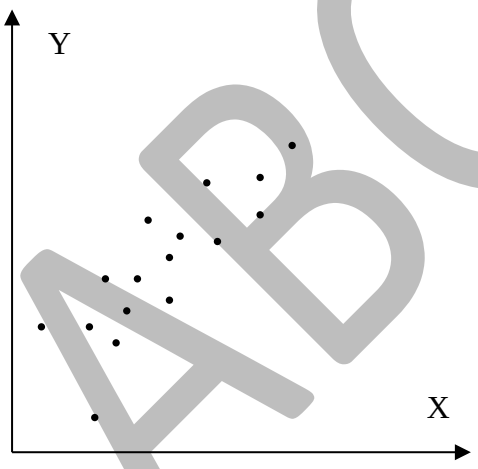
# ДИАГРАМИ НА РАЗСЕЙВАНЕ

## 3.1. Какво представляват диаграмите на разсейване

В редица случаи за дадено електронно изделие се съди не по един, а по няколко показателя, които характеризират различни страни на качеството. Като правило стойностите им имат случаен характер, т.е. те могат да бъдат разглеждани като случайни величини. Анализът на разсейването на всяка от тях може да се извърши с помощта на хистограми, но опитните данни съдържат значително повече информация, която трябва да се извлече и онагледи с подходящи средства. Едно твърде ефикасно средство са *диаграмите на разсейване*. Основното им предназначение е да изяснят дали между два показателя на качеството X и Y съществува някаква връзка (зависимост).

*Диаграмата на разсейване* (ДР), наричана още *корелационно поле*, представлява графично изображение на опитните данни в координатната система *x-y* (Фиг. 3.1). По абцисната ос се нанасят стойностите на показателя X, а по ординатната – на показателя Y. Всяка точка от диаграмата характеризира резултатите от даден опит ( наблюдение, измерване). След нанасяне на резултатите от всички опити върху диаграмата се получава един “облак” от точки, чиято конфигурация дава възможност да се проследи визуално наличието (или отсъствието) на определена зависимост между X и Y. Като правило тази зависимост не е напълно определена в смисъл, че на дадена стойност на показателя X съответства строго определена стойност на показателя Y. Обикновено тя се проявява като тенденция на изменение на едната величина при нарастване или намаляване на другата. Такъв тип зависимост се нарича *вероятностна (стохастическа) зависимост*.




**Фигура 3.1.** Общ вид на диаграма на разсейване

Обикновено ДР се построяват когато трябва да се изясни дали съществува някаква връзка и какъв е нейният характер:

1. между влияещ фактор (причина) и характеристика (следствие);
2. между две характеристики;
3. между два фактора.

Влияещият фактор (причина) понякога се нарича още факторен признак, а характеристиката (следствието) се нарича - резултативен признак..

Говорейки конкретно за качеството, такива двойки променливи най-често се отнасят до:

1. към характеристика на качеството и фактора, влияещ върху нея;
2. към две различни характеристики на качеството;
3. към два фактора, влияещи върху една характеристика на качеството.

И трите категории анализ са изключително важни, защото:

- в първия случай, при наличие на корелационна зависимост, причинният фактор оказва значително влияние върху характеристиката на

качеството и следователно, ако причинният фактор се държи под контрол, тогава е възможно, първо, да се постигне стабилност на характеристиката на качеството, и второ, да се определи нивото на контрол, необходимо за необходимия показател за качество;

- във втория случай, ако има корелационна зависимост между две различни характеристики на качеството, е възможно например да се контролира само една от тях;
- в третия случай наличието на корелация между отделни фактори значително улеснява контрола на процеса от технологична, времева и икономическа гледна точка.

Ако между съпоставяните променливи се предполага наличие на причинно-следствена връзка между сравнените двойки променливи, тогава при изграждането на диаграма на разсейване причинните фактори обикновено се обозначават с променливата ***x*** и се нанасят по хоризонталната ос (абсциса); характеристиките, като правило, се означават с променливата ***y*** и се нанасят по вертикалната ос (ординатна ос).

**3.2. Построяване на диаграмите на разсейване**

Последователност на работа:

1. Събират се двойки данни (*x*, *y*), състоящи се от стойностите на величините, съответстващи на всеки един отделен опит (наблюдение), между които искаме да проучим връзката. Желателно е да се съберат поне 25–30 двойки данни, тъй като при малък брой опити зависимостта между *X* и *Y* може да не се прояви в достатъчна степен.

Данните се подреждат в таблица по реда на получаването им, в следния ред:

- Номер на опита
- Време на провеждане на опита
- Стойност на величината *X*, стойност на величината *Y*.

2. Определят се максималните и минималните стойности за *x* и *y*. Въз основа на разликата между техните максимални и минимални стойности се задават размерите и скалите на осите и е по-добре да се направят приблизително еднакви, така че диаграмата да се чете по-лесно.

3. Изгражда се графика, върху която се нанасят данните. Ако няколко еднакви стойности попадат върху една и съща точка на графиката, тогава съответните точки се посочват с помощта на концентрични кръгове (точка в кръг, в два, три кръга) или се изчертава втора, трета точка до първата точка .

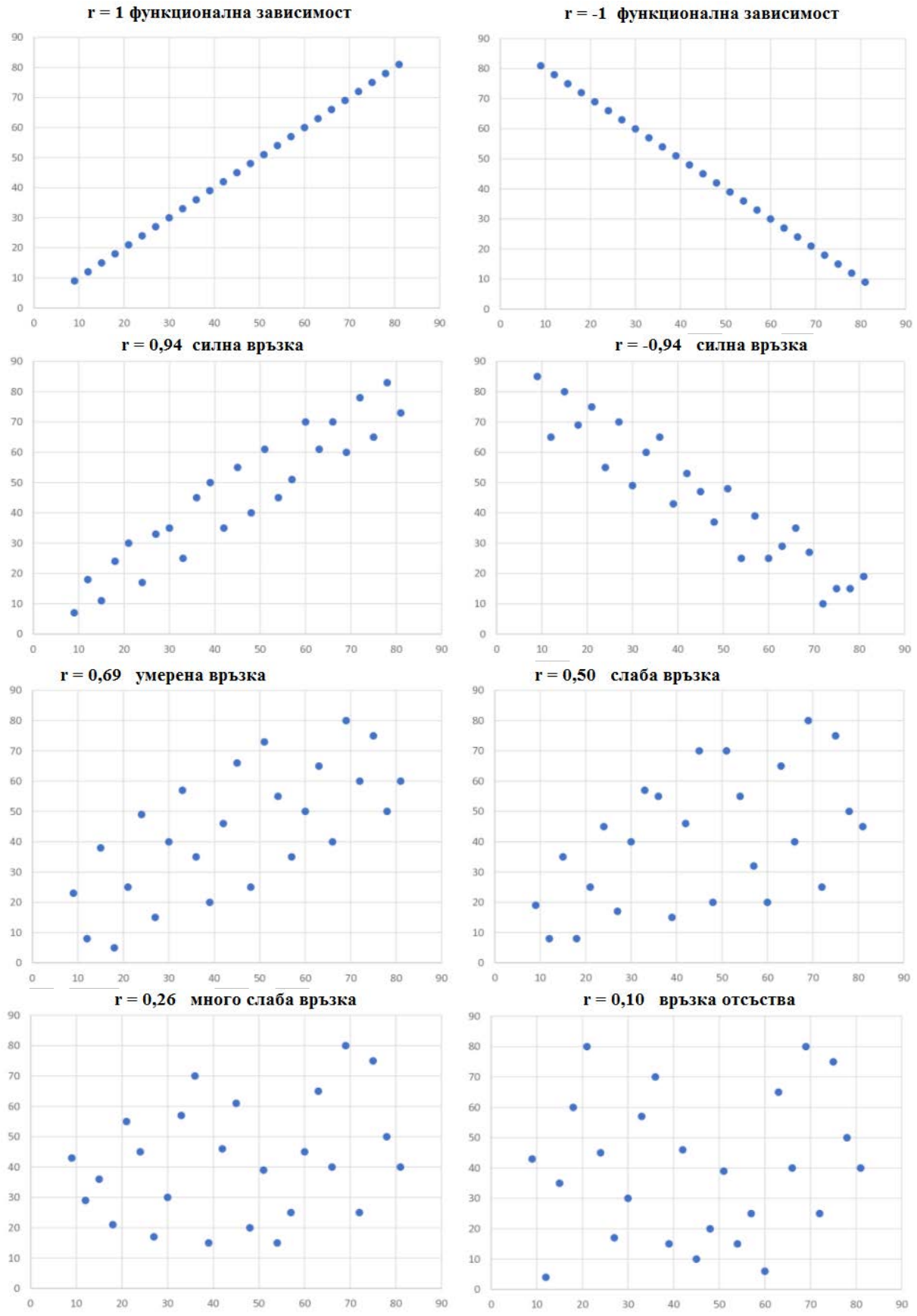
4. На графиката се прилагат всички необходими обозначения: името на диаграмата, нейния съставител, дата, интервал от време на получаване на опитните данни, брой двойки данни, мерни единици за всяка ос и т.н.:

Диаграмата може да се допълни с някои числени характеристики на разпределението на величините *x* и *y*, за които ще стане дума по-нататък.

В зависимост от стойностите на *x* и *y*, графиките могат да имат различен външен вид и изградените графики трябва да могат да се четат. Нека да видим как се прави това.

По-долу на Фиг. 2 са представени различни видове графики. Графиката ни позволява да видим от първа ръка естеството и колко силна е връзката между съответните променливи *x* и *y*. По-долу ще научим и как да определим каква е степента на силата на връзката, наречена *коефициент на корелация*.

Коефициентът на корелация *r* може да приеме стойности от -1 до +1, т.е. $-1 \le r \le 1$. В този случай, колкото по-близо е стойността на коефициента до ± 1,




Фиг. 2. Различни видове диаграми на разпределение, в зависимост от силата на връзката

толкова по-силна е връзката. Колкото по-близо е до нулата, толкова по-слаба е връзката. В ± 1 връзката е пълна (нарича се още функционална, тъй като всяка стойност на x съответства на строго определена стойност на y). При нула изобщо няма връзка.

Знакът плюс или минус посочва посоката на връзката - права или обратна: при плюс стойността на y се увеличава с увеличаване на стойността на *x*; с минус, напротив, намалява.

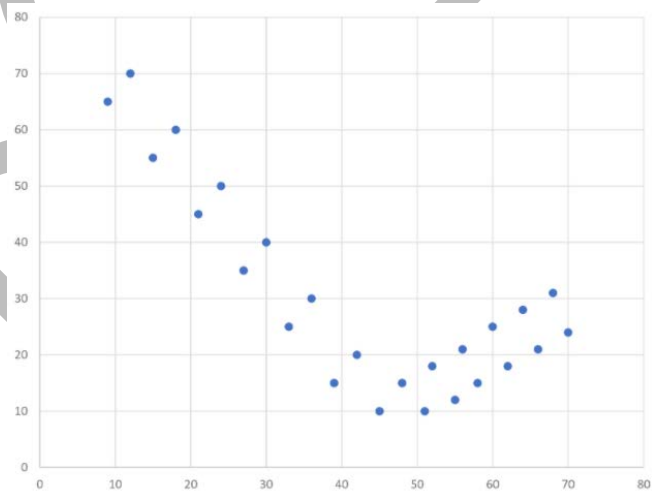
Що се отнася до оценката на силата на връзката,тогава в различни източници има различни класификации (градации). Например, може да се разглежда следната класификация

от ±0,81 до ±1,0 – силна връзка;
от ±0,61 до ±0,8 – умерена връзка;
от ±0,41 до ±0,6 – слаба връзка;
от ±0,21 до ±0,4 – много слаба връзка;
от 0 до ±0,2 – връзка отсъства.

Фиг. 2, показва различни видове диаграми на разсейване, със съответните стойности на коефициента на корелация *r*, посочени по-горе.

При липса на връзка (корелация) между изследваните параметри точките на диаграмата са разположени хаотично. Виждаме практически същата картина със слаба сила на връзката. Умерената силаа на връзката се характеризира с по-голяма степен на подреденост и доста равномерно разстояние на нанесените точки от въображаемата средна линия. Силната връзка клони в по-голяма степен към такава въображаема линияр а при r = 1, графиката всъщност представлява линия.

В случаите, показани на фиг. 2, корелацията е линейна (въображаемата средна линия е права линия), но в реалнии условия графиката може да има различна, нелинейна (криволинейна) форма, например такава, както показаната на фиг. 3.




Фиг. 3.

Най-лесният начин да изчислим коефициента на корелация е в MS Excel.

Математическата формула за изчисляване на коефициента на корелация може да се изчисли и без MS Excel или друга подобна програма. Всички съответни изчисления се правят в рамките на така наречения *корелационен анализ*.

### *3.3. Откриване на груби грешки*

В някои случаи, малка част от точките са разположени значително встрани от основния “облак” от точки. Това се дължи обикновено или на груби грешки при

измерването (определянето) на стойностите на величините, или на съществени изменения в условията на провеждане на съответните на тези точки опити. Във всички случаи такива точки изискват внимателно оглеждане на условията на експеримента и при констатиране на една от горепосочените причини те се отстраняват от комплекта данни при по-нататъшен анализ.

**3.4. Корелационен анализ**

Коефициентът на корелация се изчислява по формулата:

$$r = \frac{\sum xy - \sum x \times \sum y / n}{\sqrt{(\sum x^2 - (\sum x)^2 / n) \times (\sum y^2 - (\sum y)^2 / n)}}$$

Трябва да се отбележи, че този инструмент (диаграма на разсейване и пресмятане на коефициента на корелация ) не е 100% гаранция, че две променливи с висок коефициент на корелация са наистина свързани помежду си: има така наречените фалшиви корелации, при които изчислената стойността на коефициента на корелация е висока, но при това няма зависимост на една характеристика от друга. Причините за появата на фалшиви корелации могат да бъдат много разнообразни, например наличието на някаква друга характеристика, скрита от нас, която едновременно засяга и двете характеристики, които изучаваме.

Възможни са и обратни ситуации: връзката наистина съществува, но не се установява с този инструмент. Причините за това отново могат да бъдат много различни - от недостатъчен брой събрани данни до прекомерно голяма грешка в измерването.

Но това не означава, че този инструмент не може да се използва! Напротив, той е доста прост, но ефективен инструмент за статистически анализ. Необходимо е само да се вземе предвид, първо, че само тези, които са запознати с изследвания процес, могат правилно да оценят диаграмата на разсейване и коефициента на корелация; второ, коефициентът на корелация, получен по този начин, е произволна стойност и не е физическа константа.

С други думи, използването на този инструмент изисква известна доза предпазливост, внимание към детайлите и познаване на същността на проблема.

Друг важен момент е, че коефициентът на корелация ни позволява да оценим степента на сила на връзката между резултативния признак (*y*) и въздействащия му фактора (*x*), но не отговаря на въпроса: с колко единици ще се измени резултативния признак при измененияе на фактора с една единица. Отговорът на този въпрос може да се получи с помощта на друг инструмент – *регресионен анализ*.

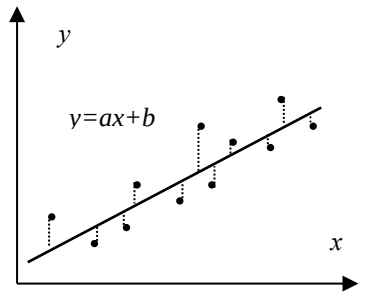
**3.4. Линия на регресия**

Констатацията, че между две величини съществува значителна корелация, веднага поражда въпроса дали може да се определи такава зависимост между *X* и *Y*, че при зададена стойност на величината *X* да се получи (макар и приблизително) каква ще бъде стойността на величината *Y* – *да се запише емпирична математическа зависимост*. Точките от диаграмата на разсейване се разполагат при наличие на линейна корелация около/на права линия. Уравнението на тази права в общ вид е

$$y = a + bx \qquad (1)$$

и се нарича *регресионно уравнение*. Линията, която съответства на (1) се нарича *линия (права) на регресия*. Променливата *x* се нар. независима променлива, а *y* - зависима променлива. Параметърът *a* е константа, а параметърът *b* се нар. коефициент на регресия. С помощта на регресионното уравнение може да се предскаже стойността на зависимата променлива *y* при произволна стойност на независимата променлива *x*.

Определянето на параметрите *a* и *b* е задача на регресионния анализ. Това става като се използват опитните данни. Критерият, заложен при определянето им, се състои в минимизиране на сумата на квадратите на отклоненията по координатата *y* на опитните резултати от линията на регресия (фиг. 4.). Методът, основаващ се на този критерий, е известен като метод на най-малките квадрати (МНК).




Фиг. 4. Регресионна линия

### *3.5. Построяване на линията на регресия*

Нека се разполага с *n* чифта опитни стойности на променливите *x* и *y* : ($x_1$,$y_1$); ($x_2$,$y_2$); … ; ($x_n$,$y_n$). Означават се с $\hat{a}$ и $\hat{b}$ някакви (за сега произволни) стойности на неизвестните параметри *a* и *b* в ур. (1), които се наричат *оценки*. Заместени в уравнението, те дават възможност да се определят предсказаните стойности

$$\hat{y}_i = \hat{a} + \hat{b}x_i \quad i = 1, 2,..., n \qquad (2)$$

Разликите

$$e_i = y_i - \hat{y} = y_i - \left(\hat{a} + \hat{b}x_i\right) \quad i = 1, 2,..., n \qquad (3)$$

се наричат *остатъци*. Критерият, по който се търси минимумът, може да бъде записан по следния начин:

$$Y(a,b) = \sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (y_i - a - bx)^2 = \min(a,b) \qquad (4)$$

Стойностите на $\hat{a}$ и $\hat{b}$, при които се получава минимумът на (4), е *МНК – оценки* на параметрите на линейното регресионно уравнение (2).

Получаването им може да стане с помощта на следния алгоритъм:

**1.** Получават се средните за *x* и *y* по формулите

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

**2.** С помощта на формулите долу се определят $S_x$ и $\hat{K}_{xy}$

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^{\,2} - n\bar{x}^2\right)}$$

$$S_y = \sqrt{\frac{1}{n-1}\sum(y_i - \bar{y})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^{\,2} - n\bar{y}^2\right)}$$

$$\hat{K}_{xy} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\overline{xy}\right)}$$

**3.** Пресмята се оценката на регресионния коефициент $\hat{b} = \frac{\hat{K}_{xy}}{S^2{}_x}$

**4.** Пресмята се оценката на свободния член (константата) $\hat{a} = \bar{y} - \hat{b}\bar{x}$

## ЗАДАЧИ ЗА ИЗПЪЛНЕНИЕ

**Задача 1.** Проведен е експеримент за установяване на зависимостта между количеството на нанесената паста и степента на проникване на пастата през ситото при ситопечат на дебелослойни платки. Да се начертае диаграмата на разсейване, да се начертае линията на регресия и се определи коефициента на корелация. Експериментът съдържа 26 чифта данни за количеството на нанесената паста в g/m$^2$ и степента на проникване в отн. единици. Данните са нанесени в табл. 1.

**Табл. 1.**Опитни данни, съдържащи стойностите на количеството паста (x) и степента на проникване (y)

| № | Дата | x | y | № | Дата | x | y |
|---|---|---|---|---|---|---|---|
| 1 | 11.04 | 230 | 0,90 | 14 | 12.04 | 184 | 0,76 |
| 2 | 11.04 | 136 | 0,61 | 15 | 12.04 | 205 | 1,06 |
| 3 | 11.04 | 282 | 1,09 | 16 | 12.04 | 223 | 0,94 |
| 4 | 11.04 | 212 | 0,87 | 17 | 12.04 | 239 | 0,99 |
| 5 | 11.04 | 271 | 0,98 | 18 | 12.04 | 147 | 0,70 |
| 6 | 11.04 | 228 | 0,97 | 19 | 13.04 | 213 | 0,93 |
| 7 | 11.04 | 137 | 0,65 | 20 | 13.04 | 136 | 0,45 |
| 8 | 11.04 | 217 | 0,93 | 21 | 13.04 | 224 | 0,94 |
| 9 | 12.04 | 187 | 0,88 | 22 | 13.04 | 256 | 1,01 |
| 10 | 12.04 | 241 | 0,92 | 23 | 13.04 | 171 | 0,91 |
| 11 | 12.04 | 206 | 0,76 | 24 | 13.04 | 296 | 1,05 |
| 12 | 12.04 | 144 | 0,67 | 25 | 13.04 | 201 | 0,61 |
| 13 | 12.04 | 232 | 0,98 | 26 | 13.04 | 182 | 0,74 |

**Задача 2.** За определяне на корелацията между относително увеличение на вискозитета на паста за дебелослойни ХИС в % и количество на вложената смола в т.ч. е проведен експеримент, включващ 30 опита. Една част от опитите са проведени с един вид смола (А), а друга част – с друг вид (В). Опитните данни са показани в табл. 2.

1. Постройте ДР, с използването на всички опитни данни. Определете вида на корелацията, постройте линията на регресия и определете коефициента на корелация.
2. Разделете данните на две групи (слоя), така че едната да съдържа само данните с използване на смолата А, а другата – тези на смолата В. Постройте две ДР за двата слоя. Определите вида на корелациите, ако

съществуват такива, постройте линията на регресия и определете коефициента на корелация.

**Табл.2.** Опитни данни, съдържащи стойностите на количеството смола (X) и относителното увеличаване на вискозитета на паста за дебелослоен ситопечат

| № | Вид смола | X | У | № | Вид смола | X | У |
|---|---|---|---|---|---|---|---|
| 1 | A | 8,5 | 482 | 16 | A | 9,6 | 518 |
| 2 | A | 11,3 | 531 | 17 | A | 11,8 | 536 |
| 3 | A | 10,4 | 525 | 18 | A | 10,0 | 520 |
| 4 | A | 9,2 | 490 | 19 | B | 11,5 | 516 |
| 5 | B | 12,0 | 516 | 20 | B | 10,8 | 492 |
| 6 | B | 10,8 | 495 | 21 | B | 8,5 | 455 |
| 7 | B | 10,6 | 486 | 22 | B | 10,5 | 480 |
| 8 | B | 8,2 | 465 | 23 | A | 10,6 | 520 |
| 9 | A | 9,8 | 520 | 24 | A | 9,5 | 492 |
| 10 | A | 11,0 | 545 | 25 | A | 11,2 | 543 |
| 11 | A | 10,8 | 530 | 26 | A | 9,8 | 510 |
| 12 | A | 8,0 | 491 | 27 | B | 11,2 | 502 |
| 13 | B | 8,8 | 474 | 28 | B | 8,0 | 450 |
| 14 | B | 10,2 | 502 | 29 | B | 9,8 | 475 |
| 15 | B | 9,5 | 493 | 30 | B | 10,4 | 470 |