

# Решения 2023/2024

## Домашна работа

ас.г-р Костагин Костадинов

### Упражнение 1 - Стандартизация

#### Задача 1

1. Да се определи възрастовата структура на преминалите болни през Болница X, като разполагате със следните данни:

| Възраст                  | 0-1 г. | 1-4 г. | 5-9 г. | 10-19 г. | Общо |
|--------------------------|--------|--------|--------|----------|------|
| Болни с Хепатит А        | 70     | 41     | 102    | 87       | 300  |
| Възрастова структура (%) | 23.33% | 13.67% | 34%    | 29%      | 100% |

#### Решение

Изчисляваме общия брой болни: 300

- 0-1 г. -  $(70/300) \cdot 100 = 23.33\%$
- 1-4 г. -  $(41/300) \cdot 100 = 13.67\%$
- 5-9 г. -  $(102/300) \cdot 100 = 34\%$
- 10-19 г. -  $(87/300) \cdot 100 = 29\%$

## Задача 2

Да се изчислят стандартизираните показатели за плодовитостта в районите А и Б. За стандарт да се приеме възрастовият състав на жените от **район Б**. Да се анализират получените резултати.

| Район А |        |            | Район Б |            |
|---------|--------|------------|---------|------------|
| Възраст | Жени   | Живородени | Жени    | Живородени |
| 15-19   | 1 000  | 18         | 1 200   | 22         |
| 21-30   | 9 000  | 225        | 7 000   | 175        |
| 31-49   | 8 000  | 128        | 10 000  | 160        |
| Всичко  | 18 000 | 371        | 18200   | 357        |

## Решение

### Стъпка 1 Изчисляване на нестандартизираните показатели (НС)

| Район А |        |            |       | Район Б |            |        |
|---------|--------|------------|-------|---------|------------|--------|
| Възраст | Жени   | Живородени | НС    | Жени    | Живородени | НС     |
| 15-19   | 1 000  | 18         | 0.018 | 1 200   | 22         | 0.018  |
| 21-30   | 9 000  | 225        | 0.025 | 7 000   | 175        | 0.025  |
| 31-49   | 8 000  | 128        | 0.016 | 10 000  | 160        | 0.016  |
| Всичко  | 18 000 | 371        | 0.021 | 18200   | 357        | 0.0196 |

Нестандартизираните показател за район А е 21‰, а в район **Б 19.6‰**.

### Стъпка 2 Изчисляване на стандарт (S)

| Район Б Стандарт |        |                      |
|------------------|--------|----------------------|
| Възраст          | Жени   |                      |
| 15-19            | 1 200  | $1200/18200 = 0.07$  |
| 21-30            | 7 000  | $7000/18200 = 0.38$  |
| 31-49            | 10 000 | $10000/18200 = 0.55$ |
| Всичко           | 18200  |                      |

### Стъпка 3 Изчисляване на стандартизираните показатели (С)

За район А

| Район А |        |            |       | Стандарт |            |        |       |
|---------|--------|------------|-------|----------|------------|--------|-------|
| Възраст | Жени   | Живородени | НС    | S        | S*НС       | C      | C (‰) |
| 15-19   | 1 000  | 18         | 0.018 | 0.07     | 0.018*0.07 | 0.0013 | 1.3   |
| 21-30   | 9 000  | 225        | 0.025 | 0.38     | 0.025*0.38 | 0.0095 | 9.5   |
| 31-49   | 8 000  | 128        | 0.016 | 0.55     | 0.016*0.55 | 0.0088 | 8.8   |
| Всичко  | 18 000 | 371        | 0.021 |          |            |        | 19.6  |

За район Б

| Район Б |        |            |        | Стандарт |            |        |       |
|---------|--------|------------|--------|----------|------------|--------|-------|
| Възраст | Жени   | Живородени | НС     | S        | S*НС       | C      | C (‰) |
| 15-19   | 1 200  | 22         | 0.018  | 0.07     | 0.018*0.07 | 0.0013 | 1.3   |
| 21-30   | 7 000  | 175        | 0.025  | 0.38     | 0.025*0.38 | 0.0095 | 9.5   |
| 31-49   | 10 000 | 160        | 0.016  | 0.55     | 0.016*0.55 | 0.0088 | 8.8   |
| Всичко  | 18200  | 357        | 0.0196 |          |            |        | 19.6  |

Стандартизираните показатели за двата района са равни - 19.6‰. Няма разлика в плодовитостта, след стандартизация по възрастови групи на жените в двата района.

### Задача 3

Да се изчислят стандартизираните показатели за леталитета в две градски болници - Болница А и Болница Б. За стандарт да се приеме съставът на болните в **болница Б**. Да се анализират получените резултати.

| Болест              | Преминали | Починали | Преминали | Починали |
|---------------------|-----------|----------|-----------|----------|
|                     | (А)       | (А)      | (Б)       | (А)      |
| Хипертонична болест | 180       | 4        | 200       | 4        |
| Рак на стомаха      | 100       | 30       | 90        | 27       |
| Инфаркт на миокарда | 120       | 8        | 160       | 10       |
| Всичко              | 400       | 42       | 450       | 41       |

## Решение

### Стъпка 1 Изчисляване на нестандартизираните показатели (НС)

| Болест              | Преминали | Починали | НС леталитет | Преминали | Починали | НС леталитет |
|---------------------|-----------|----------|--------------|-----------|----------|--------------|
|                     | (А)       | (А)      | (А)          | (Б)       | (Б)      | (Б)          |
| Хипертонична болест | 180       | 4        | 0.02         | 200       | 4        | 0.02         |
| Рак на стомаха      | 100       | 30       | 0.30         | 90        | 27       | 0.30         |
| Инфаркт на миокарда | 120       | 8        | 0.07         | 160       | 10       | 0.06         |
| Всичко              | 400       | 42       | 0.1          | 450       | 41       | 0.09         |

Нестандартизираният леталитет в болница А е 10 %, а в болница Б 9 %.

### Стъпка 2 Изчисляване на стандарт (S)

| Болест              | Преминали | Стандарт (S)     |
|---------------------|-----------|------------------|
|                     | (Б)       |                  |
| Хипертонична болест | 200       | $200/450 = 0.44$ |
| Рак на стомаха      | 90        | $90/450 = 0.2$   |
| Инфаркт на миокарда | 160       | $160/450 = 0.36$ |
| Всичко              | 450       |                  |

### Стъпка 3 Изчисляване на стандартизираните показатели (СА)

За болница А

| Болест              | Преминали | Починали | НС леталитет | Стандарт |             |      |        |
|---------------------|-----------|----------|--------------|----------|-------------|------|--------|
|                     | (А)       | (А)      | (А)          | S        | S*НС        | СА   | СА (%) |
| Хипертонична болест | 180       | 4        | 0.02         | 0.44     | $0.02*0.44$ | 0.01 | 1      |
| Рак на стомаха      | 100       | 30       | 0.30         | 0.2      | $0.3*0.2$   | 0.06 | 6      |
| Инфаркт на миокарда | 120       | 8        | 0.07         | 0.36     | $0.07*0.36$ | 0.03 | 3      |
| Всичко              | 400       | 42       | 0.1          |          |             |      | 10     |

За болница Б

| Болест              | Преминали | Починали | НС леталитет | Стандарт |                   |      |        |
|---------------------|-----------|----------|--------------|----------|-------------------|------|--------|
|                     | (Б)       | (Б)      | (Б)          | S        | $S \cdot HC$      | СЛ   | СЛ (%) |
| Хипертонична болест | 200       | 4        | 0.02         | 0.44     | $0.02 \cdot 0.44$ | 0.01 | 1      |
| Рак на стомаха      | 90        | 27       | 0.30         | 0.2      | $0.3 \cdot 0.2$   | 0.06 | 6      |
| Инфаркт на миокарда | 160       | 10       | 0.06         | 0.36     | $0.06 \cdot 0.36$ | 0.02 | 2      |
| Всичко              | 450       | 41       | 0.09         |          |                   |      | 9      |

## Задача 4

Нестандартизираните показатели за леталитет в две болници А и Б са съответно 10‰ и 12‰. След стандартизация, спрямо болница А, леталитетът в болница А е 10‰, а този в болница Б 8‰. Изчислете стандартизираните леталитети, ако за стандарт се ползва болница Б.

## Решение

Правило 1: Стандартизираният показател, за структурата която е избрана за стандарт е **еднакъв** с нестандартизирания.

След стандартизация за болница Б, стандартизираният показател остава 12‰.

Правило 2: Разликата в стандартизираните показатели остава същата, независимо промяната в стандарта.

Разликата между стандартизираните показатели е 2‰. Това означава, че стандартизираният показател за болница Б е 14‰.

## Упражнение 2 Дескриптивна и инферентна статистика

### Задача 1

След провеждане на тренировъчен режим с група спортисти в края на периода е измерено теглото в килограми. Получени са следните стойности: 64, 76, 68, 84, 80, 72, 74, 72 kg. Да се определи средното тегло на групата и границите на 95% от интервал на доверителност за средното тегло на спортистите в популацията.

## Решение

### Стъпка 1 Изчисляване на средното тегло

Прилагаме формата за средна аритметична при негрупирани данни:

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{64 + 76 + 68 + 84 + 80 + 72 + 74 + 72}{8} = \frac{590}{8} = 73.75$$

### Стъпка 2 Изчисляване на стандартното отклонение

Прилагаме формулата за стандартно отклонение при негрупирани данни:

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$SD = \sqrt{\frac{(64-73.75)^2 + (76-73.75)^2 + (68-73.75)^2 + (84-73.75)^2 + (80-73.75)^2 + (72-73.75)^2 + (74-73.75)^2 + (72-73.75)^2}{8}} =$$

$$SD = \sqrt{\frac{(-9.75^2) + (2.25^2) + (-5.75^2) + (10.25^2) + (6.25^2) + (-1.75^2) + (0.25^2) + (-1.75^2)}{8}} =$$

$$SD = \sqrt{\frac{95.0625 + 5.0625 + 33.0625 + 105.0625 + 39.0625 + 3.0625 + 0.0625 + 3.0625}{8}} =$$

$$SD = \sqrt{\frac{283.5}{8}} = \sqrt{35.44} = 6$$

### Стъпка 3 Изчисляване на средната грешка SE

Прилагаме формулата за средна грешка при негрупирани данни:

$$SE_m = \frac{SD}{\sqrt{n}}$$

$$SE_m = \frac{6}{\sqrt{8}} = 2.12$$

### Стъпка 4 Определяме стойността на гаранционния множител z

За 95% интервал  $z = 1.96$  (от листа с формулите)

### Стъпка 5 Определяме границите на интервала на доверителност

$$\bar{x} \pm z * SE_m$$

$$73.75 \pm 1.96 * 2.12 = 73.75 \pm 4.16$$

- Горна граница: 77.91
- Долна граница: 69.59

С 95% сигурност можем да твърдим, че средното тегло на спортистите в популацията е между 69.59 и 77.91 кг.

## Задача 2

Изследвана е съобразителността на 40 студента при работа с компютри по определена психологична задача. Да се изчисли средното време, употребено от студентите за решаването и при следните резултати. Да се определят границите на 95% - овия доверителен интервал за средната аритметична. Времето е количествена променлива.

| Употребено време (мин) | Брой студенти |
|------------------------|---------------|
| 3                      | 1             |
| 4                      | 3             |
| 5                      | 4             |
| 6                      | 7             |
| 7                      | 8             |
| 8                      | 9             |
| 9                      | 8             |
| общо                   | 40            |

## Решение

### Стъпка 1 Изчисляване на средното време

Използваме формулата за средна аритметична при групирани данни в степенен ред:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Таблично:

| Употребено време (мин) | Брой студенти | $x_i f_i$ |
|------------------------|---------------|-----------|
| 3                      | 1             | 3         |
| 4                      | 3             | 12        |
| 5                      | 4             | 20        |
| 6                      | 7             | 42        |
| 7                      | 8             | 56        |
| 8                      | 9             | 72        |
| 9                      | 8             | 72        |

$$\bar{x} = \frac{3 + 12 + 20 + 42 + 56 + 72 + 72}{40} = \frac{277}{40} = 7$$

## Стъпка 2 Изчисляване на стандартното отклонение

Използваме формулата за стандартно отклонение при групирани данни в степенен ред:

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$$

Таблично:

| Употребено време<br>(мин) | Брой<br>студенти | $x_i f_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $(x_i - \bar{x})^2 f_i$ |
|---------------------------|------------------|-----------|-------------------|---------------------|-------------------------|
| 3                         | 1                | 3         | -4                | 16                  | 16                      |
| 4                         | 3                | 12        | -3                | 9                   | 27                      |
| 5                         | 4                | 20        | -2                | 4                   | 16                      |
| 6                         | 7                | 42        | -1                | 1                   | 7                       |
| 7                         | 8                | 56        | 0                 | 0                   | 0                       |
| 8                         | 9                | 72        | 1                 | 1                   | 9                       |
| 9                         | 8                | 72        | 2                 | 4                   | 32                      |

$$SD = \sqrt{\frac{16 + 27 + 16 + 7 + 0 + 9 + 32}{40}} = \sqrt{\frac{107}{40}} = \sqrt{2.68} = 2.7$$

## Стъпка 3 Изчисляване на средната грешка SE

Използваме формулата за средна грешка

$$SE_m = \frac{SD}{\sqrt{n}}$$

$$SE_m = \frac{2.7}{\sqrt{40}} = 0.43$$

## Стъпка 4 Определяме стойността на гаранционния множител z

За 95% интервал  $z = 1.96$  (от листа с формулите)

### Стъпка 5 Определяме границите на интервала на доверителност

$$\bar{x} \pm z * SE_m$$

$$7 \pm 1.96 * 0.43 = 7 \pm 0.84$$

- Горна граница: 7.84
- Долна граница: 6.16

С 95% сигурност можем да твърдим, че средното време за решаване на задачата е между 6.16 и 7.84 минути.

### Задача 3

Дадена е променливата "диагноза" с нейните разновидности. Преобразувайте категорийната променлива в дихотомна и изчислете границите на 95%-овия интервал на доверителност за относителния дял на пациенти с **хепатит** от популацията.

| Диагноза  | Брой пациенти |
|-----------|---------------|
| Колит     | 10            |
| Гастрит   | 18            |
| Апендицит | 22            |
| Язва      | 29            |
| Хепатит   | 38            |
| Общо      | 117           |

### Решение

#### Стъпка 1 Преобразуване на категорийната променлива в дихотомна

| Диагноза | Брой пациенти |
|----------|---------------|
| Хепатит  | 38            |
| Друга    | 79            |
| Общо     | 117           |

#### Стъпка 2 Изчисляване на относителния дял

$$p = \frac{38}{117} = 0.3247 = 32.48$$

### Стъпка 3 Изчисляване на стандартната грешка

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

$$SE_p = \sqrt{\frac{32.48 \cdot (100 - 32.48)}{117}}$$

$$SE_p = \sqrt{\frac{32.48 \cdot 67.52}{117}} =$$

$$SE_p = \sqrt{\frac{2195.05}{117}} = \sqrt{18.75} = 4.33$$

### Стъпка 4 Определяме стойността на гаранционния множител z

За 95% интервал  $z = 1.96$  (от листа с формулите)

### Стъпка 5 Определяме границите на интервала на доверителност

$$p \pm z * SE_p$$

$$32.48 \pm 1.96 * 4.33$$

$$32.48 \pm 8.48$$

- Горна граница: 40.96%
- Долна граница: 24%

С 95% сигурност можем да твърдим, че относителният дял на пациенти с хепатит от популацията е между 24% и 40.96%.

## Задача 4

По случаен принцип са подбрани две извадки с обем съответно  $n = 320$  и  $n = 540$  индивиди с белодробни проблеми. Изследваните случаи от първата извадка са третирани с антибактериален препарат "А" без витамини, докато в терапията на втората извадка са включени и витамини. Случаите с подобрение в първата група са 98, а във втората група 231. Да се изчислят границите на 95% в доверителен интервал за относителния дял на случаите с подобрение в двете извадки.

| Терапия        | Всички пациенти | Случаи с подобрение |
|----------------|-----------------|---------------------|
| А без витамини | 320             | 98                  |
| А с витамини   | 540             | 231                 |

## Решение

### Стъпка 1 Изчисляване на относителния дял

за терапия без витамини

$$p = \frac{98}{320} = 0.3062 = 30.63$$

за терапия с витамини

$$p = \frac{231}{540} = 0.4277 = 42.78$$

### Стъпка 2 Изчисляване на стандартната грешка

за терапия без витамини

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

$$SE_p = \sqrt{\frac{30.63 \cdot (100 - 30.63)}{320}} =$$

$$SE_p = \sqrt{\frac{30.63 \cdot 69.37}{320}} =$$

$$SE_p = \sqrt{\frac{2124.8}{320}} = \sqrt{6.64} = 2.6$$

за терапия с витамини

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

$$SE_p = \sqrt{\frac{42.78 \cdot (100 - 42.78)}{540}} =$$

$$SE_p = \sqrt{\frac{42.78 \cdot 57.22}{540}} =$$

$$SE_p = \sqrt{\frac{2447.9}{540}} = \sqrt{4.53} = 2.1$$

### Стъпка 3 Определяме стойността на гаранционния множител z

За 95% интервал z = 1.96 (от листа с формулите)

### Стъпка 4 Определяме границите на интервала на доверителност

за терапия без витамини

$$p \pm z * SE_p$$

$$30.63 \pm 1.96 * 2.6$$

$$30.63 \pm 5.1$$

- Горна граница: 35.73%
- Долна граница: 25.53%

за терапия с витамини

$$p \pm z * SE_p$$

$$42.78 \pm 1.96 * 2.1$$

$$42.78 \pm 4.1$$

- Горна граница: 46.88%
- Долна граница: 38.68%

С 95% сигурност можем да твърдим, че относителният дял на случаите с подобрене в терапия без витамини е между 25.53% и 35.73%, а в терапия с витамини е между 38.68% и 46.88%.

## Упражнение 3 Тест на хипотеза. Т-тест.

### Задача 1

По случаен начин са подбрани две извадки с обем съответно  $n_1 = 248$  и  $n_2 = 140$  индивиди с кардиологични проблеми. В таблицата по-долу е регистриран пулсът на пациентите. Изчислете границите на 95% доверителен интервал за средните величини на популацията за двете извадки. Докажете дали средната аритметична е статистически различна спрямо групата пациенти, използвайки Т-тест.

| Група 1 - пулс | Група 1 - брой | Група 2 - пулс | Група 2 - брой |
|----------------|----------------|----------------|----------------|
| 60             | 4              | 80             | 10             |
| 65             | 18             | 85             | 15             |
| 70             | 25             | 90             | 25             |
| 75             | 30             | 95             | 30             |
| 80             | 50             | 100            | 28             |
| 85             | 48             | 105            | 16             |
| 90             | 36             | 120            | 2              |
| 95             | 28             | 125            | 8              |
| 100            | 9              | 130            | 6              |
| Общо           | 248            | Общо           | 140            |

## Решение

Дефинираме нулевата хипотеза  $H_0: \mu_1 = \mu_2$  и алтернативната хипотеза  $H_1: \mu_1 \neq \mu_2$

Избираме ниво на значимост  $\alpha = 0.05$

### Стъпка 1 Изчисляване на средните величини на популацията за двете извадки

Използваме формулата за средна аритметична при групирани данни в степенен рег:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Таблично за група 1:

| Пулс | Брой пациенти | $x_i f_i$ |
|------|---------------|-----------|
| 60   | 4             | 240       |
| 65   | 18            | 1170      |
| 70   | 25            | 1750      |
| 75   | 30            | 2250      |
| 80   | 50            | 4000      |
| 85   | 48            | 4080      |
| 90   | 36            | 3240      |
| 95   | 28            | 2660      |
| 100  | 9             | 900       |
| Общо | 248           |           |

$$\bar{x} = \frac{240 + 1170 + 1750 + 2250 + 4000 + 4080 + 3240 + 2660 + 900}{248} = \frac{20290}{248} = 81.8$$

Таблично за група 2:

| Пулс | Брой пациенти | $x_i f_i$ |
|------|---------------|-----------|
| 80   | 10            | 800       |
| 85   | 15            | 1275      |
| 90   | 25            | 2250      |
| 95   | 30            | 2850      |
| 100  | 28            | 2800      |
| 105  | 16            | 1680      |
| 120  | 2             | 240       |
| 125  | 8             | 1000      |
| 130  | 6             | 780       |
| Общо | 140           |           |

$$\bar{x} = \frac{800 + 1275 + 2250 + 2850 + 2800 + 1680 + 240 + 1000 + 780}{140} = \frac{13675}{140} = 97.68$$

## Стъпка 2 Изчисляване на стандартното отклонение

Използваме формулата за стандартно отклонение при групирани данни в степенен рег:

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$$

Таблично за група 1:

| Пулс   | Брой пациенти | $x_i f_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $(x_i - \bar{x})^2 f_i$ |
|--------|---------------|-----------|-------------------|---------------------|-------------------------|
| 60,00  | 4,00          | 240       | -21,80            | 475,24              | 1900,96                 |
| 65,00  | 18,00         | 1170      | -16,80            | 282,24              | 5080,32                 |
| 70,00  | 25,00         | 1750      | -11,80            | 139,24              | 3481                    |
| 75,00  | 30,00         | 2250      | -6,80             | 46,24               | 1387,2                  |
| 80,00  | 50,00         | 4000      | -1,80             | 3,24                | 162                     |
| 85,00  | 48,00         | 4080      | 3,20              | 10,24               | 491,52                  |
| 90,00  | 36,00         | 3240      | 8,20              | 67,24               | 2420,64                 |
| 95,00  | 28,00         | 2660      | 13,20             | 174,24              | 4878,72                 |
| 100,00 | 9,00          | 900       | 18,20             | 331,24              | 2981,16                 |

$$SD = \sqrt{\frac{1900.96 + 5080.32 + 3481 + 1387.2 + 162 + 491.52 + 2420.64 + 4878.72 + 2981.16}{248}}$$

$$SD = \sqrt{\frac{22783.52}{248}} = \sqrt{91.87} = 9.6$$

Таблично за група 2:

| Пулс   | Брой<br>пациенти | $x_i f_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $(x_i - \bar{x})^2 f_i$ |
|--------|------------------|-----------|-------------------|---------------------|-------------------------|
| 80,00  | 10,00            | 800       | -17,68            | 312,58              | 3125,8                  |
| 85,00  | 15,00            | 1275      | -12,68            | 160,78              | 2411,7                  |
| 90,00  | 25,00            | 2250      | -7,68             | 58,98               | 1474,5                  |
| 95,00  | 30,00            | 2850      | -2,68             | 7,18                | 215,4                   |
| 100,00 | 28,00            | 2800      | 2,32              | 5,38                | 150,64                  |
| 105,00 | 16,00            | 1680      | 7,32              | 53,58               | 857,28                  |
| 120,00 | 2,00             | 240       | 22,32             | 498,18              | 996,36                  |
| 125,00 | 8,00             | 1000      | 27,32             | 746,38              | 5971,04                 |
| 130,00 | 6,00             | 780       | 32,32             | 1044,58             | 6267,48                 |

$$SD = \sqrt{\frac{3125.8 + 2411.7 + 1474.5 + 215.4 + 150.64 + 857.28 + 996.36 + 5971.04 + 6267.48}{140}}$$

$$SD = \sqrt{\frac{21470,2}{140}} = \sqrt{153.36} = 12.38$$

### Стъпка 3 Изчисляване на стандартната грешка

Използваме формулата за средна грешка

$$SE_m = \frac{SD}{\sqrt{n}}$$

За група 1:

$$SE_m = \frac{9.6}{\sqrt{248}} = \frac{9.6}{15.75} = 0.6$$

За група 2:

$$SE_m = \frac{12.38}{\sqrt{140}} = \frac{12.38}{11.83} = 1.05$$

#### Стъпка 4 Изчисляване на стойността на t

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{SE_{m_1}^2 + SE_{m_2}^2}}$$

$$t = \frac{|81.8 - 97.68|}{\sqrt{0.6^2 + 1.05^2}} = \frac{15.88}{\sqrt{0.36 + 1.1}} = \frac{15.88}{\sqrt{1.46}} = \frac{15.88}{1.2} = 13.23$$

#### Стъпка 5 Определяне стойността на p (възможността нулевата хипотеза да е вярна)

Поглеждаме таблицата с t - разпределението.

| df | p = 0.1   | <b>p = 0.05</b>  | p = 0.01  |
|----|-----------|------------------|-----------|
| 30 | t = 1.645 | <b>t = 1.960</b> | t = 2.576 |

Стойността t = 13.23 е надясно (по-голяма от) 1,96. Това означава, че вероятността p е по-малка от 0,05.

#### Стъпка 7 Приемане или отхвърляне на нулевата хипотеза

Тъй като  $p < \alpha$ , тоест  $p < 0.05$ , то нулевата хипотеза се **отхвърля**. Има статистически значима разлика в средните стойности на пулса на пациентите от двете групи.

## Задача 2

Да се определи съществено ли е различието във виталния капацитет при индивиди от мъжки пол на 19 годишна възраст в две независими извадки с еднакъв брой случаи. В едната извадка са включени баскетболисти, а в другата младежи, трениращи плувен спорт. Да се приложи t тестът, като необходимите статистически параметри – средна аритметична и стандартно отклонение да се изчислят както при групирани данни.

| <b>Баскетболисти</b> |             | <b>Плувци</b>     |             |
|----------------------|-------------|-------------------|-------------|
| Витален капацитет    | Честота (f) | Витален капацитет | Честота (f) |
| 2601-2800            | 12          | 3401-3600         | 13          |
| 2801-3000            | 21          | 3601-3800         | 26          |
| 3001-3200            | 35          | 3801-4000         | 35          |
| 3201-3400            | 62          | 4001-4200         | 58          |
| 3401-3600            | 33          | 4201-4400         | 30          |
| 3601-3800            | 19          | 4401-4600         | 20          |
| 3801-4000            | 15          | 4601-4800         | 15          |
| общо:                | 197         | общо:             | 197         |

## Решение

Дефинираме нулевата хипотеза  $H_0: \mu_1 = \mu_2$  и алтернативната хипотеза  $H_1: \mu_1 \neq \mu_2$

Избираме ниво на значимост  $\alpha = 0.05$

### Стъпка 1 Изчисляване на средните величини на популацията за двете извадки

Използваме формулата за средна аритметична при групирани данни в интервален рег:

$$\bar{x} = \frac{\sum x_u \cdot f}{\sum f}$$

Таблично за баскетболисти:

| Витален капацитет | Честота | $x_u$   | $x_u f$   |
|-------------------|---------|---------|-----------|
| 2601-2800         | 12,00   | 2700,50 | 32406,00  |
| 2801-3000         | 21,00   | 2900,50 | 60910,50  |
| 3001-3200         | 35,00   | 3100,50 | 108517,50 |
| 3201-3400         | 62,00   | 3300,50 | 204631,00 |
| 3401-3600         | 33,00   | 3500,50 | 115516,50 |
| 3601-3800         | 19,00   | 3700,50 | 70309,50  |
| 3801-4000         | 15,00   | 3900,50 | 58507,50  |

$$\bar{x} = \frac{32406 + 60910.50 + 108517.50 + 204631 + 115516.50 + 70309.50 + 58507.50}{197} =$$

$$\frac{650798.5}{197} = 3303.5$$

Таблично за плувци:

| Витален капацитет | Честота | $x_u$   | $x_u f$   |
|-------------------|---------|---------|-----------|
| 3401-3600         | 13,00   | 3500,50 | 45506,50  |
| 3601-3800         | 26,00   | 3700,50 | 96213,00  |
| 3801-4000         | 35,00   | 3900,50 | 136517,50 |
| 4001-4200         | 58,00   | 4100,50 | 237829,00 |
| 4201-4400         | 30,00   | 4300,50 | 129015,00 |
| 4401-4600         | 20,00   | 4500,50 | 90010,00  |
| 4601-4800         | 15,00   | 4700,50 | 70507,50  |

$$\bar{x} = \frac{45506.50 + 96213 + 136517.50 + 237829 + 129015 + 90010 + 70507.50}{197} =$$

$$\frac{805598.5}{197} = 4089.3$$

## Стъпка 2 Изчисляване на стандартното отклонение

Използваме формулата за стандартно отклонение при групирани данни в интервален рег:

$$SD = \sqrt{\frac{\sum (x_u - \bar{x})^2 \cdot f}{\sum f}}$$

Таблично за баскетболисти:

| Витален<br>капацитет | Честота | $x_u$  | $x_u f$ | $(x_u - \bar{x})$ | $(x_u - \bar{x})^2$ | $(x_u - \bar{x})^2 f$ |
|----------------------|---------|--------|---------|-------------------|---------------------|-----------------------|
| 2601-2800            | 12      | 2700,5 | 32406   | -603              | 363609              | 4363308               |
| 2801-3000            | 21      | 2900,5 | 60921   | -403              | 162409              | 3410589               |
| 3001-3200            | 35      | 3100,5 | 108516  | -203              | 41209               | 1442315               |
| 3201-3400            | 62      | 3300,5 | 204671  | -3                | 9                   | 558                   |
| 3401-3600            | 33      | 3500,5 | 115516  | 197               | 38809               | 1280697               |
| 3601-3800            | 19      | 3700,5 | 70395   | 397               | 157609              | 2994571               |
| 3801-4000            | 15      | 3900,5 | 58507   | 597               | 356409              | 5346135               |

$$SD = \sqrt{\frac{4363308 + 3410589 + 1442315 + 558 + 1280697 + 2994571 + 5346135}{197}} =$$

$$SD = \sqrt{\frac{18838173}{197}} = \sqrt{95625,24} = 309,23$$

Таблично за плувци:

| Витален<br>капацитет | Честота | $x_u$  | $x_u f$  | $(x_u - \bar{x})$ | $(x_u - \bar{x})^2$ | $(x_u - \bar{x})^2 f$ |
|----------------------|---------|--------|----------|-------------------|---------------------|-----------------------|
| 3401-3600            | 13      | 3500,5 | 45506,5  | -588,8            | 346685,44           | 4506910,72            |
| 3601-3800            | 26      | 3700,5 | 96213    | -388,8            | 151165,44           | 3930301,44            |
| 3801-4000            | 35      | 3900,5 | 136517,5 | -188,8            | 35645,44            | 1247590,4             |
| 4001-4200            | 58      | 4100,5 | 237829   | 11,2              | 125,44              | 7275,52               |
| 4201-4400            | 30      | 4300,5 | 129015   | 211,2             | 44605,44            | 1338163,2             |
| 4401-4600            | 20      | 4500,5 | 90010    | 411,2             | 169085,44           | 3381708,8             |
| 4601-4800            | 15      | 4700,5 | 70507,5  | 611,2             | 373565,44           | 5603481,6             |

$$SD = \sqrt{\frac{4506910,72 + 3930301,44 + 1247590,4 + 7275,52 + 1338163,2 + 3381708,8 + 5603481,6}{197}} =$$

$$SD = \sqrt{\frac{20015431,68}{197}} = \sqrt{101601,18} = 318.75$$

### Стъпка 3 Изчисляване на стандартната грешка

Използваме формулата за средна грешка

$$SE_m = \frac{SD}{\sqrt{n}}$$

За баскетболисти:

$$SE_m = \frac{309.23}{\sqrt{197}} = \frac{309.23}{14.04} = 22.02$$

За плувци:

$$SE_m = \frac{318.75}{\sqrt{197}} = \frac{318.75}{14.04} = 22.69$$

### Стъпка 4 Изчисляване на стойността на t

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{SE_{m_1}^2 + SE_{m_2}^2}}$$

$$t = \frac{|3303.5 - 4089.3|}{\sqrt{22.02^2 + 22.69^2}}$$

$$t = \frac{785.8}{\sqrt{484.9 + 514.8}}$$

$$t = \frac{785.8}{\sqrt{999.7}} = \frac{785.8}{31.61} = 24.9$$

### Стъпка 5 Определяне стойността на $p$ (възможността нулевата хипотеза да е вярна)

Поглеждаме таблицата с  $t$  - разпределението.

| df | $p = 0.1$   | <b><math>p = 0.05</math></b>  | $p = 0.01$  |
|----|-------------|-------------------------------|-------------|
| 30 | $t = 1.645$ | <b><math>t = 1.960</math></b> | $t = 2.576$ |

Стойността  $t = 24.9$  е надясно (по-голяма от) 1,96. Това означава, че вероятността  $p$  е по-малка от 0,05.

### Стъпка 7 Приемане или отхвърляне на нулевата хипотеза

Тъй като  $p < \alpha$ , тоест  $p < 0.05$ , то нулевата хипотеза се **отхвърля**. Има статистически значима разлика в средните стойности на виталния капацитет на спортистите от двете групи.

## Упражнение 4 & 5 Изследване на връзки. Корелация. Асоциация

### Задача 1

Като използвате **непараметричен тест хи квадрат** проверете съществува ли асоциация между продължителността трудовия стаж при миньорите и заболяемостта от вибрационна болест?

| трудова стаж | с вибрационна болест | здрав | Общо |
|--------------|----------------------|-------|------|
| до 5 г.      | 32                   | 127   | 159  |
| 6-15 г.      | 56                   | 268   | 324  |
| 16-25 г.     | 23                   | 48    | 71   |
| над 25 г.    | 41                   | 34    | 75   |
| Общо         | 152                  | 477   | 629  |

### Решение

Дефинираме нулевата хипотеза  $H_0$ : няма асоциация между продължителността трудовия стаж при миньорите и заболяемостта от вибрационна болест и алтернативната хипотеза  $H_1$ : има асоциация между продължителността трудовия стаж при миньорите и заболяемостта от вибрационна болест.

Избираме ниво на значимост  $\alpha = 0.05$

| трудоѝ стаж | с вибрационна болест | здрави  | Общо |
|-------------|----------------------|---------|------|
| до 5 г.     | (a) 32               | (e) 127 | 159  |
| 6-15 г.     | (b) 56               | (f) 268 | 324  |
| 16-25 г.    | (c) 23               | (g) 48  | 71   |
| над 25 г.   | (d) 41               | (h) 34  | 75   |
| Общо        | 152                  | 477     | 629  |

### Стъпка 1 Изчисляване на очакваните честоти

Използваме формулата за очаквана честота:

$$f_{tcell} = \frac{\sum_{columns} \cdot \sum_{rows}}{\sum_{total}}$$

за клетка (a):

$$f_{ta} = \frac{159 \cdot 152}{629} = 38$$

за клетка (b):

$$f_{tb} = \frac{324 \cdot 152}{629} = 78$$

за клетка (c):

$$f_{tc} = \frac{71 \cdot 152}{629} = 17$$

за клетка (d):

$$f_{td} = \frac{75 \cdot 152}{629} = 18$$

за клетка (e):

$$f_{te} = \frac{159 \cdot 477}{629} = 121$$

за клетка (f):

$$f_t f = \frac{324 \cdot 477}{629} = 247$$

за клетка (g):

$$f_t g = \frac{71 \cdot 477}{629} = 57$$

за клетка (h):

$$f_t h = \frac{75 \cdot 477}{629} = 58$$

## Стъпка 2 Изчисляване на статистиката на хи квадрат

Използваме формулата за статистиката на хи квадрат:

$$\chi^2 = \sum \frac{(f_o - f_t)^2}{f_t}$$

за клетка (a):

$$\chi^2 = \frac{(32 - 38)^2}{38} = 0.95$$

за клетка (b):

$$\chi^2 = \frac{(56 - 78)^2}{78} = 6.08$$

за клетка (c):

$$\chi^2 = \frac{(23 - 17)^2}{17} = 2.35$$

за клетка (d):

$$\chi^2 = \frac{(41 - 18)^2}{18} = 10.78$$

за клетка (e):

$$\chi^2 = \frac{(127 - 121)^2}{121} = 0.25$$

за клетка (f):

$$\chi^2 = \frac{(268 - 247)^2}{247} = 1.71$$

за клетка (g):

$$\chi^2 = \frac{(48 - 57)^2}{57} = 1.26$$

за клетка (h):

$$\chi^2 = \frac{(34 - 58)^2}{58} = 10.45$$

Общо:

$$\chi^2 = 0.95 + 6.08 + 2.35 + 10.78 + 0.25 + 1.71 + 1.26 + 10.45 = 33.83$$

### Стъпка 3 Определяне степените на свобода

Използваме формулата за степените на свобода:

$$df = (r - 1)(c - 1)$$

където  $r$  е броят на редовете, а  $c$  е броят на колоните.

$$df = (4 - 1)(2 - 1) = 3$$

### Стъпка 4 Определяне стойността на $p$ (възможността нулевата хипотеза да е вярна)

Поглеждаме таблицата с хи квадрат разпределението.

| df | p = 0.1          | p = 0.05         | p = 0.01          |
|----|------------------|------------------|-------------------|
| 3  | $\chi^2 = 6.251$ | $\chi^2 = 7.815$ | $\chi^2 = 11.345$ |

Стойността  $\chi^2 = 33.83$  е надясно (по-голяма от) 7,815. Това означава, че вероятността  $p$  е по-малка от 0,05.

### Стъпка 5 Приемане или отхвърляне на нулевата хипотеза

Тъй като  $p < \alpha$ , тоест  $p < 0.05$ , то нулевата хипотеза се **отхвърля**. Има статистически значима асоциация между продължителността трудовия стаж при миньорите и заболяемостта от вибрационна болест.

## Задача 2

При проучване на влиянието на вида използван дюшек при получаване на декубитусни рани, са получени следните резултати:

- При *low air-loss* дюшек при **7 от 56** изследвани случаи имат декубитусни рани
- При *воден дюшек* при **13 от 71** изследвани лица
- При дюшек *на вълни* с декубитусни рани са **25 от 72** пациенти
- Определете дали видът на използвания дюшек е фактор за появата на декубитусни рани?

## Решение

Дефинираме нулевата хипотеза  $H_0$ : видът на използвания дюшек не е фактор за появата на декубитусни рани и алтернативната хипотеза  $H_1$ : видът на използвания дюшек е фактор за появата на декубитусни рани.

Избираме ниво на значимост  $\alpha = 0.05$

| Вид дюшек | с декубитусни рани | без декубитусни рани | Общо |
|-----------|--------------------|----------------------|------|
| low air   | (a) 7              | (d) 49               | 56   |
| воден     | (b) 13             | (e) 58               | 71   |
| на вълни  | (c) 25             | (f) 47               | 72   |
| Общо      | 45                 | 154                  | 199  |

### Стъпка 1 Изчисляване на очакваните честоти

Използваме формулата за очаквана честота:

$$f_{tcell} = \frac{\sum_{columns} \cdot \sum_{rows}}{\sum_{total}}$$

за клетка (a):

$$f_{ta} = \frac{56 \cdot 45}{199} = 13$$

за клетка (b):

$$f_{tb} = \frac{71 \cdot 45}{199} = 16$$

за клетка (c):

$$f_{tc} = \frac{72 \cdot 45}{199} = 16$$

за клетка (d):

$$f_{td} = \frac{56 \cdot 154}{199} = 43$$

за клетка (e):

$$f_{te} = \frac{71 \cdot 154}{199} = 55$$

за клетка (f):

$$f_{tf} = \frac{72 \cdot 154}{199} = 58$$

Таблично:

| вид гъшек | с декубитусни рани | без декубитусни рани | Общо |
|-----------|--------------------|----------------------|------|
| low air   | (a) 13             | (d) 43               | 56   |

| Вид гъшек | с декубитусни рани | без декубитусни рани | Общо |
|-----------|--------------------|----------------------|------|
| воден     | (b) 16             | (e) 55               | 71   |
| на вълни  | (c) 16             | (f) 58               | 72   |
| Общо      | 45                 | 154                  | 199  |

## Стъпка 2 Изчисляване на статистиката на хи квадрат

Използваме формулата за статистиката на хи квадрат:

$$\chi^2 = \sum \frac{(f_o - f_t)^2}{f_t}$$

за клетка (a):

$$\chi^2 = \frac{(7 - 13)^2}{13} = 2.77$$

за клетка (b):

$$\chi^2 = \frac{(13 - 16)^2}{16} = 0.56$$

за клетка (c):

$$\chi^2 = \frac{(25 - 16)^2}{16} = 3.06$$

за клетка (d):

$$\chi^2 = \frac{(49 - 43)^2}{43} = 0.84$$

за клетка (e):

$$\chi^2 = \frac{(58 - 55)^2}{55} = 0.16$$

за клетка (f):

$$\chi^2 = \frac{(47 - 58)^2}{56} = 1.62$$

Общо:

$$\chi^2 = 2.77 + 0.56 + 3.06 + 0.84 + 0.16 + 1.62 = 8.01$$

### Стъпка 3 Определяне степените на свобода

Използваме формулата за степените на свобода:

$$df = (r - 1)(c - 1)$$

където  $r$  е броят на редовете, а  $c$  е броят на колоните.

$$df = (3 - 1)(2 - 1) = 2$$

### Стъпка 4 Определяне стойността на $p$ (възможността нулевата хипотеза да е вярна)

Поглеждаме таблицата с хи квадрат разпределението.

| df | $p = 0.1$        | <b><math>p = 0.05</math></b> | $p = 0.01$       |
|----|------------------|------------------------------|------------------|
| 2  | $\chi^2 = 4.605$ | $\chi^2 = 5.991$             | $\chi^2 = 9.210$ |

Стойността  $\chi^2 = 8.01$  е надясно (по-голяма от) 5,991. Това означава, че вероятността  $p$  е по-малка от 0,05.

### Стъпка 5 Приемане или отхвърляне на нулевата хипотеза

Тъй като  $p < \alpha$ , тоест  $p < 0.05$ , то нулевата хипотеза се **отхвърля**. Видът на използвания глؤшек е фактор за появата на декубитосни рани.

### Задача 3

При проучване на нов анестетик е отчетено времето от подаването му във венозната система до реакцията на пациента и съответната ефективна доза, определена в  $\mu\text{g/kg}$ .

| доза       | време за въздействие |
|------------|----------------------|
| 1,3        | 25                   |
| 1,9        | 13                   |
| 1,5        | 19                   |
| 1,1        | 23                   |
| 1          | 26                   |
| 1,3        | 21                   |
| 2,1        | 11                   |
| 1,7        | 30                   |
| 1,6        | 14                   |
| 0,9        | 24                   |
| общо: 14,4 | общо: 206            |

#### Първо:

Изчислете 95% -вия интервал на доверителност на ефективна доза в генералната съвкупност.

#### Решение

##### Определяме средната аритметична

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{14.4}{10} = 1.44$$

##### Определяме стандартното отклонение

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

| гоза           | $x - \bar{x}$ | $(x - \bar{x})^2$           |
|----------------|---------------|-----------------------------|
| 1,3            | -0,14         | 0,02                        |
| 1,9            | 0,46          | 0,21                        |
| 1,5            | 0,06          | 0,01                        |
| 1,1            | -0,34         | 0,12                        |
| 1              | -0,44         | 0,19                        |
| 1,3            | -0,14         | 0,02                        |
| 2,1            | 0,66          | 0,44                        |
| 1,7            | 0,26          | 0,07                        |
| 1,6            | 0,16          | 0,03                        |
| 0,9            | -0,54         | 0,29                        |
| $\bar{x}$ 1,44 |               | $\sum (x - \bar{x})^2$ 1,39 |

$$SD = \sqrt{\frac{1.39}{10}} = 0.39$$

**Определяме стандартната грешка**

$$SE_m = \frac{SD}{\sqrt{n}}$$

$$SE_m = \frac{0.39}{\sqrt{10}} = 0.12$$

**Определяме стойността на гаранционния множител z**

За 95% интервал  $z = 1.96$  (от листа с формулите)

**Определяме границите на интервала на доверителност**

$$\bar{x} \pm z * SE_m$$

$$1.44 \pm 1.96 * 0.12$$

$$1.44 \pm 0.24$$

- Долна граница на интервала на доверителност: 1.2
- Горна граница на интервала на доверителност: 1.68

С 95% вероятност ефективната доза в генералната съвкупност е между 1.2 и 1.68.

### Второ:

Изчислете 95% -вия интервал на доверителност за времето до ефект в генералната съвкупност.

### Решение

#### Определяме средната аритметична

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{206}{10} = 20.6$$

#### Определяме стандартното отклонение

$$SD = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}$$

| време          | $y - \bar{y}$ | $(y - \bar{y})^2$            |
|----------------|---------------|------------------------------|
| 25             | 4,4           | 19,36                        |
| 13             | -7,6          | 57,76                        |
| 19             | -1,6          | 2,56                         |
| 23             | 2,4           | 5,76                         |
| 26             | 5,4           | 29,16                        |
| 21             | 0,4           | 0,16                         |
| 11             | -9,6          | 92,16                        |
| 30             | 9,4           | 88,36                        |
| 14             | -6,6          | 43,56                        |
| 24             | 3,4           | 11,56                        |
| $\bar{y}$ 20,6 |               | $\sum (y - \bar{y})^2$ 350,4 |

$$SD = \sqrt{\frac{350.4}{10}} = \sqrt{35.04} = 5.92$$

**Определяме стандартната грешка**

$$SE_m = \frac{SD}{\sqrt{n}}$$

$$SE_m = \frac{5.92}{\sqrt{10}} = 1.87$$

**Определяме стойността на гаранционния множител z**

За 95% интервал  $z = 1.96$  (от листа с формулите)

**Определяме границите на интервала на доверителност**

$$\bar{x} \pm z * SE_m$$

$$20.6 \pm 1.96 * 1.87$$

$$20.6 \pm 3.67$$

- Долна граница на интервала на доверителност: 16.93
- Горна граница на интервала на доверителност: 24.27

С 95% вероятност времето до ефект в генералната съвкупност е между 16.93 и 24.27.

**Трето:**

Докажете дали съществува връзка между явленията **доза** и **време за реакция**.

## Решение

Прилагаме формулата за корелационен коефициент на Пийрсън:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

За целта:

**Изчисляваме средната аритметична за гозата**

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{14.4}{10} = 1.44$$

**Изчисляваме средната аритметична за времето за реакция**

$$\bar{y} = \frac{\sum y}{n}$$

$$\bar{y} = \frac{206}{10} = 20.6$$

**Изчисляваме отклоненията  $dx$  и  $dy$**

$$dx = x - \bar{x}$$

$$dy = y - \bar{y}$$

Таблично:

| гоза | време за въздействие | dx    | dy    | dx*dy |
|------|----------------------|-------|-------|-------|
| 1,30 | 25,0                 | -0,14 | 4,40  | -0,62 |
| 1,90 | 13,0                 | 0,46  | -7,60 | -3,50 |
| 1,50 | 19,0                 | 0,06  | -1,60 | -0,10 |
| 1,10 | 23,0                 | -0,34 | 2,40  | -0,82 |
| 1,00 | 26,0                 | -0,44 | 5,40  | -2,38 |
| 1,30 | 21,0                 | -0,14 | 0,40  | -0,06 |
| 2,10 | 11,0                 | 0,66  | -9,60 | -6,34 |
| 1,70 | 30,0                 | 0,26  | 9,40  | 2,44  |
| 1,60 | 14,0                 | 0,16  | -6,60 | -1,06 |
| 0,90 | 24,0                 | -0,54 | 3,40  | -1,84 |

**Изчисляваме сумата на dx\*dy**

$$\sum dx * dy = (-0.62) + (-3.50) + (-0.10) + (-0.82) + (-2.38) + (-0.06) + (-6.34) + 2.44 + (-1.06) + (-1.84) = -14.28$$

**Изчисляваме сумата на dx^2 и dy^2**

Таблично:

| гоза | време за въздействие | dx    | dy    | dx^2 | dy^2  |
|------|----------------------|-------|-------|------|-------|
| 1,30 | 25,00                | -0,14 | 4,40  | 0,02 | 19,36 |
| 1,90 | 13,00                | 0,46  | -7,60 | 0,21 | 57,76 |
| 1,50 | 19,00                | 0,06  | -1,60 | 0,00 | 2,56  |
| 1,10 | 23,00                | -0,34 | 2,40  | 0,12 | 5,76  |
| 1,00 | 26,00                | -0,44 | 5,40  | 0,19 | 29,16 |
| 1,30 | 21,00                | -0,14 | 0,40  | 0,02 | 0,16  |
| 2,10 | 11,00                | 0,66  | -9,60 | 0,44 | 92,16 |
| 1,70 | 30,00                | 0,26  | 9,40  | 0,07 | 88,36 |
| 1,60 | 14,00                | 0,16  | -6,60 | 0,03 | 43,56 |
| 0,90 | 24,00                | -0,54 | 3,40  | 0,29 | 11,56 |

$$\sum dx^2 = 0.02 + 0.21 + 0.00 + 0.12 + 0.19 + 0.02 + 0.44 + 0.07 + 0.03 + 0.29 = 1.39$$

$$\sum dy^2 = 19.36 + 57.76 + 2.56 + 5.76 + 29.16 + 0.16 + 92.16 + 88.36 + 43.56 + 11.56 = 350.2$$

**Изчисляваме корелационния коефициент на Пиърсън**

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$\rho = \frac{-14.28}{\sqrt{1.39} \sqrt{350.2}}$$

$$\rho = \frac{-14.28}{1.18 \cdot 18.71}$$

$$\rho = \frac{-14.28}{22.07}$$

$$\rho = -0.65$$

Корелационният коефициент на Пиърсън е -0.65. Това означава, че има **умерена обратна корелация** между гозата и времето за реакция.

## Упражнение 6 Регресия

### Задача 1

Данните регистрирани в таблицата по-долу са на пациенти, участници в проучване на коронарната болест на сърцето. Изградете регресионен модел с който да предскажете стойността на холестерола в кръвта на пациентите спрямо теглото им.

| Тегло (kg) (x) | Холестерол (mg/dl) (y) |
|----------------|------------------------|
| 84             | 135                    |
| 107            | 222                    |
| 94             | 244                    |
| 83             | 252                    |
| 131            | 269                    |
| 72             | 286                    |
| 87             | 294                    |
| 88             | 302                    |
| 77             | 311                    |
| 90             | 311                    |
| 79             | 312                    |
| 82             | 353                    |
| 70             | 403                    |
| 88             | 403                    |

### Решение

#### Изчисляване на регресионния коефициент b

$$\beta = \frac{(n \cdot \sum (X \cdot Y)) - (\sum X \cdot \sum Y)}{(n \cdot \sum X^2) - (\sum X)^2}$$

## Изчисляване на $\sum X, \sum Y, \sum XY$

Таблично:

| Тегло (kg)    | Холестерол (mg/dl) | $X*Y$            |
|---------------|--------------------|------------------|
| 84            | 135                | 11340            |
| 107           | 222                | 23754            |
| 94            | 244                | 22936            |
| 83            | 252                | 20916            |
| 131           | 269                | 35239            |
| 72            | 286                | 20592            |
| 87            | 294                | 25578            |
| 88            | 302                | 26576            |
| 77            | 311                | 23947            |
| 90            | 311                | 27990            |
| 79            | 312                | 24648            |
| 82            | 353                | 28946            |
| 70            | 403                | 28210            |
| 88            | 403                | 35464            |
| $\sum X$ 1232 | $\sum Y$ 4097      | $\sum XY$ 356136 |

$$\sum X = 84 + 107 + 94 + 83 + 131 + 72 + 87 + 88 + 77 + 90 + 79 + 82 + 70 + 88 = 1232$$

$$\sum Y = 135 + 222 + 244 + 252 + 269 + 286 + 294 + 302 + 311 + 311 + 312 + 353 + 403 + 403 = 4097$$

$$\sum XY = 11340 + 23754 + 22936 + 20916 + 35239 + 20592 + 25578 +$$

$$26576 + 23947 + 27990 + 24648 + 28946 + 28210 + 35564 = 356136$$

## Изчисляване на $\sum X^2$

Таблично:

| Тегло (kg)    | Холестерол (mg/dl) | $X^2$             |
|---------------|--------------------|-------------------|
| 84            | 135                | 7056              |
| 107           | 222                | 11449             |
| 94            | 244                | 8836              |
| 83            | 252                | 6889              |
| 131           | 269                | 17161             |
| 72            | 286                | 5184              |
| 87            | 294                | 7569              |
| 88            | 302                | 7744              |
| 77            | 311                | 5929              |
| 90            | 311                | 8100              |
| 79            | 312                | 6241              |
| 82            | 353                | 6724              |
| 70            | 403                | 4900              |
| 88            | 403                | 7744              |
| $\sum X$ 1232 | $\sum Y$ 4097      | $\sum X^2$ 111526 |

$$\begin{aligned}\sum X^2 &= 7056 + 11449 + 8836 + 6889 + 17161 + 5184 + 7569 + 7744 + 5929 + \\ &8100 + 6241 + 6724 + 4900 + 7744 = 111526\end{aligned}$$

## Изчисляване на регресионния коефициент $b$

$$\beta = \frac{(n \cdot \sum (X \cdot Y)) - (\sum X \cdot \sum Y)}{(n \cdot \sum X^2) - (\sum X)^2}$$

$$\beta = \frac{(14 \cdot 356136) - (1232 \cdot 4097)}{(14 \cdot 111526) - (1232)^2}$$

$$\beta = \frac{4985904 - 5047504}{1561364 - 1517824}$$

$$\beta = \frac{-61600}{43540}$$

$$\beta = -1.415$$

Всяка единица повишение на теглото води до намаляване на холестерола с 1.415 мг/дл.

**Изчисляване на константата a**

$$a = \bar{y} - (\beta \cdot \bar{x})$$

**Изчисляване на  $\bar{x}$**

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{1232}{14} = 88$$

**Изчисляване на  $\bar{y}$**

$$\bar{y} = \frac{\sum y}{n}$$

$$\bar{y} = \frac{4097}{14} = 292.64$$

**Изчисляване на константата a**

$$a = \bar{y} - (\beta \cdot \bar{x})$$

$$a = 292.64 - (-1.415 \cdot 88)$$

$$a = 417.16$$

## Изчисляване на регресионната уравнение

$$y = a + b \cdot x$$

$$y = 417.16 - 1.415 \cdot x$$

## Задача 2

При проучване на нов анестетик е отчетено времето от подаването на медикамента до реакцията на пациента и съответната ефективна доза, определена в мкг/кг. Данните са показани в таблицата по-долу. Изградете регресионен модел с който да предсказвате времето за реакция в зависимост от дозата на медикамента.

| Доза (мкг/кг) | Време (сек) |
|---------------|-------------|
| 1,3           | 25          |
| 1,9           | 13          |
| 1,5           | 19          |
| 1,1           | 23          |
| 1             | 26          |
| 1,3           | 21          |
| 2,1           | 11          |
| 1,7           | 30          |
| 1,6           | 14          |
| 0,9           | 24          |

## Решение

### Изчисляване на регресионния коефициент b

$$\beta = \frac{(n \cdot \sum (X \cdot Y)) - (\sum X \cdot \sum Y)}{(n \cdot \sum X^2) - (\sum X)^2}$$

### Изчисляване на $\sum X, \sum Y, \sum XY$

Таблично:

| Доза (мкг/кг) | Време (сек) | X*Y  |
|---------------|-------------|------|
| 1,3           | 25          | 32,5 |

| Доза (мкз/кг) | Време (сек)  | X*Y             |
|---------------|--------------|-----------------|
| 1,9           | 13           | 24,7            |
| 1,5           | 19           | 28,5            |
| 1,1           | 23           | 25,3            |
| 1             | 26           | 26              |
| 1,3           | 21           | 27,3            |
| 2,1           | 11           | 23,1            |
| 1,7           | 30           | 51              |
| 1,6           | 14           | 22,4            |
| 0,9           | 24           | 21,6            |
| $\sum X$ 14,4 | $\sum Y$ 206 | $\sum XY$ 282,4 |

$$\sum X = 1.3 + 1.9 + 1.5 + 1.1 + 1 + 1.3 + 2.1 + 1.7 + 1.6 + 0.9 = 14.4$$

$$\sum Y = 25 + 13 + 19 + 23 + 26 + 21 + 11 + 30 + 14 + 24 = 206$$

$$\sum XY = 32.5 + 24.7 + 28.5 + 25.3 + 26 + 27.3 + 23.1 + 51 + 22.4 + 21.6 = 282.4$$

**Изчисляване на  $\sum X^2$**

Таблично:

| Доза (мкз/кг) | Време (сек)  | X^2              |
|---------------|--------------|------------------|
| 1,3           | 25           | 1,69             |
| 1,9           | 13           | 3,61             |
| 1,5           | 19           | 2,25             |
| 1,1           | 23           | 1,21             |
| 1,0           | 26           | 1                |
| 1,3           | 21           | 1,69             |
| 2,1           | 11           | 4,41             |
| 1,7           | 30           | 2,89             |
| 1,6           | 14           | 2,56             |
| 0,9           | 24           | 0,81             |
| $\sum X$ 14,4 | $\sum Y$ 206 | $\sum X^2$ 22,12 |

$$\sum X^2 = 1.69 + 3.61 + 2.25 + 1.21 + 1 + 1.69 + 4.41 + 2.89 + 2.56 + 0.81 = 22.12$$

### Изчисляване на регресионния коефициент $\beta$

$$\beta = \frac{(n \cdot \sum (X \cdot Y)) - (\sum X \cdot \sum Y)}{(n \cdot \sum X^2) - (\sum X)^2}$$

$$\beta = \frac{(10 \cdot 282.4) - (14.4 \cdot 206)}{(10 \cdot 22.12) - (14.4)^2}$$

$$\beta = \frac{2824 - 2966.4}{221.2 - 207.36}$$

$$\beta = \frac{-142.4}{13.84}$$

$$\beta = -10.29$$

Всяка единица повишение на дозата с 1 мкг/кг води до намаляване на времето за реакция с 10.29 секунди.

### Изчисляване на константата $a$

$$a = \bar{y} - (\beta \cdot \bar{x})$$

### Изчисляване на $\bar{x}$

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{14.4}{10} = 1.44$$

### Изчисляване на $\bar{y}$

$$\bar{y} = \frac{\sum y}{n}$$

$$\bar{y} = \frac{206}{10} = 20.6$$

**Изчисляване на константата a**

$$a = \bar{y} - (\beta \cdot \bar{x})$$

$$a = 20.6 - (-10.29 \cdot 1.44)$$

$$a = 35.42$$