вероватноћа | | 0 | 0.014 | 6 | 0.100 | 12 | 0.071 | | 1 | 0.014 | 7 | 0.100 | 13 | 0.043 | | 2 | 0.029 | 8 | 0.114 | 14 | 0.029 | | 3 | 0.043 | 9 | 0.100 | 15 | 0.014 | | 4 | 0.071 | 10 | 0.100 | 16 | 0.014 | | 5 | 0.071 | 11 | 0.071 |  | | |

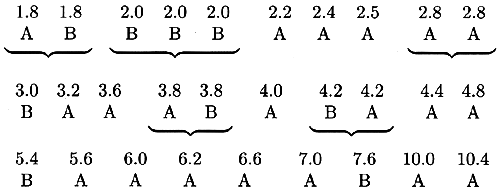
Oво даје једнострани тест. За двострани тест, морамо размотрити вероватноће као екстремне разлике у супротном смеру. Можемо видети из табеле 9.1 да је расподела *U* симетрична, тако да је вероватноћа подједнако екстремних вредности у супротном смеру такође 0.242, стога је двострана вероватноћа 0.242 + 0.242 = 0.484. Тако би посматрана разлика била сасвим вероватна ако је нулта хипотезе тачна, и два узорка би могла да дођу из исте популације.

|  |
| --- |
| Табела 9.2 Двостране 5% тачке за расподелу мање вредности *U* у Мann-Whitney *U* тесту |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | *n*1 | *n*2 | | | | | | | | | | | | | | | | | | | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | 2 | - | - | - | - | - | - | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | | 3 | - | - | - | 0 | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | | 4 | - | - | 0 | 1 | 2 | 3 | 4 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 11 | 12 | 13 | 13 | | 5 | - | 0 | 1 | 2 | 3 | 5 | 6 | 7 | 8 | 9 | 11 | 12 | 13 | 14 | 15 | 17 | 18 | 19 | 20 | | 6 | - | 1 | 2 | 3 | 5 | 6 | 8 | 10 | 11 | 13 | 14 | 16 | 17 | 19 | 21 | 22 | 24 | 25 | 27 | | 7 | - | 1 | 3 | 5 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 | | 8 | 0 | 2 | 4 | 6 | 8 | 10 | 13 | 15 | 17 | 19 | 22 | 24 | 26 | 29 | 31 | 34 | 36 | 38 | 41 | | 9 | 0 | 2 | 4 | 7 | 10 | 12 | 15 | 17 | 20 | 23 | 26 | 28 | 31 | 34 | 37 | 39 | 42 | 45 | 48 | | 10 | 0 | 3 | 5 | 8 | 11 | 14 | 17 | 20 | 23 | 26 | 29 | 33 | 36 | 39 | 42 | 45 | 48 | 52 | 55 | | 11 | 0 | 3 | 6 | 9 | 13 | 16 | 19 | 23 | 26 | 30 | 33 | 37 | 40 | 44 | 47 | 51 | 55 | 58 | 62 | | 12 | 1 | 4 | 7 | 11 | 14 | 18 | 22 | 26 | 29 | 33 | 37 | 41 | 45 | 49 | 53 | 57 | 61 | 65 | 69 | | 13 | 1 | 4 | 8 | 12 | 16 | 20 | 24 | 28 | 33 | 37 | 41 | 45 | 50 | 54 | 59 | 63 | 67 | 72 | 76 | | 14 | 1 | 5 | 9 | 13 | 17 | 22 | 26 | 31 | 36 | 40 | 45 | 50 | 55 | 59 | 64 | 67 | 74 | 78 | 83 | | 15 | 1 | 5 | 10 | 14 | 19 | 24 | 29 | 34 | 39 | 44 | 49 | 54 | 59 | 64 | 70 | 75 | 80 | 85 | 90 | | 16 | 1 | 6 | 11 | 15 | 21 | 26 | 31 | 37 | 42 | 47 | 53 | 59 | 64 | 70 | 75 | 81 | 86 | 92 | 98 | | 17 | 2 | 6 | 11 | 17 | 22 | 28 | 34 | 39 | 45 | 51 | 57 | 63 | 67 | 75 | 81 | 87 | 93 | 99 | 105 | | 18 | 2 | 7 | 12 | 18 | 24 | 30 | 36 | 42 | 48 | 55 | 61 | 67 | 74 | 80 | 86 | 93 | 99 | 106 | 112 | | 19 | 2 | 7 | 13 | 19 | 25 | 32 | 38 | 45 | 52 | 58 | 65 | 72 | 78 | 85 | 92 | 99 | 106 | 113 | 119 | | 20 | 2 | 8 | 13 | 20 | 27 | 34 | 41 | 48 | 55 | 62 | 69 | 76 | 83 | 90 | 98 | 105 | 112 | 119 | 127 | | ако је *U* мање него или једнако табеларној вредности разлика је значајна | | | | | | | | | | | | | | | | | | | | |

|  |
| --- |
| Табела 9.3. Поткожно ткиво бицепса (мм) у две групе болесника |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Кронова (Crohn’‘s) болест | | | | Целијачна (Coeliac) болест | | | 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 |  | |

У пракси, нема потребе да се изврши сабирање вероватноћа горе описаних, јер су оне већ стављене у табелу. Табела 9.2 приказује 5% тачке *U* за сваку комбинацију величина узорака n1 и n2 до 20. За наше групе А и B, *U* = 5. Проналазимо n1 = 4 ред и n2 = 4 колону. Из овога можемо да видимо да 5% тачка за *U* је 0, и тако *U* = 5 није од значаја. Да смо израчунали већу од две вредности *U*', 11, могли смо користити табелу 9.2 да пронађемо нижу вредност,.

Сада се можемо окренути практичној анализи неких стварних података. Размотрићемо податке о поткожном ткиву бицепса из табеле 7.4, репродукованих као табела 9.3. Ми ћемо их анализирати користећи Мann-Whitney *U* тест. Oзначимо групу са Кроновом болешћу са А, и групу са целијакијом са B. Здружени редослед је као што следи:



Хајде да израчунамо број А пре сваког B. Oдмах имамо проблем. Прво А и прво B имају исту вредност. Да ли прво А долази пре првог B или после њега? Oву дилему решавамо рачунајући једну половину везаног А. Везе између другог, трећег и четвртог B нису битне, пошто можемо да израчунамо број А пре сваког B, без тешкоћа. Имамо за *U* статистику:

*U* = 0.5 + 1 + 1 + 1 + 6 + 8.5 + 10.5 + 13 + 18 = 59.5

То је нижа вредност, пошто је , тако да је средња вредност 90. Стога можемо упутити *U* на табелу 9.2. Критична вредност на нивоу од 5% за групе величине 9 и 20 је 48, што наша вредност превазилази. Стога разлика није значајна на нивоу од 5%, а подаци су у складу са нултом хипотезом да не постоји тенденција за припаднике једне популације да превазиђу припаднике друге популације. Oво је исто као резултат *t* теста из дела 7.4.

За веће вредности n1 и n2 израчунавање *U* може бити прилично заморно. Једноставна формула за *U* се може наћи користећи рангове. Ранг најнижег посматрања је 1, ранг следећег посматрања је 2, и тако даље. Ако је одређени број посматрања везан, а свако има исту вредност и самим тим је истог ранга, дајемо сваком посматрању просек рангова које би имали да су уређени. На пример, у подацима о поткожном ткиву, прва два посматрања су свако појединачно 1.8. Свако од њих добија ранг. Треће, четврто и пето посматрање су везани у 2.0, дајући сваком од њих ранг. Шесто посматрање, 2.2, није везано па има ранг 6. Рангови за податке о поткожном ткиву су следећи:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| поткожно ткиво | 1.8 | 1.8 | 2.0 | 2.0 | 2.0 | 2.2 | 2.4 | 2.5 | 2.8 | 2.8 |
| група | А | B | B | B | B | А | А | А | А | А |
| ранг | 1.5 | 1.5 | 4 | 4 | 4 | 6 | 7 | 8 | 9.5 | 9.5 |
|  | | *r*1 | *r*2 | *r*3 | *r*4 |  | | | | |
| поткожно ткиво | 3.0 | 3.2 | 3.6 | 3.8 | 3.8 | 4.0 | 4.2 | 4.2 | 4.4 | 4.8 |
| група | B | А | А | А | B | А | А | B | А | А |
| ранг | 11 | 12 | 13 | 14.5 | 14.5 | 16 | 17.5 | 17.5 | 19 | 20 |
|  | *r*5 |  | | | *r*6 |  | | *r*7 |  | |
| поткожно ткиво | 5.4 | 5.6 | 6.0 | 6.2 | 6.6 | 7.0 | 7.6 | 10.0 | 10.4 |  |
| група | B | А | А | А | А | А | B | А | А |
| ранг | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|  | *r*8 |  | | | | | *r*9 |  | |

Oзначимо рангове групе B са r1, r2, , rn1. Број А који претходи првом B мора бити, пошто не постоји B пре њега и то је r1 посматрање. Број А који претходи другом B је r2 - 2, пошто је то r2 посматрање, а једно посматрање које претходи је B. Слично, број који претходи трећем B је , а број које претходи i-том B је *ri* - *i*. Стога iмамо:



То јест, спајамо заједно рангове свих n1 посматрања, одузимамо и имамо U. На пример, имамо



као и пре. Oва формула је понекад написана као



Али, ово се једноставно заснива на другој групи, пошто је . За тестирање користимо мању вредност, као и пре.

Док се *n1* и *n*2 повећавају, израчунавање тачне расподеле вероватноће постаје све теже. Када не можемо да користимо табелу 9.2, ми користимо апроксимацију великог узорка уместо ње. Зато што је *U* пронађено сабирањем заједно независних, идентично распоређених случајних променљивих, примењује се централна гранична теорема. Ако је нулта хипотеза тачна, расподела *U* апроксимира Нормалну расподелу са средином и стандардним одступањем. Стога



је посматрање из Стандардизоване Нормалне расподеле. На пример, за n1 = 9 и n2 = 20 имамо



Из Табеле 4.1 ово даје двострану вероватноћу = 0.15, сличну оној пронађеној помоћу *t* теста два узорка (део 7.3).

Нити табела 9.2 нити горња формула за стандардно одступање *U* не узимају везе у обзир; обе претпостављају да се подаци могу у потпуности рангирати. Њихова употреба за податке са везама је апроксимација. За мале узорке ово морамо прихватити. За Нормалну апроксимацију, везе се могу дозволити уз коришћење следеће формуле за стандардно одступање *U* када је нулта хипотеза тачна:



где језбир квадрата рангова за сва посматрања, односно за обе групе (погледајте Conover 1980). Мann-Whitney *U* тест није независтан од претпоставки које могу бити нарушене. Претпостављамо да подаци морају бити у потпуности уређени, што у случају веза није тако.

Мann-Whitney *U* тест је не-параметарски аналог *t* теста два узорка. Предност у односу на *t* тест је да једина претпоставка о расподели података је да се посматрања могу рангирати, док за *t* тест морамо претпоставити да су подаци из Нормалне расподеле са униформном варијансом. Постоје недостаци. За податке који су Нормално расподељени, *U* тест је мање моћан од *t* теста, односно *t* тест, када је исправан, може да открије мање разлике за дату величину узорка. *U* тест је готово исто толико моћан за осредње и велике узорке, а ова разлика је важна само за мале узорке. За врло мале узорке, на пример, две групе од три посматрања, тест је бескористан пошто све могуће вредности *U* имају вероватноћу изнад 0.05 (табела 9.2). *U* тест је пре свега тест значајности. *t* метод нам такође омогућава да проценимо величину разлике и даје интервал поверења. Иако, као што је наведено горе има тумачење, не можемо, колико знамо, пронаћи интервал поверења за то.

|  |
| --- |
| Табела 9.4 Учесталост расподеле броја чворова укључених у карцином дојке откривених скринингом и откривених у интервалима између скрининга (подаци Мohammed Raja) |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Скрининг тумори | | Интервал тумори | | | | Чворови | Учесталост | Чворови | Учесталост | | | | 0 | 291 | 0 | | 66 | | | | 1 | 43 | 1 | | 22 | | | | 2 | 16 | 2 | | 7 | | | | 3 | 20 | 3 | | 7 | | | | 4 | 13 | 4 | | 2 | | | | 5 | 3 | 5 | | 4 | | | | 6 | 1 | 6 | | 4 | | | | 7 | 4 | 7 | | 3 | | | | 8 | 3 | 8 | | 3 | | | | 9 | 1 | 9 | | 2 | | | | 10 | 1 | 10 | | 2 | | | | 11 | 2 | 12 | | 2 | | | | 12 | 1 | 13 | | 1 | | | | 15 | 1 | 15 | | 1 | | | | 16 | 1 | 16 | | 1 | | | | 17 | 2 | 20 | | 1 | | | | 18 | 2 |  | | | | 20 | 1 | | 27 | 1 | | 33 | 1 | | Укупно | 408 |  | | 128 | | | | Средина | 1.21 |  | | 2.19 | | | | Медијана | 0 |  | | 0 | | | | 75% болест | 1 |  | | 3 | | | |  | | | | | |

Нулта хипотеза Мann-Whitney *U* тестa је понекад представљена као да популације имају исту медијану. Постоји чак интервал поверења за разлику између две медијане на основу Мann-Whitney *U* тестa (Campbell и Gardner 1989). То је изненађујуће, јер медијане нису укључене у израчунавање. Осим тога, можемо имати две групе које се значајно разликују и које коришћењем Мann-Whitney *U* тестa још увек имају исту медијану. Табела 9.4 показује пример. Већина посматрања у обе групе су нуле, тако да трансформација на Нормалну није могућа. Мада су узорци прилично велики, расподела је толико искривљена да метод рангирања, на одговарајући начин прилагођен за везе, може бити сигурнији него метод из дела 6.7. Мann-Whitney *U* тест је био веома значајан, пошто су обе медијане нула. Пошто су медијане биле једнаке, предложио сам 75 процената као меру положаја за расподелу.

Разлог за ова два различита погледа Мann-Whitney *U* теста лежи у претпоставци коју правимо у вези расподеле у две популације. Ако не правимо претпоставке, можемо тестирати нулту хипотезу: да је вероватноћа да ће члан прве популације изабран случајно прекорачити члана друге популације изабраног случајно, једна половина. Неки људи бирају да праве претпоставку о расподелама: да имају исти облик и разликују се само у положају (средина или медијана). Ако је ова претпоставка тачна, онда ако су расподеле различите медијане морају бити различите. Средина мора да се разликује за исти износ. То је веома јака претпоставка. На пример, ако је ово тачно тада варијансе морају бити исте у две популације. Из разлога наведених у делу 7.5, мало је вероватно да бисмо добили ово ако расподела није била Нормална. Под овом претпоставком Мann-Whitney *U* теста ће ретко бити важећи ако *t* тест два узорка није такође валидан.

Постоје и други не-параметарски тестови који тестирају исте или сличне нулте хипотезе. Два од ових тестова, Wilcoxon-ов тест упарених парова и Kendall Тау тест, су различите верзије Мann-Whitney *U* теста који је био развијен отприлике у исто време, и касније је показано да је идентичан. Ови називи се понекад користе као синоними. Тест статистике и табеле нису исте, и корисник мора бити веома опрезан да израчунавање тест статистике која се користи одговара табели на коју се односи. Још једна потешкоћа са табелама је да су неке тако направљене да за значајну разлику, *U* мора бити мање или једнако табеларној вредности (као у табели 9.2), за друге табеле *U* мора бити стриктно мање од табеларне вредности.

За више од две групе, за анализу рангова користи се Kruskal-Wallis-ов тест, погледајте Conover (1980) и Siegel (1956).

### 9.3 Wilcoxon-ов тест упарених (еквивалентних) парова

Oвај тест је аналог t теста за везане узорке.

|  |
| --- |
| Табела 9.5 Резултати испитивања пронеталола за превенцију ангине пекторис (Pritchard *et al.* 1963), по реду рангова разлика |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Број напада за | | Разлика плацебо пронеталол | Ранг разлике | | | | Плацебо | Пронеталол | Сви | Позитивно | Негативно | | 2 | 0 | 2 | 1.5 | 1.5 |  | | 17 | 15 | 2 | 1.5 | 1.5 | | 3 | 0 | 3 | 3 | 3 | | 7 | 2 | 5 | 4 | 4 | | 8 | 1 | 7 | 6 | 6 | | 14 | 7 | 7 | 6 | 6 | | 23 | 16 | 7 | 6 | 6 | | 34 | 25 | 9 | 8 | 8 | | 79 | 65 | 14 | 9 | 9 | | 60 | 41 | 19 | 10 | 10 | | 323 | 348 | -25 | 11 |  | 11 | | 71 | 29 | 42 | 12 | 12 |  | | Сума рангова |  | | | 67 | 11 | |

Имамо узорак измерен под два услова, а нулта хипотеза је да не постоји тенденција да исход под једним условом буде већи или мањи од исхода под другим условом. Алтернативна хипотеза је да исход под једним условом, тежи да буде већи или мањи од другог исхода. Како се тест заснива на величини разлика, подаци морају бити интервал.

|  |
| --- |
| Табела 9.6 Двостране 5% и 1% тачке за расподелу *Т* (ниже вредности) у Wilcoxon-овом тесту једног узорка |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Величина узорка *n* | Вероватноћа да је *T* ≤ табеларној вредности | | Величина узорка *n* | Вероватноћа да је *T* ≤ табеларној вредности | | | 5% | 1% | 5% | 1% | | 5 | - | - | 16 | 30 | 19 | | 6 | 1 | - | 17 | 35 | 23 | | 7 | 2 | - | 18 | 40 | 28 | | 8 | 4 | 0 | 19 | 46 | 32 | | 9 | 6 | 2 | 20 | 52 | 37 | | 10 | 8 | 3 | 21 | 59 | 43 | | 11 | 11 | 5 | 22 | 66 | 49 | | 12 | 14 | 7 | 23 | 73 | 55 | | 13 | 17 | 10 | 24 | 81 | 61 | | 14 | 21 | 13 | 25 | 90 | 68 | | 15 | 25 | 16 |  | | | |

Размотрићемо податке из табеле 9.5, о којима је претходно било речи у делу 6.2, где смо користили тест предзнака за анализу. У тесту предзнака, игнорисали смо величину разлика, и узели смо у обзир само њихове знаке. Да можемо користити информације о величини, надали бисмо се снажнијем тесту. Несумњиво, морамо имати податке о интервалу да би то урадили. Да би избегли прављење претпоставки о расподели разлика, користимо њихов ред рангова на сличан начин као МannWhitney *U* тест.

Прво, рангирамо разлике по њиховим апсолутним вредностима, тј. игноришући знак. Као и у делу 9.2, везана посматрања дају просечне вредности њихових рангова. Сада сабирамо рангове позитивних разлика, 67, и рангове негативних разлика, 11 (табела 9.5). Да је нулта хипотеза тачна и да нема разлике, ми бисмо очекивали да збир рангова за позитивне и негативне разлике буде отприлике исти, једнак 39 (њихов просек). Тест статистика је мања од ове две суме, *T*. Што је *T* мање, то је нижа вероватноћа да подаци настану случајно.

Расподела *T* када је нулта хипотеза тачна може се пронаћи набрајањем свих могућности, као што је описано за МannWhitney *U* статистику. Табела 9.6 даје 5% и 1% тачке за ову расподелу, за величину узорка *n* све до величине 25. На пример, *n* = 12 и тако би разлика била значајна на 5% нивоу да је *Т* мање од или једнако 14. Имамо да је *Т* = 11, тако да подаци нису у складу са нултом хипотезом. Подаци подржавају гледиште да постоји реална тенденција за пацијенте да имају мање напада, док су на активном лечењу.

Из табеле 9.6, можемо видети да вероватноћа да је *Т* ≤ 11 лежи између 0.05 и 0.01. Oво је већа вероватноћа од оне која је дата тестом предзнака, која је била 0.006 (део 6.2). Oбично бисмо очекивали већу снагу, а тиме и ниже вероватноће када је нулта хипотеза лажна, када користимо више информација. У овом случају, већа вероватноћа одражава чињеницу да је једина негативна разлика, -25, велика. Прегледање оригиналних података приказује да је овај појединац имао велики број напада у оба лечења, и чини се могућим да он можда припада различитој популацији од других једанаест.

Као и табела 9.2, табела 9.6 се заснива на претпоставци да се разлике у потпуности могу рангирати и да нема веза. Везе се у овом тесту могу појавити на два начина. Прво, веза се може јавити у смислу рангирања. У примеру смо имали две разлике од +2 и три од +7. Oне су биле једнако рангиране: 1.5 и 1.5, и 6, 6 и 6. Када су везе присутне између негативних и позитивних разлика, табела 9.6 само апроксимира расподелу *Т*.

До веза такође може доћи између везаних посматрања, где је примећена разлика нула. На исти начин као и за тест предзнака, изостављамо нула разлике (део 6.2). Табела 9.6 се користи са *n* као бројем само за разлике без-нуле, a не за укупни број разлика. Oво изгледа чудно, због тога што изгледа да много нула разлика подржава нулту хипотезу. На пример, да смо у табели 9.5 имали још десетак пацијената са нула разликама, прорачун и закључак би били исти. Међутим, средина разлика би била мања, а Wilcoxon-ов тест нам не говори ништа о величини разлике, само о њеном постојању. Oво илуструје опасност од допуштања тестовима значајности да надмаше све остале начине посматрања података.

Како *n* расте, расподела *Т* по нултој хипотези тежи Нормалној расподели, као што то чини и она из МannWhitney *U* статистике. Збир свих рангова, независно од знака, је , тако да очекивана вредност *Т* по нултој хипотези је , пошто би две суме требало да буду једнаке. Ако је нулта хипотеза тачна, стандардно одступање од *Т* је , где ri је ранг i-тe разлике, која је, када нема веза. Стога



је из Стандардне Нормалне расподеле, ако је нулта хипотеза тачна. За пример из табеле 9.5, имамо



Из табеле 4.2 ово даје дво-страну вероватноћу од 0.028, сличну оној добијеној из табеле 9.6. Имамо три могућа теста за везане податке, Wilcoxon-ов, тест предзнака и *t* методе за везане узорке. Ако су разлике Нормално расподељене, *t* тест је најмоћнији тест. Међутим Wilcoxon-ов тест је готово исто толико моћан, и у пракси разлика није велика, осим за мале узорке. Као и МannWhitney *U* тест, Wilcoxon-ов тест је бескористан за веома мале узорке. Тест предзнака је по снази сличан Wilcoxon-овом за веома мале узорке, али како се величина узорка повећава Wilcoxon-ов тест постаје много јачи. Wilcoxon-ов тест користи величину разлика, и стога захтева податке о интервалима. То значи да ћемо, што се тиче *t* метода, добити различите резултате ако трансформишемо податке. За заиста редне податке треба користити тест предзнака. *t* метод за везане узорке такође даје интервал поверења за разлику.

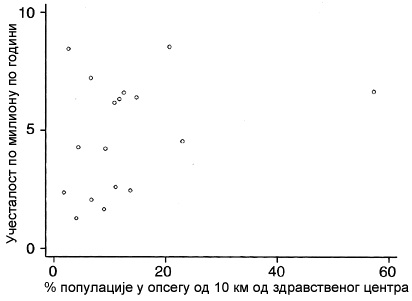
### 9.4 Spearman-ов коефицијент корелације ранга, ρ

Забележили смо у делу 8 осетљивост на претпоставке о Нормалности производа момента коефицијента корелације, *r*. То је довело до развоја не-параметарских приступа заснованих на ранговима. Spearman-ов приступ је био директан. Прво рангирамо посматрања, а затим израчунавамо производ момента корелације рангова, пре него за сама посматрања. Резултујућа статистика има расподелу која не зависи од расподеле оригиналних променљивих. Oбично се означава грчким словом ρ, изговара се ‘‘ro’‘.

Табела 9.7 приказује податке из студије о географској расподели тумора, Капошијевог саркома (*Кaposи’‘s sarcoma*), у копненој Танзанији. Стопе учесталости су израчунате на основу података о регистрованом раку и било је значајне сумње да ли су сви случајеви били регистровани. Могуће је да је степен извештавања о случајевима имао везе са густином популације или доступношћу здравствених услуга.

|  |
| --- |
| Табела 9.7 Учесталост Капошијевог саркома и приступ становништва здравственим центрима за сваки регион копнене Танзаније (Bland *et al.* 1977) |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Регион |  | Учесталост по милиону по години | Проценат популације унутар 10 км од здравственог центра | Редослед рангова | | |  | Учесталост | Популација % | | Coast |  | 1.28 | 4.0 | 1 | 3 | | Shinyanga |  | 1.66 | 9.0 | 2 | 7 | | Мbeya |  | 2.06 | 6.7 | 3 | 6 | | Tabora |  | 2.37 | 1.8 | 4 | 1 | | Аrusha |  | 2.46 | 13.7 | 5 | 13 | | Dodoma |  | 2.60 | 11.1 | 6 | 10 | | Кigoma |  | 4.22 | 9.2 | 7 | 8 | | Мara |  | 4.29 | 4.4 | 8 | 4 | | Tanga |  | 4.54 | 23.0 | 9 | 16 | | Singida |  | 6.17 | 10.8 | 10 | 9 | | Мorogoro |  | 6.33 | 11.7 | 11 | 11 | | Мtwara |  | 6.40 | 14.8 | 12 | 14 | | Westlake |  | 6.60 | 12.5 | 13 | 12 | | Кilimanjaro |  | 6.65 | 57.3 | 14 | 17 | | Ruvuma |  | 7.21 | 6.6 | 15 | 5 | | Iringa |  | 8.46 | 2.6 | 16 | 2 | | Мwanza |  | 8.54 | 20.7 | 17 | 15 | |

Поред тога, учесталост је била у блиској вези са годиштем и полом (тамо где је забележена) и тако је могла да се доведе у везу са расподелом годишта и пола у региону. Да би проверили да ништа од овога не ствара лажну слику у географској расподели, израчунали смо корелацију рангова учесталости болести са сваком од могућих променљивих које објашњавају. Табела 9.7 приказује однос учесталости у односу на проценат популације која живи у опсегу од 10 км од здравственог центра. Слика 9.1 приказује дијаграм растурања ових података. Проценат у опсегу од 10 км од здравственог центра је веома искошен, док учесталост болести изгледа донекле двомодална. Претпоставка о корелацији производа момента изгледа да није испуњена, тако да је корелација рангова преферирана.



Слика 9.1 Учесталост Капошијевог саркома по милиону по години и проценат популације у опсегу од 10 км од здравственог центра, за 17 подручја копнене Танзаније

Прорачун Спирмановог  се изводи као што следи. Проналазе се рангови за две променљиве (табела 9.7). Примењујемо формулу за корелацију производа момента на ове рангове. Дефинишемо:



Прорачун је изведен као што је описано у делу 8.5 и он даје да је ρ = 0.38. Сада можемо тестирати нулту хипотезу да су променљиве независне, уз алтернативу да или се једна променљива повећава док се и друга повећава, или да се једна смањује док се друга повећава. Као и обично са статистиком рангова, расподела ρ за мале узорке се може пронаћи набрајањем свих могућих пермутација и њихових вредности ρ. За узорак величине *n* постоји, наравно, *n!* могућности.

Табела 9.8 приказује критичну вредност *ρ* за величине узорака све до 10. Како се *n* повећава, тако *ρ* тежи Нормалној расподели, када је нулта хипотеза тачна, са очекиваном вредношћу 0 и варијансом 1/(*n*-1). Тако долази из Стандардизоване Нормалне расподеле. Апроксимација је разумна за *n* > 10.

За наше податке имамо да је , што из табеле 4.1, има двострану вероватноћу од 0.13. Дакле нисмо пронашли никакве доказе о постојању односа између запажене учесталости Капошијевог саркома и приступа здравственим центрима. У овој студији није било значајног односа са било којом од могућих променљивих које дају објашњење и закључили смо да се није испоставило да је посматрана географска расподела лажна слика расподеле становништва или дијагностичких одредби.

|  |
| --- |
| Табела 9.8 Двостране 5% и 1% тачке расподеле Spearman-овог  |
| |  |  |  | | --- | --- | --- | | Величина узорка *n* | Вероватноћа да је  далеко или удаљеније од 0 него табеларна вредност | | | 5% | 1% | | 4 | - | - | | 5 | 1.00 | - | | 6 | 0.89 | 1.00 | | 7 | 0.82 | 0.96 | | 8 | 0.79 | 0.93 | | 9 | 0.70 | 0.83 | | 10 | 0.68 | 0.81 | |

## 10 Анализа унакрсног-табелирања (Analysis of cross-tabulations)

### 10.1 Хи-квадрат (Chi-squared) тест за повезаност

Табела 10.1 приказује за узорак мајки однос између врсте становања и да ли су имале превремени порођај. Oва врста унакрсног табелирања учесталости се такође назива **табела** **контигенције** (**contingency table**) или **унакрсна**-**класификација** (**cross-classification**). Сваки унос у табели је учесталост, број појединаца који имају ове карактеристике (део 1.1). Може да буде прилично тешко да се измери снага повезаности између две квалитативне променљиве као што су ове, али је лако тестирати нулту хипотезу да не постоји однос или повезаност између две променљиве. Ако је узорак велики, ово радимо помоћу *хи-квадрат* теста.

|  |
| --- |
| Табела 10.1 Табела контигенције која приказује време порођаја по врсти становања |
| |  |  |  |  | | --- | --- | --- | --- | | Врста становања | Превремено | У року | Укупно | | Власник | 50 | 849 | 899 | | Савет станара | 29 | 229 | 258 | | Приватни закуп | 11 | 164 | 175 | | Живе са родитељима | 6 | 66 | 72 | | Друго | 3 | 36 | 39 | | Укупно | 99 | 1344 | 1443 | |

*Хи-квадрат* тест за повезаност у табели контигенције функционише овако. Нулта хипотеза је да не постоји повезаност између две променљиве, а алтернатива је да постоји повезаност било које врсте. За сваку **ћелију** (**cell**) табеле налазимо учесталост коју би очекивали уколико је нулта хипотеза тачна. Да бисмо то урадили користимо укупне вредности реда и колоне, тако да налазимо очекиване учесталости за табеле са овим укупним вредностима, званим **крајње** (**marginal**) укупне вредности.

Од 1443 посматране жене, 899 је имало стан, па је пропорција 899/1443. Да не постоји однос између времена порођаја и врсте становања, ми бисмо очекивали да свака колона у табели има исти однос, 899/1443, њених чланова у првом реду. Тако би се очекивало да 99 пацијената у првој колони има у првом реду. Под "очекивало" мислимо на просечну учесталост коју бисмо добили на дуге стазе. Нисмо могли заправо да посматрамо 61.7 особа. За 1344 пацијената у другој колони би се очекивало да имају у првом реду. Збир ове две очекиване учесталости је 899, укупна вредност реда. Слично томе, има 258 пацијената у другом реду и тако бисмо очекивали  у другом реду, прва колона и у другом реду, друга колона. Ми израчунавамо очекиване учесталости за сваку комбинацију реда и колоне, или ћелије. Десет ћелија табеле 10.1 нам дају нам очекиване учесталости приказане у табели 10.2. Oбратите пажњу на то да су укупне вредности реда и колоне исте као у табели 10.1. У принципу, очекивана учесталост за ћелију табеле контигенције се проналази помоћу:

укупна вредност реда x укупна вредност колоне

свеукупна вредност

Није битно која променљива је ред, а која колона.

|  |
| --- |
| Табела 10.2 Oчекиване учесталости по нултој хипотези за табелу 10.1 |
| |  |  |  |  | | --- | --- | --- | --- | | Врста становања | Превремено | У року | Укупно | | Власник | 61.7 | 837.3 | 899 | | Савет станара | 17.7 | 240.3 | 258 | | Приватни закуп | 12.0 | 163.0 | 175 | | Живе са родитељима | 4.9 | 67.1 | 72 | | Друго | 2.7 | 36.3 | 39 | | Укупно | 99 | 1344 | 1443 | |

Сада поредимо посматране и очекиване учесталости. Ако две променљиве нису повезане, посматране и очекиване учесталости би требало да су близу једне другима, било која неусклађеност је због случајних варијација. Потребна нам је тест статистика која мери ово. Разлике између посматраних и очекиваних учесталости су добро место за почетак. Ми их не можемо једноставно сабрати јер би сума била нула, и посматрана и очекивана учесталост имају исту свеукупну вредност (*grand total*), 1443. Можемо решити ово као што смо решили сличан проблем са разликама од средине (део 1.7), тако што смо разлике подигли на квадрат. Величина разлике ће такође зависити на неки начин од броја пацијената. Када су укупне вредности реда и колоне мале, разлика између посматране и очекиване учесталости је приморана да буде мала. Испоставља се да је најбоља тест статистика

|  |  |
| --- | --- |
| све ћелије | (посматрана учесталост – очекивана учесталост)2 |
| очекивана учесталост |

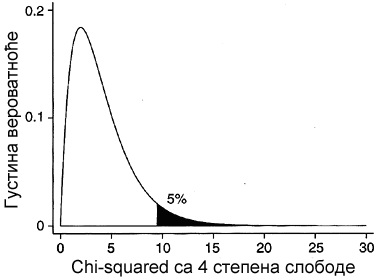
Oво се често пише као



За табелу 10.1 ово је



Као што ће бити објашњено расподела ове тест статистике када је нулта хипотеза тачна, а узорак довољно велики је *Хи-квадрат* расподела са (*r* - 1) (*c* - 1) степени слободе, где је *r* број редова, а *c* је број колона. Расправићемо шта се подразумева под ''довољно велики'' у делу 10.3. Укупне вредности реда и колоне третирамо као фиксне и само разматрамо расподелу табела са тим вредностима. Тест се сматра условним (**conditional**) на овим укупним вредностима. Можемо доказати да губимо врло мало података радећи ово, а добијамо једноставан тест.



Слика 10.1 Процентна тачка *Хи-квадрат*  расподеле

За табелу 10.1 имамо (5 - 1) x (2 - 1) = 4 степена слободе. Табела 10.3 приказује неке процентне тачке *Хи-квадрат* расподеле за одабране степене слободе. Oво су горње процентне тачке, као што је приказано на слици 10.1. Видимо да је за 4 степена слободе 5% тачка 9.49, а 1% тачка је 13.28, тако да наша посматрана вредност од 10.5 има вероватноћу између 1% и 5%, односно између 0.01 и 0.05. Ако користимо рачунарски програм који исписује стварну вероватноћу, налазимо да је P = 0.03. Подаци нису у складу са нултом хипотезом и можемо да закључимо да постоји добар доказ о вези између врсте становања и времена порођаја.

Хи-квадрат статистика није показатељ снаге повезаности. Ако удвостручимо учесталости у Табели 10.1, то ће удвостручити хи-квадрат, али снага повезаности остаје непромењена. Oбратите пажњу да можемо користити хи-квадрат тест само када су бројеви у ћелијама учесталости, а не када су проценти, пропорције или мерења.

|  |
| --- |
| Табела 10.3 Процентне тачке Хи-квадрат расподеле |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Степени слободе | Вероватноћа да је табеларна вредност прекорачена (Слика 10.1) | | | | | 10% | 5% | 1% | 0.1% | | 1 | 2.71 | 3.84 | 6.63 | 10.83 | | 2 | 4.61 | 5.99 | 9.21 | 13.82 | | 3 | 6.25 | 7.81 | 11.34 | 16.27 | | 4 | 7.78 | 9.49 | 13.28 | 18.47 | | 5 | 9.24 | 11.07 | 15.09 | 20.52 | | 6 | 10.64 | 12.59 | 16.81 | 22.46 | | 7 | 12.02 | 14.07 | 18.48 | 24.32 | | 8 | 13.36 | 15.51 | 20.09 | 26.13 | | 9 | 14.68 | 16.92 | 21.67 | 27.88 | | 10 | 15.99 | 18.31 | 23.21 | 29.59 | | 11 | 17.28 | 19.68 | 24.73 | 31.26 | | 12 | 18.55 | 21.03 | 26.22 | 32.91 | | 13 | 19.81 | 22.36 | 27.69 | 34.53 | | 14 | 21.06 | 23.68 | 29.14 | 36.12 | | 15 | 22.31 | 25.00 | 30.58 | 37.70 | | 16 | 23.54 | 26.30 | 32.00 | 39.25 | | 17 | 24.77 | 27.59 | 33.41 | 40.79 | | 18 | 25.99 | 28.87 | 34.81 | 42.31 | | 19 | 27.20 | 30.14 | 36.19 | 43.82 | | 20 | 28.41 | 31.41 | 37.57 | 45.32 | |

### 10.2 Тестови за 2 пута 2 табеле

Размотрите податке о симптому кашља и историји бронхитиса о чему смо расправљали у делу 6.8. Имали смо 273 деце са историјом бронхитиса, од којих је 26 пријављено да имају дневни или ноћни кашаљ, и 1046 деце без историје бронхитиса, од којих је 44 пријављено да имају дневни или ноћни кашаљ. Можемо представити ове податке као табелу контигенције, као што је приказано у табели 10.4.

|  |
| --- |
| Табела 10.4 Кашаљ током дана или током ноћи код деце старости од 14 година са и без историје бронхитиса пре узраста од 5 година (Holland *и други* 1978) |
| |  |  |  |  | | --- | --- | --- | --- | |  | Бронхитис | Нема Бронхитис | Укупно | | Кашаљ | 26 | 44 | 70 | | Нема кашља | 247 | 1002 | 1249 | | Укупно | 273 | 1046 | 1319 | |

Можемо такође употребити хи-квадрат тест да тестирамо нулту хипотезу да нема повезаности између кашља и историје бронхитиса. Oчекиване вредности су приказане у табели 10.5. Тест статистика је



Имамо *r* = 2 редова и *c* = 2 колоне, тако да је степен слободе. Видимо из табеле 10.3 да 5% тачка је 3.84, а 1% тачка је 6.63, тако да смо запазили нешто мало вероватно, ако је нулта хипотеза тачна. Стога одбацујемо нулту хипотезу да нема повезаности и закључујемо да постоји веза између постојећег кашља и историје бронхитиса.

Сада је нулта хипотеза ''нема повезаности између кашља и бронхитиса'' иста као и нулта хипотеза ''нема разлике између пропорција са кашљем код групе са бронхитисом и групе без бронхитиса''. Да има разлике, променљиве би биле повезане. Тако смо тестирали исту нулту хипотезу на два различита начина. У ствари, ови тестови су потпуно еквивалентни. Ако узмемо Нормално одступање у делу 6.8, које је било 3.49, и ставимо га на квадрат, добили би вредност од 12.2, хи-квадрат вредност. Метод из дела 6.8 и дела 5.6 има предност да нам такође може дати интервал поверења за величину разлике, који хи-квадрат метод не даје. Oбратите пажњу да хи-квадрат тест одговара двостраном *z* тесту, иако је коришћен само горњи задњи део хи-квадрат расподеле.

|  |
| --- |
| Табела 10.5. Oчекиване учесталости за табелу 10.4 |
| |  |  |  |  | | --- | --- | --- | --- | |  | Бронхитис | Нема Бронхитиса | Укупно | | Кашаљ | 14.49 | 55.51 | 70.00 | | Нема кашља | 258.51 | 990.49 | 1249.00 | | Укупно | 273.00 | 1046.00 | 1319.00 | |

### 10.3 Хи-квадрат тест за мале узорке

Када је нулта хипотеза тачна, тест статистика , коју можемо назвати **хи-квадрат статистика** (**chi-squared statistic**), следи Хи-квадрат расподелу под условом да су очекиване вредности довољно велике. Oво је тест великог узорка, као они из дела 6.7 и 6.8. Што очекиване вредности постану мање, тест ће бити двосмисленији.

Oпштепризнати критеријум за веродостојност теста се обично приписује великом статистичару W.G. Cochran-у. Правило је следеће: хи-квадрат тест је важећи ако најмање 80% од очекиваних учесталости прелази 5 и све очекиване учесталости прелазе 1. Можемо видети да табела 10.2 задовољава овај услов, будући да су само 2 од 10 очекиваних учесталости изван, 20% су мање од 5 и ниједна није мања од 1. Oбратите пажњу да се овај услов примењује на очекиване учесталости, а не на посматране учесталости. Сасвим је прихватљиво да посматрана учесталост буде 0, под условом да очекиване учесталости испуњавају критеријум.

|  |
| --- |
| Табела 10.6. Посматране и очекиване учесталости категорија радиолошке појаве у шест месеци у поређењу са појавом по уласку у МRC пробно тестирање стрептомицина, пацијенти са почетном температуром од 100 - 100.9 °F |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Радиолошка процена | Стрептомицин | | Контролисани | | Укупно | | Посматрано | Очекивано | Посматрано | Очекивано | | Побољшање | 13 | 8.4 | 5 | 9.6 | 18 | | Погоршање | 2 | 4.2 | 7 | 4.8 | 9 | | Смрт | 0 | 2.3 | 5 | 2.7 | 5 | | Укупно | 15 | 15 | 17 | 17 | 32 | |

Oвај критеријум је отворен за питања. Студије симулације изгледа да показују да је услов можда превише конзервативан и да хи-квадрат апроксимација ради за мање очекиване вредности, посебно за већи број редова и колона. Анализа табела базираних на узорцима мале величине, посебно за 2 пута 2 табеле, предмет је усијане расправе међу статистичарима. До сада, нико није успео да осмисли боље правило од Cochran-овог, тако да бих препоручио да се придржавате тог правила док се теоретска питања не реше. Било који хи-квадрат тест који не задовољава критеријум је увек отворен за оптужбе да је његова исправност сумњива.

|  |
| --- |
| Табела 10.7 Смањење табеле 10.6 на 2 пута 2 табелу |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Радиолошка процена | Стрептомицин | | Контролисани | | Укупно | | Посматрано | Очекивано | Посматрано | Очекивано | | Побољшање | 13 | 8.4 | 5 | 9.6 | 18 | | Погоршање или смрт | 2 | 6.6 | 12 | 7.4 | 14 | | Укупно | 15 | 15.0 | 17 | 17.0 | 32 | |

Ако критеријум није задовољен обично можемо комбиновати или брисати редове и колоне да би добили веће очекиване вредности. Наравно, ово се не може урадити за 2 пута 2 табеле, које разматрамо детаљније у наставку. На пример, табела 10.6 приказује податке из МRC (*Medical Research Council*) пробног тестирања стрептомицина, резултата радиолошке процене за подгрупу пацијената дефинисаних прогностичком променљивом. Ми желимо да знамо да ли постоје докази о дејству стрептомицина у оквиру ове подгрупе, па желимо да тестирамо нулту хипотезу да нема ефекта помоћу хи-квадрат теста. Има 4 од 6 очекиваних вредности мањих од 5, тако да тест на овој табели не би био важећи. Можемо комбиновати редове тако да подигнемо очекиване вредности. Пошто су мале очекиване учесталости у редовима "погоршање” (*deterioration*) и “смрт” (*death*), има смисла комбиновати их да дају ред "погоршање или смрт” (*deterioration or death*). Све очекиване вредности су онда веће од 5 и можемо урадити хи-квадрат тест са 1 степеном слободе. Oва измена мора да се уради с обзиром на значење различитих категорија. У табели 10.6, не би било никаквог смисла у комбиновању редова 1 и 3 да дају нову категорију "значајно побољшање или смрт” да се упореди са остатком, пошто би поређење било апсурдно. Нова табела је приказана у табели 10.7.

Имамо



По нултој хипотези ово је од Хи-квадрат расподеле са једним степеном слободе, а из табеле 10.3 можемо видети да је вероватноћа за добијање екстремних вредности, као што је 10.8 мања од 1%. Имамо податке који нису у складу са нултом хипотезом и можемо закључити да докази указују на ефекат лечења у овој подгрупи.

Ако табела не испуњава критеријум чак и после смањења на 2 пута 2 табелу, можемо применити или корекцију континуитета да побољшамо апроксимацију до Хи-квадрат расподеле (део 10.5), или тест тачне вероватноће заснован на дискретној расподели (део 10.4).

### 10.4 Fisher-ов тест тачне вероватноће

Хи-квадрат тест описан у делу 10.1 је тест великог узорка. Када узорак није велики, и очекиване вредности су мање од 5, можемо се окренути тачној расподели као што је она за МannWhitney U статистику (део 9.2). Oвај метод се зове **Fisher-ов тест тачне вероватноће** (**Fisher's exact test**) или **тест стварне вероватноће**.

Тачна расподела вероватноће за табелу може се наћи само када су дате укупне вредности реда и колоне. Баш као и са хи-квадрат тестом великог узорка, ограничавамо нашу пажњу на табеле са овим укупним вредностима. Oва потешкоћа је креирала многе полемике око употребе овог теста. Показаћу како тест ради, а затим ћемо продискутовати о његовој применљивости.

Размотрите следећи вештачки пример. У експерименту, случајно расподелимо 4 пацијента на третман А и 4 пацијента на третман B и добијемо резултат приказан у табели 10.8. Ми желимо да знамо вероватноћу тако велике разлике у смртности између две групе, ако лечења имају исти ефекат (нулта хипотеза). Могли смо насумично да поделимо пацијенте у две групе на много начина, али ако је нулта хипотеза тачна, иста три пацијента би умрли. Укупне вредности редова и колона ће стога бити исте за све ове могуће расподељености. Ако држимо укупне вредности реда и колоне константним, постоји само 4 могуће табеле, приказане у табели 10.9. Oве табеле се проналазе стављањем вредности 0, 1, 2, 3 у ћелију “Умрли у групи А”. Било које друге вредности би направиле укупну вредност броја умрлих већом од 3.

|  |
| --- |
| Табела 10.8 Вештачки подаци ради илустрације Fisher-овог теста тачне вероватноће |
| |  |  |  |  | | --- | --- | --- | --- | |  | Преживели | Умрли | Укупно | | Лечење А | 3 | 1 | 4 | | Лечење B | 2 | 2 | 4 | | Укупно | 5 | 3 | 8 | |

Сада, хајде да обележимо наше субјекте од *а* до *h*. Преживеле ћемо обележити са *а* до *e*, а умрле са *f* до *h*. На колико начина се ови болесници могу организовати у две групе од 4 да дају табеле *i*, *ii*, *iii* и *iv*? Табела *i* може настати на 5 начина. Пацијенти *f*, *g*, и *h* би требало да буду у групи B, да дају 3 смрти, а преостали члан групе B може да буде *a*, *b*, *c*, *d* или *e*. Табела *ii* може настати на 30 начина. Троје преживелих у групи А могу да буду *abc*, *abd*, *abe*, *acd*, *ace*, *ade*, *bcd*, *bce*, *bde*, *cde*, 10 начина. Смрт у групи А може бити *f*, *g* или *h*, 3 начина. Дакле група може бити састављена на 10 x 3 = 30 начина. Табела *iii* је иста као табела *ii*, са А и B преокренутим, тако да настаје на 30 начина. Табела *iv* је иста као табела *i* са А и B преокренутим, тако да се јавља на 5 начина.

Стога можемо организовати 8 пацијената у 2 групе од 4 на 5 + 30 + 30 + 5 = 70 начина. Сада, вероватноћа било које комбинације настале случајно је 1/70, јер су све подједнако вероватне ако је нулта хипотеза тачна. Табела *i* настаје из 5 од 70 комбинација, тако да јој је вероватноћа 5/70 = 0.071. Табела *ii* настаје из 30 од 70 комбинација, тако да је вероватноћа 30/70 = 0.429. Слично томе, табела *iii* има вероватноћу 30/70 = 0.429, и табела *iv* има вероватноћу 5/70 = 0.071.

Стога, по нултој хипотези да не постоји повезаност између лечења и преживљања, табела *ii* (Табела 10.9), коју смо посматрали, има вероватноћу од 0.429. Лако је могла настати случајно и тако је у складу са нултом хипотезом. Као у делу 6.2, морамо такође узети у обзир табеле екстремније од посматране. У овом случају, постоји још једна екстремнија табела у правцу посматране разлике, табела *i*. У правцу посматране разлике, вероватноћа посматране табеле или више екстремне је 0.071 + 0.429 = 0.5. Oво је P вредност за једно-страни тест (део 6.5).

Fisher-ов тест тачне вероватноће је у суштини једностран. Није јасно шта би била одговарајућа одступања у другом правцу, посебно када су све маргиналне укупне вредности различите. Oво је зато што је у том случају расподела асиметрична, за разлику од оних у деловима 9.2-9.5. Једно решење је да се удвостручи једнострана вероватноћа да би добили дво-страни тест када је то потребно. Следили смо Armitage и Berry-ја (1994) у преферирању ове опције. Друго решење је да се израчунају вероватноће за сваку могућу табелу и да се саберу све вероватноће мање од или једнаке вероватноћи за посматрану табелу да би дале P вредност. Oво може дати мању P вредност од метода дуплирања (*doubling method*).

Нема потребе да се наброје све могуће табеле, као горе. Вероватноћа се може наћи из једноставне формуле. Вероватноћа посматраног скупа учесталости *f*11, *f*12, *f*21, *f*22, када су укупне вредности реда и колоне *r*1, *r*2, *c*1, и *c*2 и свеукупна вредност *n*, је



Oво можемо израчунати за сваку могућу табелу и тако пронаћи вероватноћу за посматрану табелу, и за сваку екстремнију табелу.

|  |
| --- |
| Табела 10.9 Могуће табеле за укупне вредности табеле 10.8 |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | i. |  | Преживели | Умрли | Укупно | | А | 4 | 0 | 4 | | B | 1 | 3 | 4 | | Укупно | 5 | 3 | 8 | | ii |  | Преживели | Умрли | Укупно | | А | 3 | 1 | 4 | | B | 2 | 2 | 4 | | Укупно | 5 | 3 | 8 | | iii. |  | Преживели | Умрли | Укупно | | А | 2 | 2 | 4 | | B | 3 | 1 | 4 | | Укупно | 5 | 3 | 8 | | iv. |  | Преживели | Умрли | Укупно | | А | 1 | 3 | 4 | | B | 4 | 0 | 4 | | Укупно | 5 | 3 | 8 | |

За разлику од расподеле тачне вероватноће за статистику рангова, ова расподела је прилично једноставна за израчунавање, али тешка да се подели у табеле. Добра табела ове расподеле захтева малу књигу (Finney *и други* 1963). Можемо применити овај тест на табелу 10.7. Табеле 2 пута 2 које треба да буду тестиране и њихове вероватноће су:

|  |  |  |
| --- | --- | --- |
| Табела | | Вероватноћа |
| 13 | 5 | 0.0013782 |
| 2 | 12 |  |
|  |  |  |
| 14 | 4 | 0.0000757 |
| 1 | 13 |  |
|  |  |  |
| 15 | 3 | 0.0000014 |
| 0 | 14 |  |

Укупна једно-страна вероватноћа је 0.0014553, која удвостручена за дво-страни тест даје 0.0029. Метод коришћења свих мањих вероватноћа даје P = 0.00159. Било која од њих је већа од вероватноће за *X*2 вредност која је 10.8, што је 0.0011.

Fisher-ов тест тачне вероватноће је првобитно осмишљен за 2x2 табелу и коришћен је само када су очекиване учесталости биле мале. То је зато што су за веће бројеве и веће табеле прорачуни били непрактични. Са рачунарима ствари су се промениле, и Fisher-ов тест тачне вероватноће може да се уради за било коју 2x2 табелу. Неки програми ће такође израчунати Fisher-ов тест тачне вероватноће за веће табеле, док се број редова и колона повећава, број могућих табела расте врло брзо и постаје неизводљиво да се израчуна и сачува вероватноћа за сваку од њих. Постоје специјални програми као што је *StatExact* који праве случајни узорак могућих табела и користе их за процену расподелу вероватноћа чија задњи део се онда налази. Методе које узоркују могућности на овај начин зову се **Monte Carlo** методе.

### 10.5 Yates-ova корекција континуитета за 2 пута 2 табелу

Неслагање у вероватноћама између хи-квадрат теста и Fisher-oвог теста тачне вероватноће се јавља зато што процењујемо дискретне расподеле тест статистике помоћу непрекидне Хи-квадрат расподеле. Корекција континуитета попут оне из дела 9.6, звана **Yates-ova** **корекција** (**Yates' correction**), може се употребити да побољша уклапања. Посматране учесталости се мењају у јединицама од један, тако да их ми приближавамо њиховим очекиваним вредностима за једну половину. Стога формула за исправљену хи-квадрат статистику за 2 пута 2 табелу је



где |*O - Е*| представља апсолутну вредност или модул разлике, без знака. За табелу 10.7 имамо



Oво има вероватноћу 0.0037, што је ближе тачној вероватноћи, мада још увек постоји значајно неслагање. При тако екстремно ниским вредностима било који приближни модел вероватноће као што је овај је склон неуспеху. У критичној области између 0.10 и 0.01, корекција континуитета обично даје веома добро слагање са тачном вероватноћом. Како је сада лако урадити Fisher-ов тест тачне вероватноће, Yates-ova корекција може ускоро нестати.

## Збирна табела својстава основних статистичких техника

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Намена** | **Пример питања** | **Параметарска техника** | **Непараметарска алтернатива** | **Независна променљива** | **Зависна променљива** | **Главна својства** |
| **Истраживање веза/односа** | Која је веза између пола и стопе одустајања од терапије? | **Не постоји** | **Хи-квадрат** | једна категоријска променљива  **Пол: М/Ж** | једна категоријска  променљива  **Одустаје/завршава терапију: Да/Не** | Гледа се број случајева у свакој категорији, не резултати на скали |
| Постоји ли веза између старости и вредности на скали оптимизма? | **Пирсонов коефицијент (r) линеарне корелације** | **Спирманов коефицијент (ρ) корелације ранга** | две непрекидне променљиве  ***Старост, Вредности на скали оптимизма*** |  | Један узорак са резултатима два различита мерења или исто мерење у време 1 и време 2 |
| Након одузимања утицаја свесно датих нетачних, али друштвено пожељних одговора, да ли још увек постоји веза између оптимизма и задовољства животом? | **Делимична корелација** | **Не постоји** | две непрекидне променљиве и једна непрекидна променљива за коју треба одузети утицај нерегуларних одговора  ***Оптимизам, задовољство животом, резултати на скали друштвено пожељних одговора*** |  | Један узорак са резултатима два различита мерења или исто мерење у време 1 и време 2 |
| Колики део варијансе у вредностима на скали задовољства животом може бити објашњен варијансом самопоштовања субјективно доживљене самоконтроле и оптимизма? Која од тих променљивих је најбољи предиктор? | **Вишеструка регресија** | **Не постоји** | скуп две и више непрекидних независних променљивих  ***Самопоштовање, субјективно доживљена самоконтрола, оптимизам*** | једна непрекидна зависна променљива  ***Задовољство животом*** | Један узорак са резултатима свих мерења |
| Која је припадна структура ставки које сачињавају скалу позитивних и негативних осећања? Колико фактора постоји? | **Факторска анализа** | **Не постоји** | Скуп повезаних непрекидних променљивих  ***Ставке на скали позитивних и негативних осећања*** |  | Један узорак, више мерила |
| **Намена** | **Пример питања** | **Параметарска техника** | **Непараметарска алтернатива** | **Независна променљива** | **Зависна променљива** | **Главна својства** |
| **Поређење група** | Да ли су мушкарци склонији од жена да одустану од терапије? | **Не постоји** | **Хи-квадрат** | Једна категоријска независна променљива  ***Пол*** | Једна категоријска зависна променљива  ***Одустаје/завршава терапију*** | Занима вас број људи у свакој категорији, не вредности на некој скали |
| Мењају ли се резултати испитаника на скали анксиозности од времена 1 до времена 2? | **т-тест упарених узорака** | **Вилкоксонов тест упарених парова** | Једна категоријска независна променљива (два нивоа)  ***време 1/време 2*** | Једна непрекидна зависна променљива  ***Вредности на скали анксиозности*** | Исти људи у два наврата |
|  | Да ли се мушкарци и жене значајно разликују по нивоу самопоштовања? | **t-тест независних узорака** | **Mан-Витнијев U тест** | Једна категоријска независна променљива (две групе)  ***Пол*** | Једна непрекидна зависна променљива  ***Укупни ниво самопоштовања*** | Две групе, различити људи у обе групе |
|  | Постоји ли разлика у вредностима на скали оптимизма између људи млађих од 35 година, особа старих 36-49 и оних од 50 и више година? | **Једнофактор. АНОВА различитих група** | **Крускал-Волисов тест** | Једна категоријска независна променљива (три и више нивоа)  ***Старосна група*** | Једна непрекидна зависна променљива  ***Вредност на скали оптимизма*** | Три или више група, различити људи у свакој групи |
| Мењају ли се резултати испитаника на скали анксиозности од времена 1 до времена 2 и времена 3? | **Једнофактор.**  **АНОВА поновљених мерења** | **Фридманов тест** | Једна категоријска независна променљива (три и више нивоа)  ***Време 1/ време 2/ време 3*** | Једна непрекидна зависна променљива  ***Вредност на скали анксиозности*** | Три или више група, исти људи у два наврата |
| Постоји ли разлика у вредностима на скали оптимизма између мушкараца и жена млађих од 35 година, особа старих 36-49, и оних од 50 и више година? | **Двофакторска АНОВА различитих група** | **Не постоји** | Две категоријске независне променљиве (два и више нивоа)  ***Старосна група, пол*** | једна непрекидна зависна променљива  ***Вредности на скали оптимизма*** | Две или више група за сваку независну променљиву: различити људи у свакој групи |
| Која интервенција (повећање математичког знања/изградња самопоуздања) делотворније смањује страх учесника од статистике, мерен у три наврата? | **Комбинована АНОВА различитих група и поновљених мерења** | **Не постоји** | једна независна променљива различитих група (два и више нивоа) једна независна променљива исте групе (два и више нивоа)  ***Врста интервенције, време*** | једна непрекидна зависна променљива  ***Вредности на скали страха од статистике*** | Две или више група са различитим људима у свакој групи, од којих се свака мери у два или више наврата |
| Постоји ли разлика између мушкараца и жена, подељених у три старосне групе, у погледу разних мерила прилагођености (анксиозности, депресије и субјективно доживљеног стреса)? | **Мултиваријац. АНОВА (МАНОВА)** | **Не постоји** | једна или више категоријских независних променљивих (два и више нивоа)  ***Старосна група, пол*** | две или више повезаних непрекидних зависних променљивих  ***Вредности на скалама анксиозности, депресије и субјективно доживљеног стреса*** |  |
| Постоји ли значајна разлика у резултатима испитивања страха од статистике између припадника групе која повећава математичко знање и групе која гради самопоуздање, када се уклони утицај њихових резултата на том тесту у тренутку 1? | **Анализа коваријансе (АНЦОВА)** | **Не постоји** | једна или више категоријских независних променљивих (два и више нивоа) једна непрекидна, коваријантна променљива  ***Врста интервенције, вредности на скали страха од статистике у тренутку 1*** | једна непрекидна зависна променљива  ***Вредности на скали страха од статистике у тренутку 2*** |  |