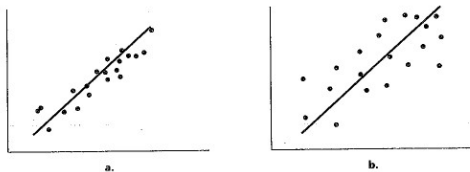


Регресија и корелација

1. Власник фирме жели да проучи везу између наплате од стране купаца у првом месецу X и укупне годишње наплате од купца Y . Добијени су подаци:

| | | | | | |
|-----|-----|----|-----|-----|-----|
| x | 10 | 5 | 25 | 100 | 30 |
| y | 120 | 75 | 250 | 800 | 360 |

- а) Нацртати дијаграм распршености.
 - б) Да ли изгледа да су корелисане X и Y , и ако јесу, како?
 - в) Одредити оцену корелације.
2. Нека x представља број недеља у којима је особа била на дијети, а Y број килограма које је изгубила. Добијен је узорак: $(1, 0.6)$, $(2, 0.7)$, $(3, 1)$, $(4, 1.3)$.
- а) Нацртати дијаграм распршености.
 - б) Одредити једначину линеарне регресије Y у односу на x .
 - в) Предвидети губитак након 5 година.
 - г) Одредити резидуале и SSE .
 - д) Одредити суму резидуала.
 - ђ) Одредити $\hat{\sigma}^2$.
 - е) Одредити R^2 - коефицијент детерминације модела.
3. За који од следеће две групе података се очекује да има већу оцењену дисперзију?



4. Проучава се однос између x , масе аутомобила, и Y , потрошње бензина. Подаци су дати у бази *mtcars*.
- а) Приказати дијаграм распршености између одговарајућих обележја.
 - б) Направити модел линеарне регресије зависности потрошње од масе. Одредити коефицијент нагиба регресионе праве и додати је на дијаграм распршености.
 - в) Оценити σ^2 и σ .
 - г) Одредити 95% интервал поверења просечне потрошње горива за све аутомобиле тежине 2 тоне.
 - д) Одредити 95% интервал предвиђања просечне потрошње горива аутомобила тежине 2 тоне.
 - ђ) Тестирати да ли маса има значајан ефекат на потрошњу са нивоом значаности $\alpha = 0.01$.
 - е) Одредити који проценат дисперзије у потрошњи је објашњен разликама у маси аутомобила.

Задаци за вежбу за семинарски - вишеструка линеарна регресија

5. Проучавана је веза цене књиге у зависности од броја страница и укоричености. Подаци су дати у бази *E1.19* у пакету *SenSrivastava*.
- а) Направити модел зависности. Да ли укориченост значајно утиче на цену књиге?

- б) Колико је у просеку укоричена књига скупља од папирне књиге, ако имају исти број страна?
- в) Предвидети цену укоричене књиге са 100 страна.
- г) Ако Пера жели да купи неукоричену верзију Гетеовог Фауста, која има 200 страна, одредити колико би новца морао да издвоји, да би био 90% сигуран да може да купи књигу.
6. Вршено је истраживање зависности плата од пола, за професора једног Америчког универзитета. Подаци се налазе у пакету *carData* у бази *Salaries*.
- а) Да ли постоји значајна разлика у платама између полова? Тестирати са нивоом значајности $\alpha = 0.05$.
- б) Да ли се закључак мења ако тестирамо контролишући по струци, звању и годинама радног стажа?

Задаци за вежбу

1. Истраживач је заинтересован за проучавање везе између броја сати у току дана проведених испред телевизора, X , и нивоу читања у трећем разреду, Y . Добијени су следећи подаци:

| | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|
| x | 3 | 5 | 6 | 2 | 0 | 1 | 3 | 5 | 6 | 2 |
| y | 4 | 3 | 2 | 4 | 5 | 5 | 3 | 2 | 1 | 5 |

- а) Нацртати дијаграм распршености.
- б) Да ли изгледа да су корелисане X и Y , и ако јесу, како?
- в) Одредити оцену корелације.
2. Нека x представља температуру у степенима целзијуса између 15 и 16 часова, а Y потрошњу електричне енергије изражене на скали од 1 до 10. Добијени су следећи подаци: (21,4), (31,7), (25,5), (34,8), (30,7), (24,4.5), (32,6), (30,6), (38,9), (38,9.5).
- а) Нацртати дијаграм распршености.
- б) Одредити линију регресије.
- в) Предвидети вредност потребе за електричном енергијом ако је температура $35^{\circ}C$.
3. Нека x представља касе који раде у супермаркету, а Y време потребно да купац дође на ред. Добијени су подаци:

| | | | | | | | |
|-----|---|----|----|----|----|----|----|
| x | 3 | 4 | 5 | 10 | 5 | 7 | 8 |
| y | 9 | 11 | 14 | 26 | 15 | 17 | 17 |

- а) Одредити једначину линеарне регресије.
- б) Предвидети вредност Y за свако x .
- в) Одредити резидуале и SSE .
- г) Одредити $\hat{\sigma}^2$.
- д) Одредити суму резидуала.
4. Проучава се веза између прихода домаћинства у хиљадама (x) и потрошње електричне енергије (Y). Добијени су подаци:

| | | | | | | | | |
|-----|------|------|------|------|------|------|------|------|
| x | 21.0 | 30.0 | 40.0 | 55.0 | 60.0 | 75.0 | 88.0 | 95.0 |
| y | 1.9 | 3.0 | 4.5 | 5.0 | 6.5 | 7.0 | 9.0 | 9.5 |

- а) Оценити коефицијент нагиба регресионе линије.
- б) Оценити σ^2 и σ .

- в) Одредити 90% интервал поверења просечне потрошње електричне енергије породице са примањима 50000 годишње.
- г) Одредити 90% интервал предвиђања потрошње електричне енергије породице са примањима 50000 годишње.
5. Конобар жели да одреди колико је просечно време које особа треба да чека да буде послужена у зависности од броја тренутно заузетих столова. Подаци су:

| | | | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|----|----|
| x | 5 | 5 | 10 | 12 | 15 | 20 | 20 | 25 | 27 | 30 |
| y | 10 | 12 | 14 | 13 | 18 | 19 | 21 | 23 | 30 | 35 |

- а) Нацртати дијаграм распршености.
- б) Тестирати $H_0(\beta = 0)$ против $H_1(\beta \neq 0)$, са нивоом $\alpha = 0.05$. Ако се одбацује нулта хипотеза, оценити параметре лиенарне регресије.
- в) Који проценат дисперзије у времену послужења је објашњен разликама у броју заузетих столова?