

## 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 по-

ложително 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 –  $y = a + bx$  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 (-) е 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 |
|----------------------|--------------------------------------|--------------------------------------|
| слаба                | до 0,3                               | под 0.3                              |
| умерена              | от 0,31 до 0,5                       | от 0.31 до 0.7                       |
| значителна           | от 0,51 до 0,7                       | над 0.7                              |
| голяма               | от 0,71 до 0,9                       |                                      |
| изключително голяма  | над 0,9                              |                                      |

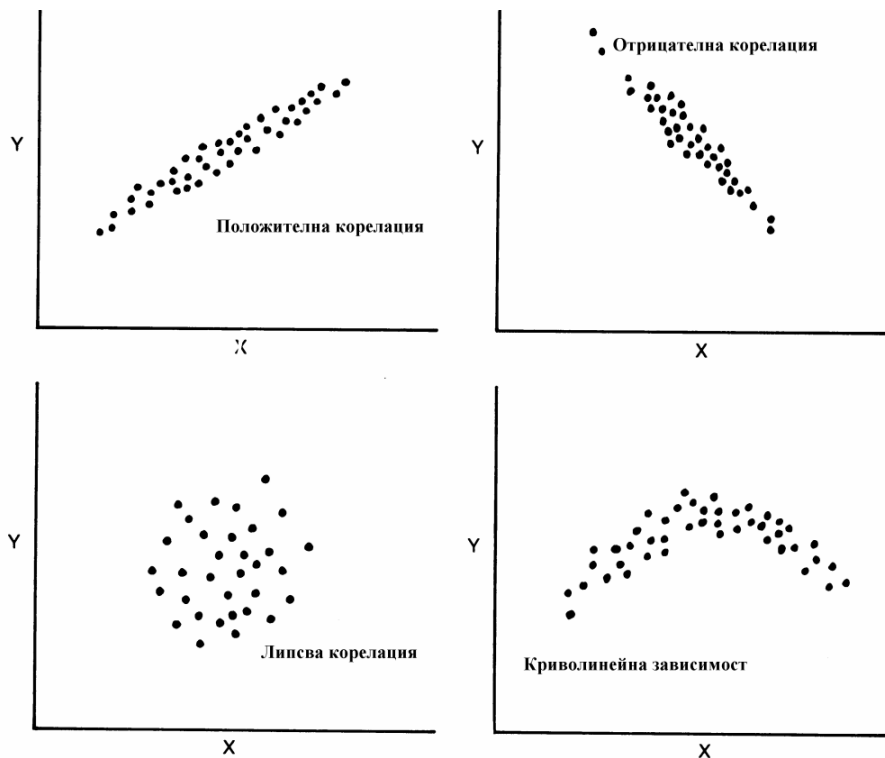
#### 12.1.4. Диаграма на разсейване

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 зависимите 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  $x$  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  $2 \times 2$ :

|         |              |              |                |
|---------|--------------|--------------|----------------|
| $x / y$ | $y_1$        | $y_2$        | Общо           |
| $x_1$   | <b>a</b>     | <b>b</b>     | <b>a + b</b>   |
| $x_2$   | <b>c</b>     | <b>d</b>     | <b>c + d</b>   |
| Общо    | <b>a + c</b> | <b>b + d</b> | <b>a+b+c+d</b> |

Следователно:

**a** е съчетание на  $x_1 y_1$

**b** е съчетание на  $x_1 y_2$

**c** е съчетание на  $x_2 y_1$

**d** е съчетание на  $x_2 y_2$

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

$$C = \sqrt{\frac{\chi^2_{em}}{n}}, \text{ където:}$$

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

$\chi^2_{em}$  - изчислената стойност от емпиричните данни (виж раздел 12.5.2)

Когато от четирикратната таблица има изчислена стойност на  $\chi^2$ , 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, умирање и др.) при експонирани (изложени на рисков фактор) и при неекспонирани лица.

Използвайки посочените символи в клетките на четирикратната таблица:

$$OR = \frac{ad}{bc}$$

**Ако OR е 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 експонирани 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 ( $y$ ) при две групи болни ( $x$ ): лекувани само с антибиотици ( $x_1$ ) и лекувани с антибиотици и витамини ( $x_2$ ). 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 ( $y_1$ ) | Без странични явления ( $y_2$ ) | Общо |
|--------------------------------------|---------------------------------|---------------------------------|------|
| С антибиотици и витамини - ( $x_1$ ) | 9                               | 57                              | 66   |
| Само с антибиотици - ( $x_2$ )       | 16                              | 29                              | 45   |
| Общо                                 | 25                              | 86                              | 111  |

$$OR = \frac{ad}{bc} = \frac{29 \times 9}{16 \times 57} = 0.29$$

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

Стойностите на **C** и **r** ( $\varphi$ ) 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** ( $\varphi$ ) 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. IBM 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 ( $\rho$ ) 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:

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}, \text{ където:}$$

$n$  - брой изследвани случаи

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

Стойностите на ранговия коефициент за корелация ( $\rho$ ) са от -1 до +1. Степента и посоката на връзката се определя по скалата в зависимост от знака пред  $\rho$  - 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 $x$ | Ранг на $y$ | Разлика м/у $x$ и $y$ ( $d$ )     | ( $d^2$ ) |
|-----------------|-------------|-------------|-----------------------------------|-----------|
| 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         |
| <b>Общо</b>     |             |             | <b><math>\sum d^2 = 24</math></b> |           |

За да се определи силата и посоката на връзката между променливите  $x$  и  $y$ , 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:

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 24}{8 \times (8^2 - 1)} = 0.72$$

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

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

Изчисляването на  $r$  при количествени променливи е 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  $x$  (независимата променлива) трябва да съответства определено значение на  $y$  (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  $x$  се свързва с варирането на втората променлива  $y$ . Означава се с  $K_{\text{ДЕТ}}$  и се изчислява по формулата:

$$(K_{\text{ДЕТ}}) = 100r^2$$

Например, ако  $r = 0,9$ , то  $K_{\text{ДЕТ}} = 100 \times 0,9^2 = 100 \times 0,81 = 81\%$ , т. е. 81% от промените в  $y$  се дължат на промени в  $x$ .

## 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  $y$  и в диаграмата на разсейването стойностите ѝ се нанасят върху вертикалната ос на координатната система. Променливата, която оказва влияние върху зависимата, се нарича **независима променлива**, отбелязва се с  $x$  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  $y$  за всяка единица промяна в  $x$ .

-  **$a$ , пресечната точка на линията**, т. е. откъде започва линията. Това е стойността на  $y$  при  $x$  равно на нула.

На тази основа се съставя уравнението на правата линия:  $y = a + b x$  при положителна или  $y = a - b x$  при отрицателна корелация.

Следователно, описанието на връзката се свежда до намиране на стойностите на  $a$  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  $a$  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  $y$  (зависимата променлива) при промяна на  $x$  с единица*. Положителният знак пред коефициента  $b$  показва, че с нарастване (намаляване) на  $x$  с единица  $y$  нараства (намалява) със стойността на  $b$ . Обратно - отрицателният знак показва, че с нарастване (намаляване) на  $x$  с единица  $y$  намалява (нараства) точно с толкова, колкото е стойността на  $b$ .

### Видове регресионни модели

#### 1. Според броя на включените в анализа фактори:

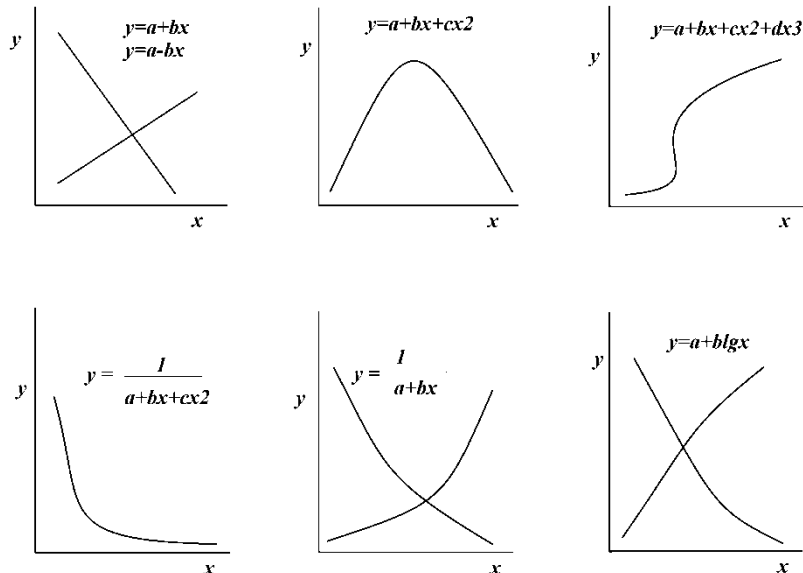
- **единични или еднофакторни** - изследва се връзката между едно следствие ( $y$ ) и една причина ( $x$ ) и се представят:  $y = f(x, e)$
- **множествени или многофакторни** - изследват връзките между едно следствие (зависима променлива) и две или повече причини (фактори) -  $x_1, x_2, \dots, x_n$ . Представят се с израза:  $y = f(x_1, x_2, \dots, x_n, e)$

#### 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  $y = a + v_i x + e$ , където:
  - $y$  и  $x$  са изследваните явления;
  - $v_1 \dots v_i$  са параметри на модела (регресионни коефициенти) и се представят в същите мерни единици както резултатът  $y$ ;
  - $a$  е свободен член в уравнението, несвързан пряко с някоя величина, няма определен съдържателен смисъл, защото се формира от влиянието на много и разнопосочно действащи причини (в стойността му е акумулирано влиянието на фактори, невключени в модела);
  - $e$  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  $y = a + v_1 x + v_2 x^2 + e$

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

- За определяне коефициента на корелация между  $x$  и  $y$  е необходимо да разполагаме с:
  - група лица, някои от които притежават променливата  $x$ , а останалите - променливата  $y$
  - измервания на променливата  $x$  при една група лица и променливата  $y$  при друга група
  - измервания на променливите  $x$  и  $y$  при всяко лице в една и съща група
- Диаграмата на разсейването представлява графично средство за откриване и анализиране на връзка между две променливи.
  - вярно
  - невярно
- 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?

- A. 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  $r = 0.5$ ?
- A. 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?
- A.  $r = 1.18$       Б.  $r = -0.77$       В.  $r = 0.68$       Г.  $r = 0.45$
6. Ако корелацията между възрастта на колите и средствата за ремонт е  $+0.90$ . то:
- A. 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:
- A. 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?
- A. 20      Б. 16      В. 40      Г. 80
9. Ако в даден експеримент  $r = 0.70$ , то тогава 49% от варирането на зависимата променлива  $Y$  може да се обясни с различията в променливата  $X$ .
- A. вярно      Б. невярно
10. Коефициентът на корелация може да има стойности между  $-1$  и  $+1$ .
- A. 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 характеризира връзката между променливите “брой отработени часове” и “количество извършена работа”?

- A.  $r = 0.9$
- Б.  $r = 0.4$
- В.  $r = 0.0$

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

- A. 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:

- A. 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:

- A. 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.

- A. 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.

- A. 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.

- A. 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  $x$ , нараства  $y$

Б. с нарастване на  $x$ , намалява  $y$

В. с намаляване на  $x$ , намалява  $y$

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  $y$  намалява:

А. при нарастване на  $x$

Б. запазване на  $x$  на едно и също ниво

В. при намаляване на  $x$

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