

معرفی روش های آماری و نرم افزار Minitab

دوره پایتون و یادگیری ماشین



</Monolearn>

مباحث پایه در آمار

- میانه یا Median
- میانگین یا Mean
- واریانس و کوواریانس یا Variance and covariance

میانہ

{۲۶, ۱۰, ۱۵, ۱۶, ۱۱, ۲۳, ۱۳}

{۲۶, ۲۳, ۱۶, ۱۵, ۱۳, ۱۱, ۱۰}

{۲۹, ۲۳, ۵۶, ۱۲, ۱۴, ۲۳, ۴۰, ۲۳, ۲۳, ۲۱, ۵, ۷, ۱۳, ۳}

{۵۶, ۴۰, ۲۹, ۲۳, ۲۳, ۲۳, ۲۳, ۲۱, ۱۴, ۱۳, ۱۲, ۷, ۵, ۳}

$$۲۱ + ۲۳ = ۴۴$$

$$۴۴ \div ۲ = ۲۲$$

میانگین

۲۹, ۲۳, ۵۶, ۱۲, ۱۴, ۲۳, ۴۰, ۲۳, ۳۹, ۲۳, ۲۰, ۱۳, ۵, ۷, ۳

$$۳۳۰ \div ۱۵ = ۲۲$$

تفاوت واریانس و کوواریانس

- واریانس شاخصی برای پراکندگی نمونه ها حول یک میانگین است.
- کوواریانس شاخصی برای تغییرات یک متغیر نسبت به متغیر دیگر است.

برای تحلیل رابطه دو متغیر معمولاً از **کوواریانس** استفاده می شود.
اگر دو متغیر یکسان باشند مقدار واریانس و کوواریانس برابر خواهد بود.

محاسبه کوواریانس

فرض کنید مجموعه آماری دارای دو صفت خاصه ی A و B باشد. آنگاه مجموعه داده ها به صورت زیر خواهد بود:

$\{ (a_1, b_1) \dots (a_n, b_n) \}$

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

محاسبہ کوواریانس

x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
12	20	-9.3	-21.2	197.16
30	60	8.7	18.8	163.56
15	27	-6.3	-14.2	89.46
24	50	2.7	8.8	23.76
14	21	-7.3	-20.2	147.46
18	30	-3.3	-11.2	36.96
28	61	6.7	19.8	132.66
26	54	4.7	12.8	60.16
19	32	-2.3	-9.2	21.16
27	57	5.7	15.8	90.06
$\bar{x} = 21.3$	$\bar{y} = 41.2$			$\Sigma = 962.4$

محاسبہ کوواریانس

x	y	$(x_i - \bar{x})(y_i - \bar{y})$
12	20	197.16
30	60	163.56
15	27	89.46
24	50	23.76
14	21	147.46
18	30	36.96
28	61	132.66
26	54	60.16
19	32	21.16
27	57	90.06
$\bar{x} = 21.3$	$\bar{y} = 41.2$	$\Sigma = 962.4$

$$Cov(x, y) = s_{xy} = \frac{962.4}{n - 1}$$

$$\frac{962.4}{9}$$

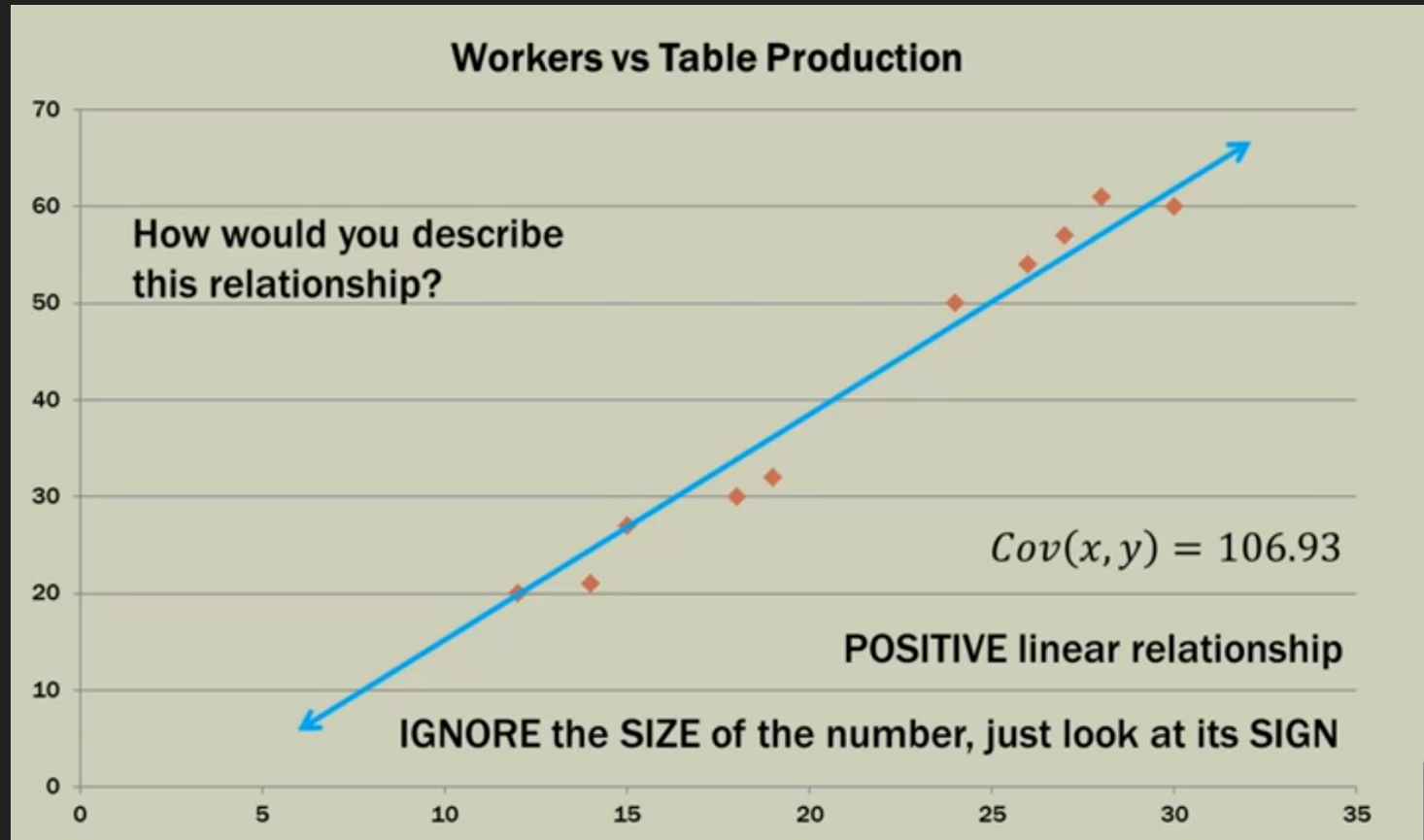
$$Cov(x, y) = 106.93$$

کوواریانس مثبت و منفی

مقدار کوواریانس اهمیت چندانی ندارد بلکه علامت آن مهم است.

کوواریانس مثبت بیانگر تغییرات خطی مثبت و منفی بیانگر تغییرات خطی منفی است.

محاسبہ کوواریانس



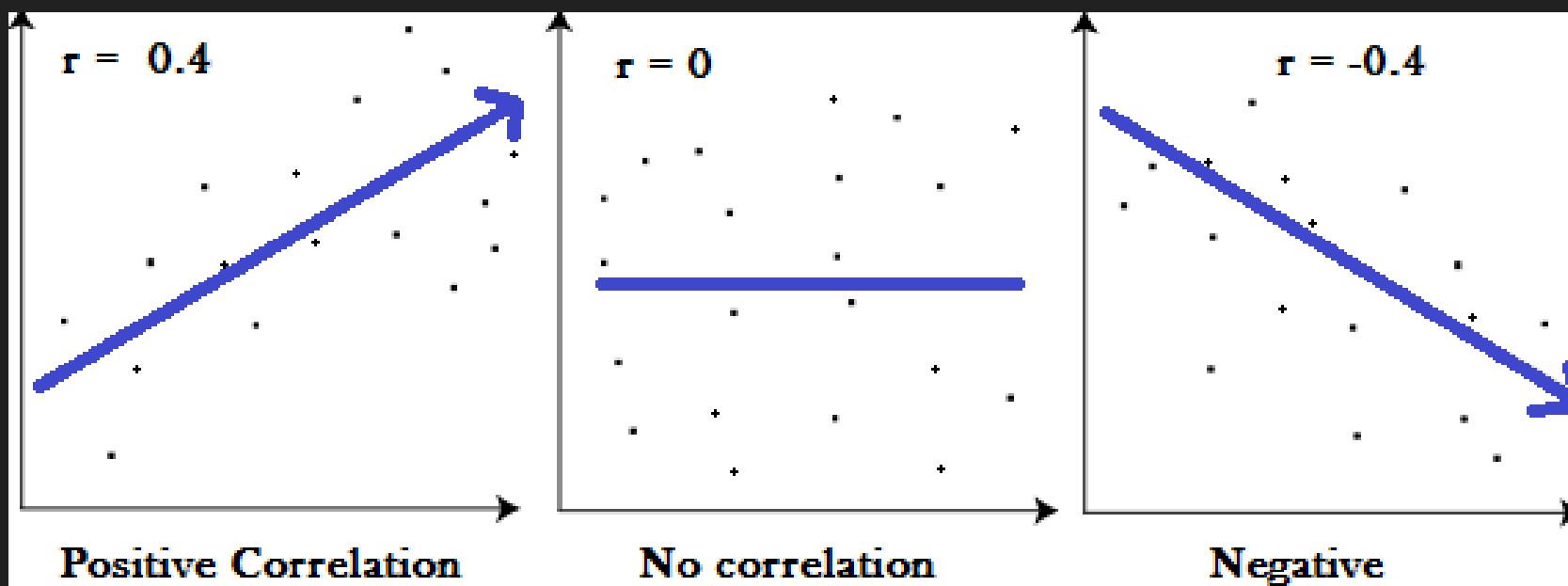
ضریب همبستگی

روش آماری Pearson's Product Moment Coefficient برای محاسبه همبستگی میان داده های عددی کاربرد دارد.

کوواریانس به تنهایی معنای خاصی ندارد زیرا که قدرت رابطه ی میان صفات خاصه را نمایش نمی دهد. از این جهت، باید از ضریب همبستگی استفاده شود.

ضریب همبستگی

- مقدار همبستگی می تواند از -1 تا +1 باشد.
- مقدار صفر یعنی همبستگی نداریم.
- هر چه عدد به +1 نزدیک تر باشد، همبستگی مثبت بیشتر است و هر چه به -1 نزدیک باشد همبستگی منفی بیشتر خواهد بود.



محاسبه ضریب همبستگی

روش آماری Pearson's Product Moment Coefficient برای محاسبه همبستگی میان داده های عددی کاربرد دارد.

کوواریانس به تنهایی معنای خاصی ندارد زیرا که قدرت رابطه ی میان صفات خاصه را نمایش نمی دهد. از این جهت، باید از ضریب همبستگی استفاده شود.

محاسبه ضریب همبستگی

Subject	Age x	Glucose Level y	xy	x ²	y ²
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

محاسبه ضریب همبستگی

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

$$\sum x = 247$$

$$\sum y = 486$$

$$\sum xy = 20485$$

$$\sum x^2 = 11409$$

$$\sum y^2 = 40022$$

$$n = 6$$

The Correlation coefficient =

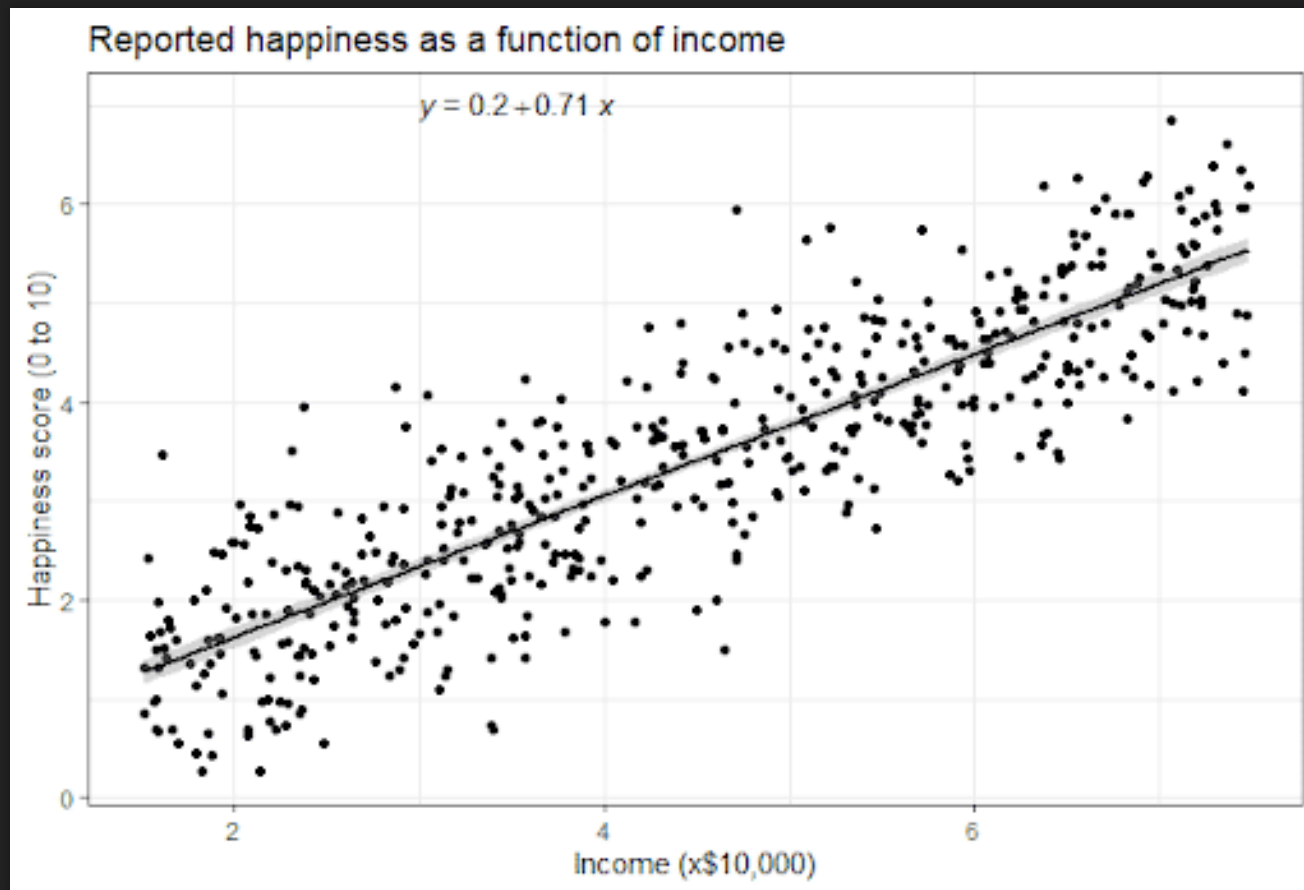
$$\begin{aligned} & 6(20485) - (247 \times 486) / [\sqrt{[6(11409) - (247^2)] \times} \\ & [6(40022) - 486^2]}] \\ & = 0.5298 \end{aligned}$$

رگرسیون یا Regression

برای پیش بینی مقادیر پیوسته مانند قیمت، میزان درآمد، قد، وزن و ... کاربرد دارد.

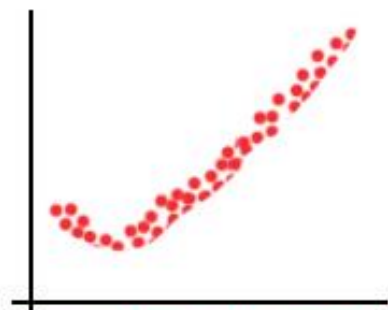
در مسائل علوم مختلف، مانند رشته ی مکانیک، اکثر مسائل از طریق روش های رگرسیون قابل حل است.

رگرسیون یا Regression



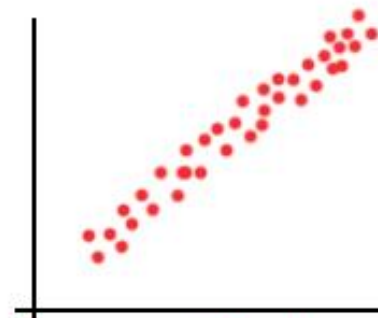
$$Y = a + bX$$

رگرسیون یا Regression



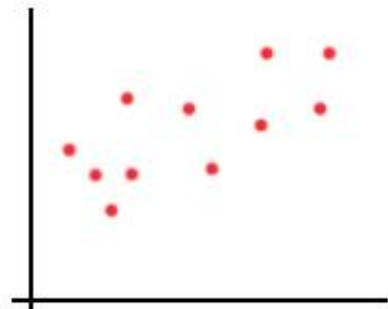
a

very strong but not linear



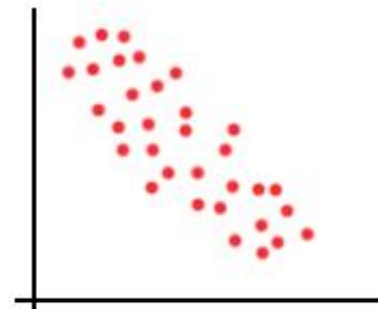
b

strong positive linear correlation



c

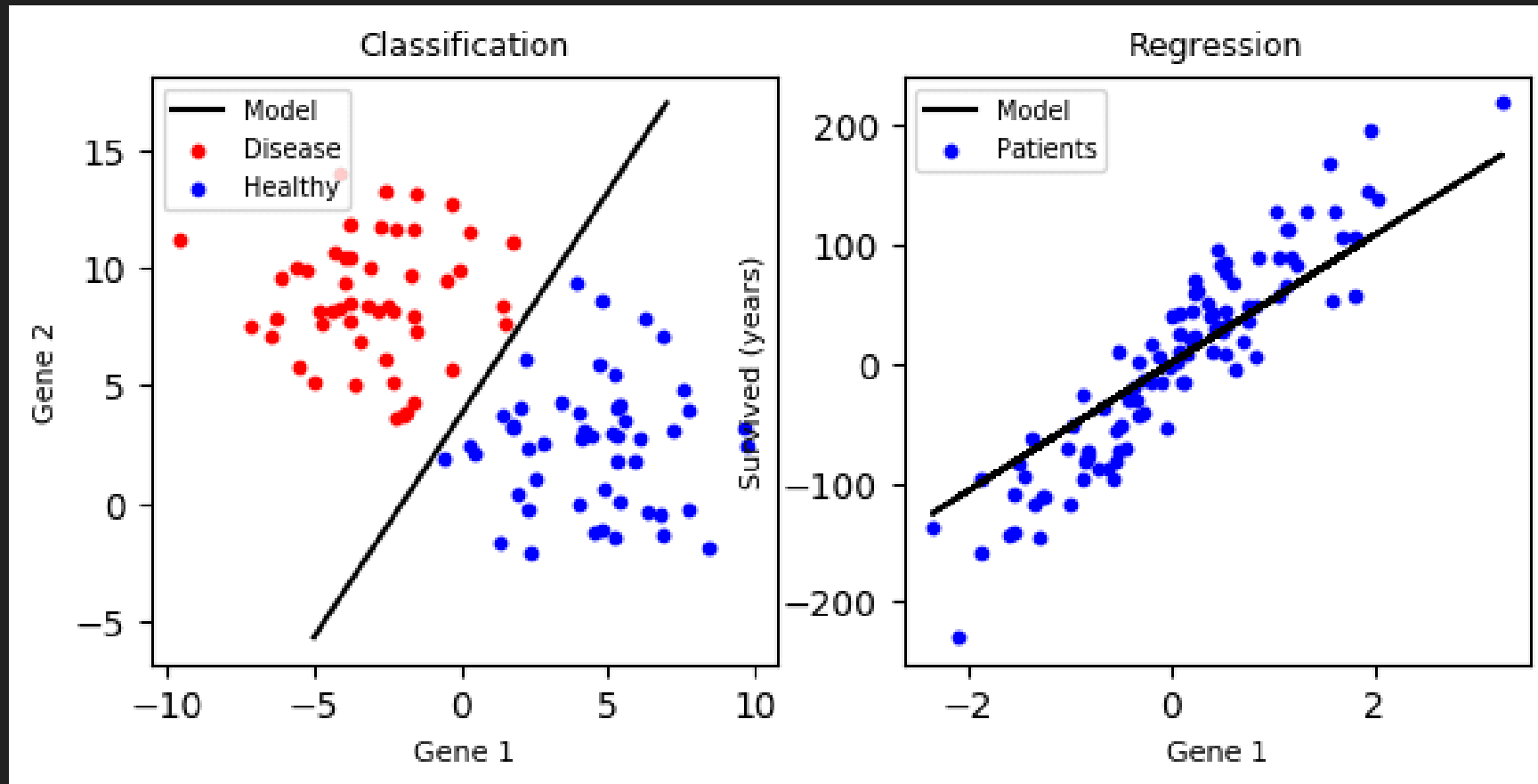
weak positive correlation



d

weak negative linear correlation

تفاوت Classification و Regression



طرح آزمایش یا Design Of Experiments

فرض کنید آزمایشی داریم شامل ۴ فاکتور (عامل، پارامتر یا متغیر) که بر یکدیگر اثر می گذارند، اگر هر فاکتور سه مقدار a ، b و c (سطح یا Level) را بتواند داشته باشد، تعداد کل آزمایش های قابل انجام برابر است با:

۳ به توان ۴ یا ۸۱ آزمایش یا L^4 که L بیانگر level و F بیانگر فاکتور است.

حالتی که تمام حالت های آزمایش را در نظر می گیرد Full Factorial و حالتی که بخشی از آن ها را در نظر می گیرد Fractional Factorial نامیده می شود.

حال فرض کنید تنها با انجام ۱۶ آزمایش یعنی ۲۰٪ از کل آزمایش ها بتوان به همان نتایج قبلی دست یافت، این امر همان چیزی است که منجر به ایجاد روش های مختلف طراحی آزمایش شده است.

با استفاده از طراحی آزمایش می توان ماتریس پارامترهای ورودی را (مانند حالت رندوم) ایجاد کرد و سپس با انجام آزمایش ها نتایج را بدست آورد و تحلیل کرد.

مزایای روش طراحی آزمایش در مرحله توسعه

۱. کاهش نوسانات خروجی فرآیند
۲. کاهش زمان توسعه فرآیند
۳. کاهش هزینه های کلی

مزایای روش طراحی آزمایش در مرحله طراحی مهندسی

۱. ارزیابی و مقایسه انواع طرحهای پایه
۲. ارزیابی انواع مواد
۳. انتخاب عوامل طراحی به طوری که محصول در شرایط مختلف کار کند
۴. تعیین عوامل کلیدی طراحی که بر کارایی محصول تأثیر می گذارند

روش Response Surface Methodology

روش RSM یا Response Surface Methodology یا روش سطح پاسخ زمانی استفاده می شود که تعداد آزمایش ها به دلیل تعداد زیاد متغیرهای مستقل انتخاب شده (تعداد فاکتورهای بالا)، زیاد باشد.

در این حالت دو مشکل وجود دارد. اول اینکه هزینه آزمایش های بسیار زیاد است. و دیگر اینکه برای انجام تمام آزمایش ها با محدودیت زمانی مواجه خواهیم شد. (با Full Factorial نتوان آزمایش را انجام داد.)

روش Response Surface Methodology

در روش سطح پاسخ، سطوح را با عددهای ۰، ۱ و -۱ به صورت کدگذاری شده می توان نشان داد که عدد ۰ سطح وسط، ۱ سطح بالا و -۱ سطح پایین است.

برای کار با RSM باید آزمایش ها را بعد از یک مطالعه عمیق برای شناسایی سطوح مناسب و عوامل مناسب و یافتن پاسخ های مناسب به منظور طراحی درست آزمایش، برنامه ریزی نمود.

روش Response Surface Methodology

به طور خلاصه این روش به طراحی یک آزمایش تجربی می پردازد و اثرات چندگانه متغیرها را مدلسازی می کند. سپس با ارائه ی یک مدل رگرسیون به برقراری ارتباطی بین جواب و فاکتورها می پردازد.

از مهمترین اهداف آن:

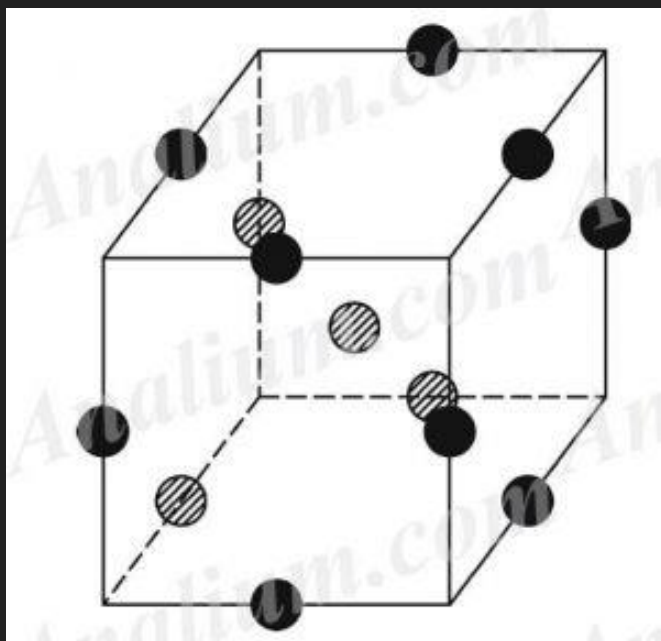
- بهبود فرآیند یا یافتن ورودی های بهینه
- رفع مشکلات و نقاط ضعف فرآیند
- پایدارسازی

برخی از روش های Response Surface Methodology

- Central composite design
- Box-Behnken design

روش Box-Behnken Design

شکل مقابل نمونه ای از طراحی Box-Behnken برای سه فاکتور است.



تعداد آزمایش های برای این روش از فرمول زیر محاسبه می شود که N تعداد آزمایش، K تعداد فاکتورها و C_0 تعداد نقاط مرکزی است که معمولاً برابر ۳ است:

$$N = 2k(k-1) + C_0$$

پایان

با تشکر از توجه تان، اوقات خوشی را برایتان آرزومندم.