**Глава 12**

**мЕТОДИ ЗА ИЗУЧАВАНЕ НА ПРИЧИННИ ЗАВИСИМОСТИ**

*Г. Грънчарова, П. Христова*

***В тази глава:***

*12.1. Основни понятия на корелационния анализ*

*12.1.1. Функционална и корелационна зависимост*

*12.1.2. Видове корелационни зависимости*

*12.1.3. Коефициент на корелация – същност и оценка*

*12.1.4. Диаграма на разсейване*

*12.2. Методи за изчисляване на коефициенти на корелация*

*12.2.1. Коефициент на корелация при качествени променливи величини*

*12.2.2. Рангов коефициент на корелация на Спирман*

*12.2.3. Коефициент на корелация на Пирсон при количествени променливи величини*

*12.2.4. Коефициент на детерминация*

*12.3. Регресионен анализ*

*12.4. Въпроси за самоподготовка*

**12.1. КОРЕЛАЦИОНЕН АНАЛИЗ**

**12.1.1. Функционална и корелационна зависимост**

 Ед­на от гла­в­ни­те за­да­чи на ста­ти­с­ти­че­с­кия ана­лиз в медицината и здра­ве­о­па­з­ва­не­то е раз­к­ри­ва­не и изу­ча­ва­не на вза­и­мо­в­ръ­з­ки­те ме­ж­ду яв­ле­нията и при­чин­на­та обу­с­ло­ве­ност на ед­ни яв­ле­ния и про­це­си от дру­ги. От та­зи гле­д­на то­ч­ка мо­же да се раз­г­ле­ж­да­т ***яв­ле­ния при­чи­ни и яв­ле­ния сле­д­с­т­вия*.**

 ***Яв­ле­нията при­чи­ни*** пре­д­с­та­в­ля­ват ус­ло­вие за въз­ни­к­ва­не­то, съ­ще­с­т­ву­ва­не­то или про­ме­ни­те на дру­ги яв­ле­ния. Променливите величини, които ги характеризират се наричат ***независими променливи*** и се отбелязват с ***x***.

 ***Яв­ле­нията сле­д­с­т­вия*** са ре­зул­тат от въз­дей­с­т­ви­е­то на яв­ле­ни­я­та при­чи­ни. Променливите, които ги характеризират се наричат ***зависими променливи*** и се отбелязват с ***y.***

 На основата на установените при­чин­но-сле­д­с­т­ве­ни­ връ­з­ки ме­ж­ду яв­ле­ни­я­та мо­гат да се раз­ра­бо­т­ва­т и реализират ме­рки за от­с­т­ра­ня­ва­не или намаляване на от­ри­ца­тел­но вли­я­е­щи­те фа­к­то­ри, ка­к­то и ме­рки за из­по­л­з­ва­не на по­ло­жи­тел­но вли­я­е­щи­те вър­ху здра­ве­то фа­к­то­ри. Ка­то пра­ви­ло, при раз­г­ле­ж­да­не на при­чин­ната обу­с­ло­ве­ност на здра­ве­то се сблъ­с­к­ва­ме не с еди­ни­ч­ни, изо­ли­ра­ни фа­к­то­ри, а с мно­же­с­т­во ком­п­ле­к­с­но дей­с­т­ва­щи при­чи­ни, ко­е­то зна­чи­тел­но ус­ло­ж­ня­ва ста­ти­с­ти­че­с­кия ана­лиз и осъ­ще­с­т­вя­ва­не­то на ефе­к­ти­в­ни оз­д­ра­ви­тел­ни ме­ро­п­ри­я­тия.

 Раз­ли­ча­ва­ме две фор­ми на про­я­в­ле­ние на при­чин­но-сле­д­с­т­ве­ни­те връ­з­ки ме­ж­ду яв­ле­ни­я­та: ***фун­к­ци­о­нал­на и ко­ре­ла­ци­он­на за­ви­си­мо­с­ти.***

 ***Фун­к­ци­о­нал­на за­ви­си­мост*** е та­зи, при която на вся­ко зна­че­ние на независимата променлива съ­о­т­ве­т­с­т­ва то­ч­но оп­ре­де­ле­но зна­че­ние на зависимата променлива, описваща резултата. Фун­к­ци­о­нал­на­та за­ви­си­мост е ха­ра­к­тер­на за ма­те­мати­ка­та, фи­зи­ко-ма­те­ма­ти­че­с­ки­те про­це­си и дру­ги то­ч­ни на­у­ки. Тя мо­же да се из­ра­зи с то­ч­на ма­те­ма­ти­че­с­ка фор­му­ла. В медицината и здра­ве­о­па­з­ва­не­то функционални за­ви­си­мости по­ч­ти не се сре­щат.

 ***Ко­ре­ла­ци­он­на за­ви­си­мост*** е та­зи, при ко­я­то оп­ре­де­ле­но из­ме­не­ние в яв­ле­ни­е­то при­чи­на не ви­на­ги во­ди до то­ч­но оп­ре­де­ле­на про­мя­на в резултата, т.е. на оп­ре­де­ле­но зна­че­ние на независимата променлива могат да съ­о­т­ве­т­с­т­ват ня­кол­ко зна­че­ния на зависимата променлива. При то­ва из­ме­не­ни­я­та в зависимата променлива са­мо от­ча­с­ти за­ви­сят от съ­о­т­ве­т­ни­те из­ме­не­ния във проучваните фа­к­то­ри, не са пре­до­пре­де­ле­ни от тях, ка­к­то при фун­к­ци­о­нал­на­та за­ви­си­мост. При да­де­на стойност или категория на независимата променлива са въз­мо­ж­ни и се на­б­лю­да­ват че­с­то раз­ли­ч­ни стойности на резултата. На­пр., те­г­ло­то на чо­ве­ка за­ви­си гла­в­но от не­го­вия ръст. Но ос­вен ръ­с­та, вър­ху те­г­ло­то оказват влияние мно­го дру­ги фа­к­то­ри ка­то пол, въз­раст, хра­не­не, енер­го­ра­з­ход, здра­в­но съ­с­то­я­ние и т. н. За­то­ва при ли­ца с ед­на­къв ръст, и да­же на ед­наква въз­раст, ря­д­ко се сре­щат ли­ца с напълно ед­на­к­во те­г­ло - ко­ле­ба­ни­я­та ва­ри­рат в оп­ре­де­ле­ни гра­ни­ци, т.е. има съ­о­т­но­си­тел­ност.

 Тъй ка­то пра­к­ти­че­с­ки е не­въз­мо­ж­но да се изу­чи вли­я­ни­е­то на вси­ч­ки фа­к­то­ри вър­ху да­де­но яв­ле­ние или про­цес, то при из­мер­ва­не на при­чин­но-сле­д­с­т­ве­на­та връ­з­ка е ну­ж­но да се от­ди­фе­рен­ци­рат ос­но­в­ни­те от вто­ро­с­те­пен­ни­те фа­к­то­ри. Не­об­хо­ди­мо е да се тър­си ло­ги­ч­на­та връ­з­ка ме­ж­ду яв­ле­ни­я­та и да се из­бя­г­ва сме­с­ва­не­то на ре­ал­ни­те ко­ре­ла­ци­он­ни за­ви­си­мо­с­ти с т. н. ***ус­по­ре­д­ност в из­ме­не­ни­я­та*** ме­ж­ду две или по­ве­че яв­ле­ния, ко­я­то мо­же да бъде ре­зул­тат от слу­чай­но съ­в­па­де­ние на ня­кои об­с­то­я­тел­с­т­ва, не свър­за­ни ед­но с дру­го, т. е. ва­ж­но е ***да се от­ли­ча­ва има­ги­нер­на­та (при­ви­д­на­та, лъ­ж­ли­ва­та) от ре­ал­на­та ко­ре­ла­ци­он­на връ­з­ка.***

**12.1.2. Видове корелационни зависимости**

 Ко­ре­ла­ци­он­ни­те за­ви­си­мо­с­ти могат да се класифицират:

***1.*** ***Според формата на проявлението си корелацията бива:***

• ***праволинейна***

 • ***криволинейна****.*

 ***Пра­во­ли­ней­на ко­ре­ла­ция*** - при нея ра­в­но­мер­ни­те из­ме­не­ния на фа­к­то­ри­те се при­д­ру­жа­ват с ра­в­но­мер­ни из­ме­не­ния на сле­д­с­т­ви­е­то (ре­зул­та­та), т.е. на­ли­це е ня­ка­к­ва по­с­то­ян­на про­по­р­ци­о­нал­ност ме­ж­ду аб­со­лю­т­ни­те раз­ме­ри на съ­о­т­ве­т­с­т­ва­щи­те из­ме­не­ния във фа­к­то­ри­те и сле­д­с­т­ви­е­то. Ма­те­ма­ти­че­с­ки тя се представя чрез ура­в­не­ни­ята на пра­вата ли­ния – ***у = а + bx*** или ***y = a - bx.***

 ***Кри­во­ли­ней­на ко­ре­ла­ция*** - при нея ра­в­но­мер­ни­те из­ме­не­ния на на­б­лю­да­ва­ни­те фа­к­то­ри се при­д­ру­жа­ват от не­ра­в­но­мер­ни из­ме­не­ния в раз­ме­ра на сле­д­с­т­ви­е­то. В те­зи слу­чаи въз­дей­с­т­ви­е­то на на­б­лю­да­ва­ни­те фа­к­то­ри не е ед­на­к­во при все­ки те­хен раз­мер и за­то­ва се на­ри­ча сло­ж­но въз­дей­с­т­вие. Ма­те­ма­ти­че­с­ки кри­во­ли­ней­на­та ко­ре­ла­ция се пре­д­с­та­вя чрез ура­в­не­ни­я­та на раз­ли­ч­ни ви­до­ве кри­ви ли­нии - па­ра­бо­ла, хи­пер­бо­ла, ек­с­по­нен­ци­ал­на кри­ва, реципрочна и др.

 ***2. Според начина и посоката на влияние:***

 • ***пряка (положителна, еднопосочна***)

 • ***обратна (отрицателна, разнопосочна***)

 ***Пря­ка ко­ре­ла­ция*** се на­б­лю­да­ва, ко­га­то с уве­ли­ча­ва­не (намаляване) на стойностите на независимата променлива нарастват (намаляват) стой­но­с­ти­те на зависимата променлива. На­пр., с уве­ли­ча­ва­не сро­ка на бре­мен­но­ст­та на­ра­с­т­ват стой­но­с­ти­те на ан­т­ро­по­ме­т­ри­ч­ни­те по­ка­за­те­ли при но­во­ро­де­ни­те.

 ***Об­ра­т­на ко­ре­ла­ция*** - при нея с уве­ли­ча­ва­не (намаляване) стой­но­с­ти­те на независимата променлива на­ма­ля­ват (увеличават се) стой­но­с­ти­те на зависимата променлива. На­пр., с уве­ли­ча­ва­не сте­пен­та на об­х­ва­та на де­ца­та с иму­ни­за­ции на­ма­ля­ва ни­во­то на ос­т­ра­та за­раз­на за­бо­ля­е­мост.

***3. Според начина на изследване на връзките****:*

 •  ***обикновена (проста, единична***)

 • ***частична***

 • ***множествена***

 ***Оби­к­но­ве­на ко­ре­ла­ция*** - при нея се из­мер­ва връ­з­ка­та ме­ж­ду две про­мен­ли­ви, ха­ра­к­те­ри­зи­ра­щи ед­на при­чи­на и ед­но сле­д­с­т­вие, без да се взе­ма пред вид вли­я­ни­е­то на дру­ги фа­к­то­ри и при­чи­ни. На­при­мер за­ви­си­мост ме­ж­ду ха­ра­к­те­ра на жи­ли­щ­ни­те ус­ло­вия и здра­ве­то на чле­но­ве­те на се­мей­с­т­во­то, без да се от­чи­та вли­я­ни­е­то на мно­же­с­т­во дру­ги фа­к­то­ри.

 ***Ча­с­ти­ч­на ко­ре­ла­ция*** - при нея се из­с­ле­д­ва връ­з­ка­та ме­ж­ду две про­мен­ли­ви, ха­ра­к­те­ри­зи­ра­щи ед­на при­чи­на и ед­но сле­д­с­т­вие, при ус­ло­вие, че дру­ги­те не­за­ви­си­ми про­мен­ли­ви запазват кон­с­тан­т­но ни­во.

 ***Мно­же­с­т­ве­на ко­ре­ла­ция*** - из­мер­ва връ­з­ка­та ме­ж­ду ед­на за­ви­си­ма про­мен­ли­ва (сле­д­с­т­вие) и мно­же­с­т­во фа­к­то­ри и при­чи­ни, ко­и­то я обу­с­ла­вят. Та­зи ко­ре­ла­ция сле­до­ва­тел­но из­мер­ва це­ло­ку­п­но­то вли­я­ние на вси­ч­ки фа­к­то­ри, вклю­че­ни в да­де­но про­у­ч­ва­не.

 Мно­же­с­т­ве­на­та и ча­с­ти­ч­на ко­ре­ла­ция са зна­чи­тел­но по-сло­ж­ни в те­х­ни­че­с­ко из­пъл­не­ние, но в за­мя­на на то­ва те да­ват мно­го по-цен­на ин­фор­ма­ция за за­ви­си­мо­с­ти­те ме­ж­ду яв­ле­ни­я­та при­чи­ни и сле­д­с­т­вия.

**12.1.3. Коефициент на корелация – същност и оценка**

 При на­ли­чие на дей­с­т­ви­тел­на връ­з­ка, ус­та­но­ве­на на ос­но­ва­та на кон­к­ре­тен ана­лиз на при­ро­да­та на изу­ча­ва­ни­те яв­ле­ния, ста­ти­с­ти­ка­та позволява тя да бъде измерена и да се ус­та­но­ви сте­пен­та на за­ви­си­мост ме­ж­ду яв­ле­ни­я­та.

 Съ­ще­с­т­ве­но пре­дим­с­т­во на ко­е­фи­ци­ен­та на ко­ре­ла­ция е то­ва, че той по­с­ре­д­с­т­вом ед­но един­с­т­ве­но чи­с­ло да­ва пре­д­с­та­ва за на­п­ра­в­ле­ни­е­то и си­ла­та на връ­з­ка­та ме­ж­ду изу­ча­ва­ни­те яв­ле­ния.

 Ко­е­фи­ци­ен­тът на ко­ре­ла­ция има чи­с­ле­на стой­ност и знак (**+** или **-**) и се оз­на­ча­ва най-че­с­то с ***r.***

***Чи­с­ле­на­та стой­ност*** на ***r*** ха­ра­к­те­ри­зи­ра си­ла­та на ко­ре­ла­ци­он­на­та за­ви­си­мост, а ***зна­кът*** пред числото показва на­п­ра­в­ле­ни­е­то на връ­з­ка­та – при знак (**+**) е налице пра­ва ко­ре­ла­ци­он­на за­ви­си­мост, а при зна­к (**-**) e об­ра­т­на.

 ***Чи­с­ле­на­та стой­ност на ко­е­фи­ци­ен­та*** ***r*** мо­же да ва­ри­ра от 0 до 1.

При ***r=0*** ли­п­с­ва ко­ре­ла­ци­он­на за­ви­си­мост.

При ***r=1*** за­ви­си­мо­ст­та е фун­к­ци­о­нал­на.

Ме­ж­дин­ни­те зна­че­ния на ***r*** го­во­рят за по-сил­на или по-сла­ба връзка. Силата на връзката се оце­ня­ва по 3- или по-често по 5-­с­те­пен­на ска­ла:

|  |  |  |
| --- | --- | --- |
| **Степени на корелация**  | **Стойност на r при 5-степенна скала** | **Стойност на r при 3-степенна скала** |
| слабаумереназначителнаголямаизключително голяма | до 0,3от 0,31 до 0,5от 0,51 до 0,7от 0,71 до 0,9над 0,9 | под 0.3от 0.3І до 0.7над 0.7 |

**12.1.4. Диаграма на разсейване**

Ко­ре­ла­ци­он­ни­ят ана­лиз е при­ло­жим ка­к­то при ка­че­с­т­ве­ни, та­ка и при ко­ли­че­с­т­ве­ни про­мен­ли­ви, а съ­що та­ка не­за­ви­си­мо от ха­ра­к­те­ра на раз­пре­де­ле­ни­е­то на про­мен­ли­ви­те - при ал­тер­на­ти­в­но раз­пре­де­ле­ние, при нор­мал­но, при аси­ме­т­ри­ч­ни раз­пре­де­ле­ния и т. н.

 Съ­ще­с­т­ву­ват мно­же­с­т­во ме­то­ди за из­чи­с­ля­ва­не на ***r***. Из­бо­рът на конкретен ме­тод се оп­ре­де­ля от цел­та, ха­ра­к­те­ра и обе­ма на из­с­ле­д­ва­не­то, на­ли­чи­е­то или от­съ­с­т­ви­е­то на из­чи­с­ли­тел­на те­х­ни­ка и най-вече от начина на представяне на данните на измерителните скали.

 При количествени променливи много полезно средство за избор на подходящ метод е представянето на промените в независимата променлива и съответстващите им изменения в зависимата променлива на т.нар. ***диаграма на разсейване (Scatter plot).*** Тя позволява лесно да се определи визуално посоката на връзката (положителна или отрицателна), вида на зависимостта (праволинейна или криволинейна) и да се придобие първоначална представа за силата на връзката или за отсъствие на зависимост между променливите.

 Практическото построяване на диаграмата на разсейване е много лесно. При малък брой случаи то може да се извърши ръчно. На абсцисата се отразява независимата променлива ***x***, а на ординатата - зависимата променлива ***y***. За всяка двойка значения на ***x*** и ***y*** се отбелязва пресечната точка. Групирането на точките определя вида на корелационната връзка или липсата на такава. При голям брой случаи се използват приложни компютърни програми. За целта в електронните таблици се нанасят двойките значения на независимата и зависима променлива. Примерни диаграми на разсейване за различни видове корелационни зависимости са представени на ***фиг. 12.1.***



***Фиг. 12.1. Диаграма на разсейването при праволинейни и криволинейни връзки***

**12.2. Методи за изчисляване на коефициенти на корелация**

***Ме­то­дите за изчисляване на r*** могат да се групират най-об­що в за­ви­си­мост от ха­ра­к­те­ра на на­б­лю­да­ва­ни­те променливи:

* ***ме­то­ди за из­чи­с­ля­ва­не на r при качествени ал­тер­на­ти­в­ни променливи, т. е. при че­ти­ри­к­ра­т­ни та­б­ли­ци;***
* ***ме­то­ди за из­чи­с­ля­ва­не на r при качествени променливи с по­ве­че от две раз­но­ви­д­но­с­ти, т. е. при мно­го­к­ра­т­ни та­б­ли­ци;***
* ***ме­то­ди за из­чи­с­ля­ва­не на r при ко­ли­че­с­т­ве­ни променливи -*** дан­ните за независимите и зависими променливи са в интервална или пропорционална скала (ко­е­фи­ци­ент на ко­ре­ла­ция на Пир­сън);
* ***ран­го­ва ко­ре­ла­ция*** (ко­е­фи­ци­ент на ко­ре­ла­ция на Спир­ман).

**12.2.1. Коефициент на корелация при качествени променливи величини**

 Най-лесно приложими са **методите за изчисляване на коефициент на корелация при качествени алтернативни променливи**, т.е. когато зависимата и независимата променлива имат само по две разновидности. Посочените по-долу подходи позволяват изчисленията да бъдат извършени лесно с помощта на обикновен калкулатор.

Дан­ни­те се пре­д­с­та­вят във вид на че­ти­ри­к­ра­т­на та­б­ли­ца, в ко­я­то с ***х*** е оз­на­че­на про­мен­ли­ва­та, ха­ра­к­те­ри­зи­ра­ща яв­ле­ни­е­то при­чи­на, а с ***y*** - яв­ле­ни­е­то сле­д­с­т­вие. Вся­ка ед­на от те­зи про­мен­ли­ви има са­мо по две раз­но­ви­д­но­с­ти. Въз­мо­ж­ни са че­ти­ри съ­че­та­ния (комбинации) ме­ж­ду раз­но­ви­д­но­с­ти­те им, за­то­ва и та­б­ли­ца­та се на­ри­ча че­ти­ри­к­ра­т­на или таблица 2 х 2:

|  |  |  |  |
| --- | --- | --- | --- |
| ***х* / *у*** | ***у1*** | ***у2*** | **Общо** |
| ***х1*** | **а** | **b** | **a + b** |
| ***х2*** | **c** | **d** | **c + d** |
| **Общо** | **a + c** | **b + d** | **a+b+c+d** |

 Следователно:

 **а** е съчетание на ***х1 у1***

 **b** e съчетание на ***х1 у2***

 **c** е съчетание на ***х2 у1***

**d** е съчетание на ***х2 у2***

Ко­е­фи­ци­ен­тът на ко­ре­ла­ция при ка­че­с­т­ве­ни ал­тер­на­ти­в­ни при­з­на­ци мо­же да се из­чи­с­ли ***чрез формулата на Чупров*** за коефициента на корелация:

, където:

 **n** - брой на всички случаи в таблицата

***χ2еm*** – изчислената стойност от емпиричните данни (виж раздел 12.5.2)

Когато от четирикратната таблица има изчислена стойност на **χ2**, то с тази формула лесно може да се определи степента на взаимодействието. Този начин обикновено се използва, когато няма сигурни данни за формата на разпределението, т.е. при всички видове разпределения на променливите.

Тъй ка­то в зна­ме­на­те­ля има ква­д­ра­тен ко­рен, то из­в­ле­че­но­то от ко­ре­на чи­с­ло ще има знак плюс и ми­нус. За офор­мя­не на край­ния знак пред ко­е­фи­ци­ен­та r зна­ме­на­те­лят се взе­ма с по­ло­жи­те­лен знак и сле­до­ва­тел­но по­со­ка­та на връ­з­ка­та ще за­ви­си от раз­ли­ка­та в чи­с­ли­те­ля, но често този знак се поставя чрез логиката на изследователският проблем.

 След из­чи­с­ля­ва­не­то на ко­е­фи­ци­ен­та ***r*** се пра­ви из­вод за си­ла­та на връзката (съгласно 3-степенната или 5-степенната скала) и за посоката на зависимостта (съгласно знака на r).

 За изу­ча­ва­не на при­чин­но-сле­д­с­т­ве­на­та връ­з­ка при ка­че­с­т­ве­ни ал­тер­на­ти­в­ни при­з­на­ци мо­же да се из­чи­с­ли и друг ***спе­ци­а­лен кри­те­рий за фа­к­тор­но вли­я­ние, на­ре­чен odds ratio (OR***). Пре­во­дът на то­ва по­ня­тие на бъл­гар­с­ки е до­с­та тру­ден - ***съ­о­т­но­ше­ние на две до­пъл­ва­щи се ве­ро­я­т­но­с­ти (очаквано-закономерно към неочаквано-случайно)***. ***OR*** се из­по­л­з­ва мно­го ши­ро­ко в съ­в­ре­мен­ни­те епи­де­ми­о­ло­ги­ч­ни про­у­ч­ва­ния.

 Пре­дим­с­т­во­то на то­зи по­ка­за­тел е в то­ва, че той, ка­к­то и пре­д­хо­д­ни­ят ко­е­фи­ци­ент на ко­ре­ла­ция, се из­чи­с­ля­ва на ба­за­та на аб­со­лю­т­ни­те чи­с­ла в че­ти­ри­к­ра­т­на­та та­б­ли­ца и да­ва до­б­ра пре­д­с­та­ва за ри­с­ка за на­с­тъ­п­ва­не на ед­но или дру­го не­б­ла­го­п­ри­я­т­но яв­ле­ние (за­бо­ля­ва­не, уми­ра­не и др.) при експонирани (из­ло­же­ни на ри­с­ков фа­к­тор) и при не­е­к­с­по­ни­ра­ни ли­ца.

 Из­по­л­з­вай­ки по­со­че­ни­те символи в кле­т­ки­те на че­ти­ри­к­ра­т­на­та та­б­ли­ца:

 ****

 ***Ако OR e ра­в­но на еди­ни­ца***, то две­те гру­пи (ек­с­по­ни­ра­ни и не­е­к­с­по­ни­ра­ни) имат ед­на­к­ва ве­ро­я­т­ност за на­с­тъ­п­ва­не на не­б­ла­го­п­ри­я­т­но­то яв­ле­ние.

 ***Ако OR е по-го­лям от еди­ни­ца***, то гру­па­та на ек­с­по­ни­ра­ни­те ли­ца има по-ви­сок риск за за­бо­ля­ва­не или дру­го не­б­ла­го­п­ри­я­т­но вли­я­ние.

 ***При OR по-ма­лък от еди­ни­ца*** про­у­ч­ва­ни­ят фа­к­тор има про­те­к­ти­в­но (за­щи­т­но) дей­с­т­вие.

 **Пример:** За да се установи има ли връзка между страничните явления и метода на лечение е проведено проучване върху появяването на странични явления (***у***) при две групи болни (***х***): лекувани само с антибиотици (***х***1) и лекувани с антибиотици и витамини (***х***2). Данните от проучването са представени в четирикратна таблица (***табл. 12.1***):

***Табл. 12.1. Резултати от лечението по два метода***

|  |  |  |  |
| --- | --- | --- | --- |
| **Вид на лечението** | **Със странични явления (*у*1)** | **Без странични явления (*у*2)** | **Общо** |
| **С антибиотици и витамини - (*х*1)** | **9** | **57** | **66** |
| **Само с антибиотици - (*х*2)** | **16** | **29** | **45** |
| **Общо** | **25** | **86** | **111** |



 Тъй като стойността на OR е под 1, това показва, че лечението с антибиотици и витамини протектира появяването на странични ефекти.

Стойностите на **С** и **r** (***φ***) показват, че зависимостта е значима обратна и слаба. Знакът на коефициентите трябва да се постави и на логическа основа, а именно, че е отрицателен - колкото повече пациенти се лекуват с антибиотици и витамини, толкова по-малко ще има странични явления. При равни условия коефициентът на Пирсон ***r*** *(****φ****)* показва и посоката на връзката.

 Внимателно трябва да се тълкува и стойността на **OR**, тъй като неговата стойност зависи от начина, по който се подреждат данните в таблицата.

 **При изчисляване на *r* при качествени променливи с повече от две разновидности** обикновено се използват софтуерни продукти – напр. IBN SPSS, която предлага голямо разнообразие на подходи за определяне на различни видове корелационни зависимости.

**12.2.2. Рангов коефициент на корелация на Спирман**

 Ранговият коефициент на корелация на Спирман ***(ρ)*** се прилага при категорийни променливи, представени на ординална скала. Използва се също при количествени променливи или при една количествена и една качествена променлива. В основата на методиката лежи ранжирането, т.е. превръщането, както на качествените, така и на количествените променливи в рангове (присвояване на рангови номера). Прилага се следната формула:

 , където:

***n*** - брой изследвани случаи

***d*** - разлика между ранговете на ***x*** и ***y***

 Стойностите на ранговия коефициент за корелация (***ρ***) са от -1 до +1. Степента и посоката на връзката се определя по скалата в зависимост от знака пред ***ρ***  - еднопосочна при знак (**+**) и разнопосочна при знак (**-**).

 **Пример**: Проведено е проучване за установяване на връзката между социално-икономическия статус и тежестта на дадено заболяване при 8 лица ***(табл. 12.2).***

***Табл.12.2. Връзка между социално-икономическия статус и***

***тежестта на заболяване***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Изследвани лица** | **Ранг на *х*** | **Ранг на *у*** | **Разлика м/у *х* и *у (d)*** | ***(d2)*** |
| 1 | **6** | **5** | **1** | **1** |
| 2 | 7 | 8 | -1 | 1 |
| 3 | 2 | 4 | -2 | 4 |
| 4 | 3 | 3 | 0 | 0 |
| 5 | 5 | 7 | -2 | 4 |
| 6 | 4 | 1 | 3 | 9 |
| 7 | 1 | 2 | -1 | 1 |
| 8 | 8 | 6 | 2 | 4 |
| ***Общо*** | ***∑ d2 = 24*** |

 За да се определи силата и посоката на връзката между променливите ***х*** и ***у***, се преминава през следната последователност на изчисленията:

1. В колона 1 са изследваните лица (с номера от 1 до 8).
2. В колона 2 и 3 са представени ранговете на стойностите, характеризиращи социално-икономическия статус и заболяването: най-ниският социално-икономически статус е с ранг 1, най-високият – с ранг 8; по същия начин - най-ниската стойност на показателя за тежестта на заболяването е с ранг 1, а най-високата стойност – с ранг 8).
3. В колона 4 е разликата между ранговете (колона 2 – колона 3).
4. В колона 5 – квадрата на разликата и сумата е представена в последния ред.
5. Изчислява се коефициента за рангова корелация по формулата:

 

 Надеждността на ***ρ*** се определя като изчислената стойност се сравнява с теоретичната стойност ***(Приложение 3).*** При **k**=8–2=6 теоретичната стойност на ***ρ*** е 0,71 при Р=95% и 0,83 при Р=99%. Изчислената стойност 0.72е по-голяма от 0,71. Следователно, между социално-икономическия статус и тежестта на заболяването има значима (при Р=95%), голяма права връзка.

**12.2.3. Коефициент на корелация на Пирсон при количествени променливи величини**

 Изчисляването на ***r*** при количествени променливие доста по-сложно и за целта се използват подходящи софтуерни програми – напр. SPSS. Изследователят трябва да подготви из­хо­д­ни­те дан­ни в подходящ вид. На вся­ко зна­че­ние на ***х*** (независимата променлива) трябва да съ­о­т­ве­т­с­т­ва оп­ре­де­ле­но зна­че­ние на ***у*** (зависимата променлива). Софтуерните програми представят пряко и достоверността на изчисления коефициент без да е нужно изследователят да сравнява изчисления коефициент на корелация с таблични стойности.

**12.2.4. Коефициент на детерминация**

Коефициентът на детерминация представлява *мярка за степента, в която варирането (промените) на променливата* ***х*** *се свързва с варирането на втората променлива* ***у****.* Означава се с **KDET** и се изчислява по формулата:

 **(KDET)** = 100***r2***

 Например, ако ***r***=0,9, то **KDET** = 100 х *0,92* = 100 х 0,81 = 81%, т. е. 81% от промените в ***у*** се дължат на промени в ***х***.

# 12.3. Регресионен анализ – същност и предназначение

В пре­д­хо­д­ния раз­дел бе об­съ­ж­дан въ­п­ро­сът за из­мер­ва­не на връ­з­ка­та ме­ж­ду две ко­ли­че­с­т­ве­ни про­мен­ли­ви, ко­е­то е най-че­с­то сре­ща­на­та си­ту­а­ция при ана­ли­за на при­чин­на­та обу­с­ло­ве­ност на яв­ле­ни­я­та.

 *Про­мен­ли­ва­та, ко­я­то е резултат от въздействието е* ***за­ви­си­ма про­мен­ли­ва****, от­бе­ля­з­ва се с* ***у*** *и в ди­а­г­ра­мата на раз­сей­ва­не­то стой­но­с­ти­те й се на­на­сят вър­ху вер­ти­кал­на­та ос на ко­ор­ди­на­т­на­та си­с­те­ма. Про­мен­ли­вата, ко­я­то ока­з­ва вли­я­ние вър­ху зависимата, се на­ри­ча* ***не­за­ви­си­ма про­мен­ли­ва****, от­бе­ля­з­ва се с* ***х*** *и се изо­б­ра­зя­ва вър­ху хо­ри­зон­тал­на­та ос.*

 Гра­фи­ч­но­то пре­д­с­та­вя­не на две­те про­мен­ли­ви и из­чи­с­ля­ва­не­то на ко­е­фи­ци­ен­та на ко­ре­ла­ция да­ват пре­д­с­та­ва за на­ли­чи­е­то или от­съ­с­т­ви­е­то на вза­и­мо­в­ръ­з­ка ме­ж­ду ко­ли­че­с­т­ве­ни­те про­мен­ли­ви, но твър­де че­с­то изследователите се нуждаят от по-точно описание на зависимостите – например: По какъв на­чин въз­ра­ст­та вли­яе вър­ху ни­во­то на кръ­в­но­то на­ля­га­не? Ка­к­ва е то­ч­но връ­з­ка­та ме­ж­ду тю­тю­но­пу­ше­не­то и смър­т­но­ст­та от рак на бе­лия дроб и ка­к­ва про­мя­на в смър­т­но­ст­та мо­жем да оча­к­ва­ме при оп­ре­де­ле­но сни­же­ние на че­с­то­та­та на тю­тю­но­пу­ше­не­то в по­пу­ла­ци­я­та? и др. От­го­во­ри­те на такива въ­п­ро­си изи­с­к­ват за­дъл­бо­чено из­с­ле­д­ва­не на връ­з­ки­те ме­ж­ду за­ви­си­ми­те и не­за­ви­си­ми про­мен­ли­ви чрез мо­де­ли­ра­не с помощта на ***ре­г­ре­си­он­ен ана­ли­з, най-популярният вид на който е обикновената линейна регресия.***

 ***Същността на моделирането чрез обикновена линейна регресия е да се въз­про­из­ве­де пра­ва ли­ния, ко­я­то най-до­б­ре съ­о­т­ве­т­с­т­­ва на ди­а­г­ра­ма­та на раз­сей­ва­не, т.е. да се намери та­ка­ва пра­ва ли­ния, при ко­я­то су­ма­та от ква­д­ра­ти­те на вер­ти­кал­ни­те раз­с­то­я­ния от вся­ка то­ч­ка до ли­ни­я­та да е най-мал­ка.***

За да опи­шем ко­я­то и да е пра­ва ли­ния, тря­б­ва да зна­ем две стой­но­с­ти:

 - ***b, на­к­ло­нът на ли­ни­я­та,*** т. е. кол­ко стръм­но на­ра­с­т­ва ли­ни­я­та (по­ло­жи­те­лен на­к­лон) или на­ма­ля­ва (от­ри­ца­те­лен на­к­лон). Наклонът отразява с кол­ко на­ра­с­т­ва/на­ма­ля­ва) ***у*** за вся­ка еди­ни­ца про­мя­на в ***х***.

 - ***а, пре­се­ч­на­та то­ч­ка на ли­ни­я­та*,** т. е. от­къ­де за­по­ч­ва ли­ни­я­та. То­ва е стой­но­ст­та на ***у*** при ***х*** ра­в­но на ну­ла.

 На тази основа се съставя ура­в­не­нието на пра­ва­та ли­ния: ***у*** ***= а + b х*** при положителна или ***у*** ***= а - b х*** при отрицателна корелация.

 Сле­до­ва­тел­но, опи­са­ни­е­то на връ­з­ка­та се све­ж­да до на­ми­ра­не на стой­но­с­ти­те на ***а*** и ***b***, ко­и­то оп­ре­де­лят линията на ре­г­ре­си­я.

 Подходът за оп­ре­де­ля­не на ли­ни­я­та, ко­я­то най-до­б­ре опи­с­ва да­де­на ди­а­г­ра­ма на раз­сей­ва­не, се свежда до намиране на ли­ни­ята, ко­я­то най-до­б­ре при­ля­га към отделните то­ч­ки. Този подход се нарича ***“Метод на най-малките квадрати (Least Square Differences)“*** и се използва широко при изучаване на динамични промени и прогнозиране на бъдещото развитие на изучаваните явления***.*** Пример за това е представен в следващата глава 13.

 Стой­но­с­ти­те на ***а*** и ***b***, ко­и­то оп­ре­де­лят ли­ни­я­та на най-мал­ки­те ква­д­ра­ти, се из­чи­с­ляват с помощта на съ­о­т­ве­тни софтуерни про­дукти.

 ***Ко­е­фи­ци­ен­тът b*** е ***ко­е­фи­ци­ент на ре­г­ре­сия*** и не­го­во­то сми­с­ло­во зна­че­ние е да ***из­мер­ва ко­ли­че­с­т­ве­но с кол­ко се про­ме­ня у (за­ви­си­ма­та про­мен­ли­ва) при про­мя­на на х с еди­ни­ца***. По­ло­жи­тел­ни­ят знак пред ко­е­фи­ци­ен­та ***b*** по­ка­з­ва, че с на­ра­с­т­ва­не (намаляване) на ***х*** с еди­ни­ца ***у*** на­ра­с­т­ва (намалява) със стой­но­ст­та на ***b***. Об­ра­т­но - от­ри­ца­тел­ни­ят знак по­ка­з­ва, че с на­ра­с­т­ва­не (намаляване) на ***х*** с еди­ни­ца ***у*** на­ма­ля­ва (нараства) то­ч­но с тол­ко­ва, кол­ко­то е стой­но­ст­та на ***b***.

 **Видове регресионни модели**

 1. Според броя на включените в анализа фактори:

1. единични или еднофакторни - изследва се връзката между едно следствие ***(у)*** и една причина ***(х)*** исепредставят: ***у = f (х1, е)***
2. множествени или многофакторни - изследват връзките между едно следствие (зависима променлива) и две или повече причини (фактори) - ***х1, х2...... хn****.* Представят се с израза: ***у = f (x1 , x2 .......... xi , e)***

 2. Според формата на връзката (типа на модела):

* линейни модели - представят се с уравнението на права линия в равнината или линеен полином в пространството. Еднофакторният линеен модел се представя с израза ***y = а + вi х + е,*** където:
	+ ***y*** и***х*** са изследваните явления;
	+ ***в1* ... *вi*** са параметри на модела (регресионни коефициенти) и се представят в същите мерни единици както резултатът ***y****;*
	+ ***а*** е свободен член в уравнението, несвързан пряко с някоя величина, няма определен съдържателен смисъл, защото се формира от влиянието на много и разнопосочно действуващи причини (в стойността му е акумулирано влиянието на фактори, невключени в модела);

***- е*** случаен компонент (отчита влиянието на случайни причини).

* + - нелинейни модели - представят се с някакво уравнение на крива в равнината или с нелинеен полином в пространството.

Връзката между ***y*** и ***x*** напр. може да се представи с парабола от втора степен ***y = а + в1 х + в2 х2+ е***

На ***фиг. 12.2*** са представени някои видове регресионни линии и съответните им уравнения.



***Фиг. 12.2. Избрани регресионни линии и съответните им уравнения***

##### 12.4. Въпроси за самоподготовка

1. За определяне коефициента на корелация между ***x и y*** е необходимо да разполагаме с:

А. група лица, някои от които притежават променливата ***x***, а останалите - променливата ***y***

Б. измервания на променливата ***x*** при една група лица и променливата ***y*** при друга група

##### В. измервания на променливите *x и y* при всяко лице в една и съща група

2. Диаграмата на разсейването представлява графично средство за откриване и анализиране на връзка между две променливи.

А. вярно Б. невярно

3. Даден изследовател установява, че корелацията между личностовите черти “алчност” и “надменност” е -0.40. Какъв процент от варирането в алчността може да бъде обяснено чрез връзката с надменността?

А. 60% Б. 0% В. 16% Г. 20% Д. 40%

4. Каква част от варирането на зависимата променлива ***y*** е свързана с независимата променлива ***x***, ако коефициент на корелация между двете променливи е r = 0.5?

А. най-голямата част

Б. половината

В. много малка част

Г. една четвърт (25%)

5. Коя стойност на r показва най-силна зависимост на една променлива от друга?

А. r = 1.18 Б. r = -0.77В. r = 0.68 Г. r = 0.45

6. Ако корелацията между възрастта на колите и средствата за ремонт е +0.90. то:

##### А. 81% от изразходваните средства за ремонт се обясняват с възрастта на колите

Б. 81% от изразходваните средства за ремонт не зависят от възрастта на колите

В. 90% от изразходваните средства за ремонт може да се обяснят с възрастта на колите

7. Ако коефициентът на корелация е 0.61, това показва, че пропорцията от варирането в зависимата променлива, която се обяснява с варирането в независимата променлива е:

А. 37% Б. 61% В. 98% Г. не може да се определи

8. Да предположим, че между средния успех на учащите се в един колеж и резултатите от представянето им на тест за интелигентност е установен коефициент на корелация r *=* 0.40. Какъв общ процент вариране имат тези две променливи?

А. 20 Б. 16 В. 40 Г. 80

9. Ако в даден експеримент r = 0.70, то тогава 49% от варирането на зависимата променлива У може да се обясни с различията в променливата Х.

А. вярно Б. невярно

10. Коефициентът на корелация може да има стойности между –1 и + 1.

А. вярно Б. невярно

11. Коефициентът на детерминация може да има стойности между –1 и + 1.

А. вярно Б. невярно

12. Коефициентът на линейна корелация между ***x*** и ***y*** е –0.95. Кое от следните заключения е вярно?

А. линейната връзка между ***x*** и ***y*** е слаба и ***y*** намалява, когато ***x*** нараства.

Б. линейната връзка между ***x***и***y*** е силна и***y*** намалява, когато *x* нараства.

В. линейната връзка между ***x*** и ***y*** е силна и ***y*** нараства, когато ***x*** нараства.

13. Ако коефициентът на корелация между ***x*** и ***y*** в извадка е – 1, то:

А. няма връзка между ***x*** и ***y*** в извадката

Б. няма връзка между ***x*** и ***y*** в популацията

В. има пълна (функционална) отрицателна връзка между ***x*** и ***y*** в популацията

##### Г. има пълна (функционална) отрицателна връзка между *x* и *y* в извадката

14. При проучване на връзката между IQ (коефициент на интелигентност) и средния успех от следването в един колеж, кой от следните статистически анализи е най-подходящ?

##### А. коефициент на корелация

Б. дисперсионен анализ

В. t-критерий

15. При проучване на връзката между самооценката за подготовката и постиженията при лица в дадена извадка е изчислен коефициент на корелация r = 0.75. Това показва, че връзката между тези две променливи е:

А. слаба и положителна

Б. силна и положителна

В. слаба и отрицателна

Г. силна и отрицателна

16. Корелацията между изпълнението на сложни задачи (***x***) и състоянието на тревожност (***y***) е r = -0.73. Какво заключение може да се направи?

А. нарастването на тревожността подобрява изпълнението на сложни задачи

##### Б. подобряването на изпълнението на сложни задачи намалява тревожността

В. високо ниво на тревожност води до лошо изпълнение на сложни задачи

17. Коефициентът на корелация на Пирсон r = -0.75 означава, че:

А. индивидите с високи стойности на едната променлива проявяват тенденция към ниски стойности на другата променлива

Б. индивидите с високи резултати за едната променлива проявяват тенденция към високи резултати и за другата променлива

В. няма връзка между двете променливи

18. Изчислен е коефициент на корелация r = -1.08 между резултатите от кандидат-студентските изпити и успеваемостта на студентите през следващите години. Какъв е изводът?

А. резултатите от кандидат-студентските изпити са добър прогностичен признак за успех

##### Б. университетът трябва да си наеме друг статистик

В. резултатите от кандидат-студентските изпити са лош прогностичен признак за успех

19. Под диаграмата на разсейване е отбелязано r = +0.10. Това означава:

А. плюс и минус 10% от средните стойности включват 68% от случаите

Б. една десета от дисперсията на дадена променлива се споделя с другата променлива

В. една десета от едната променлива е причинена от другата променлива

Г. степента на линейна връзка между променливите е слабо положителна

20. Коя от посочените стойности на коефициента на корелация показва ситуация, където повече от половината от промените в едната величина са свързани с промени в другата?

А. r = -0.7 Б. r = 0.3 В. r = -0.9 Г. r = - 0.6

21. Ако коефициентът на корелация между ***x*** и ***y*** е нула, то:

А. ***x*** и ***y*** имат стандартни разпределения

Б. дисперсиите на ***x*** и ***y*** са еднакви

В. има връзка между ***x*** и ***y***

##### Г. няма линейна връзка между *x* и *y*

22. Между променливите x и y е изчислен r = - 0.90. Това показва, че:

А. изчислението е грешно, тъй като r не може да бъде отрицателно число

Б. има много слаба връзка между двете променливи

В. коефициентът на детерминация е равен на корен квадратен от 0.90

##### Г. променливата y има тенденция да намалява при нарастване на x

23. Коефициент на корелация r = - 0.80:

А. е по-малък от r = + 0.80

Б. показва същата сила на връзка както r = + 0.80

24. Коя от следните стойности най-добре характеризирала връзката между променливите “брой отработени часове” и “количество извършена работа”?

##### А. r = 0.9

Б. r = 0.4

В. r = 0.0

25. Знакът (плюс или минус) пред коефициента на корелация показва:

##### А. посоката на корелационната зависимост

Б. практическата значимост на зависимостта

В. вероятността, че степента на връзката е по-голяма от нула

26. Корелацията между резултати от тест за неврастения и тест за тревожност е силна и положителна. Следователно:

А. тревожността не причинява неврастения

Б. лицата с ниски резултати при единия тест показват високи резултати при другия тест

В. лицата с ниски резултати от единия тест показват ниски резултати и при другия

27. При корелационния анализ, когато точките се разпръскват широко около линията на регресия, това означава че корелацията е:

А. отрицателна

Б. слаба

В. силна

28. Ако r е близо до +1 или –1, казваме, че има силна корелация, подразбирайки безусловно, че говорим само за линейна зависимост.

А. вярно Б. невярно

29. Ако знаем, че коефициентът на корелация при обикновена линейна зависимост е положителен за две променливи (***x*** и ***y*** ), тогава регресията на ***y*** по отношение на ***x*** ще се представи чрез линия на регресия, която е наклонена отгоре надолу, започвайки отляво надясно.

А. вярно Б. невярно

30. Ако знаем, че коефициентът на корелация при обикновена линейна зависимост е положителен за две променливи (***x*** и ***y***), тогава регресията на ***y*** по отношение на ***x*** ще се представи чрез линия на регресия, която е наклонена отдолу нагоре, започвайки отляво надясно.

А. вярно Б. невярно

31. Когато r е изчислен от извадка, стойността, която получаваме за r е само една оценъчна стойност за истинския параметър на корелация за цялата популация.

А. вярно Б. невярно

32. При сра­в­ня­ва­не на хи-ква­д­рат и ко­е­фи­ци­ен­та на ко­ре­ла­ция (r) вяр­но е:

А. ко­е­фи­ци­ен­тът на ко­ре­ла­ция е с по-мал­ка мо­щ­ност от хи-ква­д­рат

Б. ко­е­фи­ци­ен­тът на ко­ре­ла­ция е с по-го­ля­ма мо­щ­ност от хи-ква­д­рат

В. ня­ма раз­ли­ка в мо­щ­но­ст­та на два­та кри­те­рия

33. Кое от по­со­че­ни­те твър­де­ния не е вяр­но?

А. ко­е­фи­ци­ен­тът на ко­ре­ла­ция е с по-го­ля­ма мо­щ­ност от хи-квадрат

##### Б. ко­е­фи­ци­ен­тът на ко­ре­ла­ция е с по-мал­ка мо­щ­ност от хи-квадрат

В. има раз­ли­ка в мо­щ­но­ст­та на два­та кри­те­рия

34. Ко­ре­ла­ци­он­ни­ят ана­лиз се при­ла­га:

А. са­мо при ка­че­с­т­ве­ни про­мен­ли­ви

Б. са­мо при ко­ли­че­с­т­ве­ни про­мен­ли­ви

##### В. и при два­та ви­да про­мен­ли­ви

35. При избор на коефициент за ко­ре­ла­ция за кон­к­ре­т­ни дан­ни, тря­б­ва да се взема предвид:

А. да­ли за­ви­си­мо­ст­та е ли­ней­на или не­ли­ней­на

Б. ви­да на ска­ла­та за из­мер­ва­не на вся­ка про­мен­ли­ва

##### В. и две­те ус­ло­вия

36. Най-по­д­хо­дящ за изу­ча­ва­не на вза­и­мо­в­ръ­з­ка ме­ж­ду две про­мен­ли­ви ве­ли­чи­ни е:

А. сре­д­на ари­т­ме­ти­ч­на ве­ли­чи­на

Б. ко­е­фи­ци­ент на корелация

В. ко­е­фи­ци­ент за че­с­то­та

37. Оп­ре­де­ле­те сте­пен­та и по­со­ка­та на ко­ре­ла­ци­ята при r = + 0.65, като използвате 5-степенна скала за оценка:

А. уме­ре­на пра­ва връ­з­ка

Б. уме­ре­на об­ра­т­на връ­з­ка

В. зна­чи­тел­на пра­ва връ­з­ка

38. Оп­ре­де­ле­те сте­пен­та и по­со­ка­та на ко­ре­ла­ци­он­на връ­з­ка при ко­е­фи­ци­ент на ко­ре­ла­ция r = - 0.23, ако използвате 5-степенна скала за оценка:

А. сла­ба об­ра­т­на връ­з­ка

Б. уме­ре­на пра­ва връ­з­ка

В. зна­чи­тел­на об­ра­т­на връ­з­ка

39. Ка­з­ват Ви, че съ­ще­с­т­ву­ва силна об­ра­т­на връ­з­ка ме­ж­ду про­мен­ли­ви­те “ко­ли­че­с­т­во фи­зи­че­с­ки уп­ра­ж­не­ния” и “че­с­то­та на сър­де­ч­ни за­бо­ля­ва­ния”. Кой от сле­д­ва­щи­те ко­ре­ла­ци­он­ни ко­е­фи­ци­ен­ти съ­о­т­ве­т­с­т­ва на то­ва твър­де­ние?

А. r = +0.8 Б. r = - 0.3 В. r = - 0.8 Г. r = + 0.9

40. При от­ри­ца­тел­на ко­ре­ла­ци­он­на за­ви­си­мост:

А. с на­ра­с­т­ва­не на ***х,*** на­ра­с­т­ва ***у***

##### Б. с на­ра­с­т­ва­не на *х*, на­ма­ля­ва *у*

В. с на­ма­ля­ва­не на ***х***, на­ма­ля­ва ***у***

41. Кой от сле­д­ва­щи­те ко­ре­ла­ци­он­ни ко­е­фи­ци­ен­ти от­ра­зя­ва най-сла­ба за­ви­си­мост?

А. r = - 0.60 Б. r = - 0.33 В. r = + 0.29 Г. r = + 0.5

42. Ако ко­ре­ла­цията ме­ж­ду ***x*** и ***y*** е по­ло­жи­тел­на, за­ви­си­ма­та про­мен­ли­ва ***у*** на­ма­ля­ва:

А. при на­ра­с­т­ва­не на ***х***

Б. за­па­з­ва­не на ***х*** на ед­но и съ­що ни­во

В. при на­ма­ля­ва­не на ***х***

43. Из­бе­ре­те най-по­д­хо­дя­щия ста­ти­с­ти­че­с­ки тест за оп­ре­де­ля­не на връ­з­ка­та ме­ж­ду ни­во­то на си­с­то­ли­ч­но­то на­ля­га­не и се­рум­ния хо­ле­с­те­рол ка­то има­те пре­д­вид, че и две­те про­мен­ли­ви ве­ли­чи­ни са из­ра­зе­ни ко­ли­че­с­т­ве­но:

А. ко­е­фи­ци­ент на ран­го­ва ко­ре­ла­ция на Спир­ман

##### Б. ко­е­фи­ци­ент на ко­ре­ла­ция на Пир­сон

В. хи-ква­д­рат

44. Кое от сле­д­ни­те твър­де­ния за ко­ре­ла­ци­я­та е ­вяр­но?

А. Ко­ре­ла­ци­о­нен ко­е­фи­ци­ент r = - 0.8 по­ка­з­ва по-слаба сте­пен на връ­з­ка ме­ж­ду две про­мен­ли­ви, от­кол­ко­то r = + 0.6.

Б. Ко­е­фи­ци­ен­тът на ко­ре­ла­ция на Пирсон се из­по­л­з­ва при количествени променливи

В. Ди­а­г­ра­мата на раз­сей­ва­не­ е по­ле­з­но за оце­нка на вида на връзката ме­ж­ду две про­мен­ли­ви – дали е пра­во­ли­ней­на или кри­во­ли­ней­на.

45. При из­чи­с­ля­ва­не на ко­е­фи­ци­ен­та за ран­го­ва ко­ре­ла­ция на Спир­ман из­хо­д­ни­те дан­ни тря­б­ва да бъ­дат пре­д­с­та­ве­ни в:

А. но­ми­нал­на ска­ла

##### Б. ор­ди­нал­на ска­ла

В. ин­тер­вал­на ска­ла

46. Ре­г­ре­си­он­ни­ят ана­лиз слу­жи за:

А. ус­та­но­вя­ва­не на до­с­то­вер­ност на раз­ли­чи­я­та

Б. из­мер­ва­не на цен­т­рал­на тен­ден­ция

##### В. мо­де­ли­ра­не на ли­ней­ни за­ви­си­мо­с­ти

47. За ко­ли­че­с­т­ве­но из­мер­ва­не на про­ме­ни­те в за­ви­си­ма­та про­мен­ли­ва (***y***) при еди­ни­ца про­мя­на на не­за­ви­си­ма­та про­мен­ли­ва (***x)*** се из­по­л­з­ва:

А. стан­дар­т­но от­к­ло­не­ние

Б. ко­е­фи­ци­ент на ко­ре­ла­ция

В. ко­е­фи­ци­ент на ре­г­ре­сия

48. Корелацията се дефинира като относителна разлика между две променливи.

А. вярно Б. невярно

49. Връзката между две променливи може да се представи чрез диаграма на разсейването.

А. вярно Б. невярно

50. Когато говорим за положителна зависимост, това означава, че високи стойности на едната променлива са свързани с високи стойности на другата променлива.

А. вярно Б. невярно

**Отговори на въпросите от глава 12:**

1Б; 2А; 3В; 4Г; 5Б; 6А; 7А; 8Б; 9А; 10А; 11Б; 12Б; 13Г; 14А; 15Б; 16В; 17А; 18Б; 19Г; 20В; 21Г; 22Г; 23Б; 24А; 25А; 26В; 27Б; 28А; 29Б; 30А; 31А; 32Б; 33Б; 34В; 35В; 36Б; 37В; 38А; 39В; 40Б; 41В; 42В; 43Б; 44А; 45Б; 46В; 47В; 48Б; 49А; 50А