○ امیدریاضی متغیرتصادفی (میانگین، مقدارمتوسط، ارزش انتظاری، μ_x، $E(X)$)

ـ امیدریاضی متغیرتصادفی x، حد متوسطی است که انتظار می‌رود برای متغیر تصادفی x اتفاق بیافتد.

ـ امیدریاضی همان میانگین موزون است که احتمالات درآن، نقش وزن‌ها (ضرایب) را بازی می‌کنند.

ـ دریک مجموعه از مقادیر یک جامعه امیدریاضی مرکز ثقل جامعه است.

امیدریاضی متغیر تصادفی x به صورت زیر نشان داده می‌شود:

x امیدریاضی متغیرگسستهٔ : $E(x)=\sum_{\forall x} x f_x(x)$ (به شرط همگرایی Σ)

x امیدریاضی متغیرپیوسته : $E(x)=\int_{-\infty}^{+\infty} x f_x(x)dx$ (به شرط موجود بودن ∫)

● نکات مهم امیدریاضی

۱ـ $\boxed{-\infty<E(x)<+\infty}$

۲ـ امیدریاضی یک تابع از متغیر تصادفی x با تابع توزیع احتمال $f_x(x)$ به فرم زیر است:

گسسته: $E[k(x)]=\sum_{\forall x} k(x)f_x(x)$

پیوسته: $E[k(x)]=\int_{-\infty}^{+\infty} k(x)f_x(x)dx$

۳ـ اگر a و b مقادیر ثابتی فرض شوند، روابط زیر برای امیدریاضی می‌توانند برقرار باشند:

$$E(\pm a)=\pm a$$
$$E(\pm bx)=\pm bE(x)$$
$$E(\pm ax\pm b)=\pm aE(x)\pm b$$

۴ـ اگر x و y دو متغیر تصادفی، h (x) و g (x) توابعی از x باشند، آن‌گاه روابط زیر برای امیدریاضی می‌توانند برقرار باشند:

$$\boxed{\begin{array}{l} E[g(x)\pm h(x)] = E[g(x)]\pm E[h(x)] \\ E(\ldots E(x)) = E(x) \\ E(\pm ax \pm by \pm c) = \pm aE(x)\pm bE(x)\pm c \end{array}}$$

۵ـ هنگام محاسبهٔ $E(g(x))$ تابع توزیع f (x) به هیچ شکلی تغییر نمی‌کند، به عنوان مثال:

$$E\left(x^2\right) = \sum x^2 f_x(x)$$

۶ـ هرگاه تابع چگالی به ازای x های مثبت تعریف شده باشد یعنی $x > 0$ می‌توان امید ریاضی را به طریق زیر محاسبه کرد:

$$E(x) = \int_0^\infty [1 - F(x)]dx$$

مثال ۱: تابع چگالی احتمالی متغیر تصادفی پیوسته x به صورت $f(x) = 1 - \frac{1}{2}x$ برای $0 \le x \le 2$ داده شده است. میانگین و $p(1 \le x < 2)$ به ترتیب از راست به چپ برابر است با: (اقتصاد ۷۰)

۱) $1 , \frac{3}{4}$ ۲) $\frac{2}{3} , \frac{1}{4}$ ۳) $1 , \frac{1}{4}$ ۴) $\frac{2}{3} , \frac{3}{4}$

حل : گزینه ۲ صحیح می‌باشد.

$$E(x)=\int_0^2 x f(x) dx = \int_0^2 x\left(1 - \frac{1}{2}x\right)dx = \int_0^2 \left(x - \frac{1}{2}x^2\right)dx = \left(\frac{x^2}{2} - \frac{x^3}{6}\right)\Bigg]_0^2 = \frac{4}{2} - \frac{8}{6} = \frac{4}{6} = \frac{2}{3} \Rightarrow \boxed{E(x) = \frac{2}{3}}$$

$$p(1 \le x < 2) = \int_1^2\left(1 - \frac{1}{2}x\right)dx = \left(x - \frac{x^2}{4}\right)\Bigg]_1^2 = \frac{1}{4} \Rightarrow \boxed{p(1 \le x < 2) = \frac{1}{4}}$$

مثال ۲: کدام تساوی در مورد عمل‌کننده امیدریاضی غلط است. (a و c مقادیر ثابت هستند)؟

۱) $E(c(a + x)) = a + cE(x)$ ۲) $E(a + x) = a + E(x)$

۳) $E[(cx)] = cE(x)$ ۴) $E(a) = a$

حل : گزینه ۱ صحیح می‌باشد.

$$E[c(a + x)] = E(ca + cx) = E(ca) + E(cx) = ca + cE(x)$$

مثال ۳: تابع چگالی کمیت تصادفی x به صورت زیر بیان شده است:

$$f(x) = \frac{2x}{9} \qquad 0<x<3$$

امیدریاضی کمیت تصادفی y که بر طبق رابطه $y = 2x + 1$ از کمیت x تبعیت می‌کند، چقدر است؟

۱) 3.5 ۲) 4 ۳) 2 ۴) 5

حل : گزینه ۴ صحیح می‌باشد.

برای محاسبهٔ امیدریاضی کمیت تصادفی y، ابتدا باید امیدریاضی کمیت تصادفی x را به دست آوریم:

$$E(x) = \int_0^3 x.f(x)dx = \int_0^3 x.\frac{2x}{9}dx = \int_0^3 \frac{2}{9}x^2dx = \frac{2}{27}x^3\Bigg]_0^3 = 2 \Rightarrow \boxed{E(x) = 2}$$

$$E(y) = E(2x + 1) = 2E(x) + 1 = 2(2) + 1 = 5 \Rightarrow \boxed{E(y) = 5}$$

مثال۴: تابع توزیع (توزیع تجمعی احتمال) کمیت تصادفی x به صورت زیر بیان شده است:

$$F(x)=\frac{x^2+3x}{10} \qquad 0<x<2$$

امیدریاضی کمیت تصادفی x کدام است؟ (اقتصاد ۷۷)

۱) $E(x)=0.4$ ۲) $E(x)=1$ ۳) $E(x)=1.13$ ۴) $E(x)=1.5$

حل : گزینه ۳ صحیح می‌باشد.

توجه کنید که برای بدست آوردن E (x) ، حتماً باید ابتدا تابع چگالی احتمال (f(x)) را حساب کنیم، بنابراین:

$$f_x(x)=\left(F_x(x)\right)'=\left(\frac{x^2+3x}{10}\right)'=\frac{2x+3}{10}$$

$$\int_0^2 x f_x(x)dx=\int_0^2\frac{x(2x+3)}{10}dx=\frac{1}{10}\int_0^2\left(2x^2+3x\right)dx=\frac{1}{10}\left(\frac{2x^3}{3}+\frac{3x^2}{2}\right)\Bigg]_0^2$$

$$\frac{1}{10}\left(\frac{16}{3}+\frac{12}{2}\right)=1.13$$

مثال۵: تابع احتمال (قانون توزیع احتمال‌ها)، به صورت زیر تعریف شده است:

x_i	-1	0	1	2
$p(x_i)$	0.3	0.3	0.3	0.1

امیدریاضی x^2، یعنی $E\left(x^2\right)$ کدام است؟ (مدیریت ۷۵)

۱) 0.04 ۲) 0.2 ۳) 1 ۴) 2

حل : گزینه ۳ صحیح می‌باشد.

$$E\left(x^2\right)=\sum x^2.p(x)=(-1)^2(0.3)+(0)^2(0.3)+(1)^2(0.3)+(2)^2(0.1)$$

$$E\left(x^2\right)=0.3+0.3+0.4=1$$

مثال۶: متغیر تصادفی x می‌تواند یکی از سه مقدار 5 و 4 و x_3 را انتخاب کند که احتمال آن‌ها به ترتیب 0.2، 0.5 و P_3 است. اگر میانگین متغیر تصادفی x برابر 6 باشد مقدار x_3 چقدر است؟

۱) 2 ۲) 5 ۳) 7 ۴) 10

حل : گزینه ۴ صحیح می‌باشد.

x_i	4	5	x_3	
$p_i=f_i$	0.5	0.2	P_3	$\Sigma p_i=1$

$$(I)\ \Sigma p_i=1 \Rightarrow 0.5+0.2+p_3=1 \Rightarrow 0.7+p_3=1 \Rightarrow \boxed{p_3=0.3}$$

$$(II)\ E(x)=\sum x f_x(x) \Rightarrow 4(0.5)+5(0.2)+\left(x_3\right)\left(p_3\right)=6 \Rightarrow 3+x_3p_3=6$$

$$\Rightarrow 0.3x_3=3 \Rightarrow \boxed{x_3=10}$$

مثال۷: جدول احتمال متغیر تصادفی x به صورت زیر است:

x_i	0	1	2	3
p_i	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{3}$	a

امیدریاضی x کدام است؟

۱) 1 ۲) $\frac{4}{3}$ ۳) $\frac{5}{3}$ ۴) 2

حل : گزینه ۳ صحیح می‌باشد.

باید جمع احتمالات برابر یک شود، بنابراین:

$$\sum p(x_i)=1 \Rightarrow \frac{1}{6}+\frac{1}{4}+\frac{1}{3}+a=1 \Rightarrow \boxed{a=\frac{1}{4}}$$

$$E(x)=\Sigma x_i p(x_i) \Rightarrow E(x)=0\left(\frac{1}{6}\right)+1\left(\frac{1}{4}\right)+2\left(\frac{1}{3}\right)+3\left(\frac{1}{4}\right)$$

$$E(x)=\frac{1}{4}+\frac{2}{3}+\frac{3}{4}=\frac{5}{3} \Rightarrow \boxed{E(x)=\frac{5}{3}}$$

○ واریانس (پراش،امیدمجذورانحراف از میانگین،مجذورانحراف ازمعیار،Var (x)، V (x)، D (x)، $(x)\ \sigma^2(x)$)

واریانس متغیر تصادفی گسستهٔ x که میزان پراکندگی را حول میانگین (امیدریاضی) نشان می‌دهد.

$$\boxed{\sigma_x^2=E(x-\mu)^2=E(x^2)-\mu^2}$$

واریانس را برای دو حالت متغیر تصادفی x گسسته و متغیر تصادفی x پیوسته، به صورت زیر محاسبه می‌کنیم:

$\mathrm{var}(x)=E(x^2)-(E(x))^2=\sum x^2 f(x)-\mu_x^2=\sum[x-E(x)]^2 f(x)$: متغیرتصادفی x گسسته و f(x) تابع احتمال گسسته
$\mathrm{var}(x)=E(x^2)-(E(x))^2=\int_{-\infty}^{+\infty}x^2 f(x)dx-\mu_x^2=\int_{-\infty}^{+\infty}[x-E(x)]^2 f(x)dx$:متغیرتصادفیxپیوسته وf(x)تابع احتمال‌پیوسته

مثال۱: اگر میانگین و انحراف معیار x برابر 2 باشد، میانگین x^2 چقدر است؟ (اقتصاد ۸۲)

۱) 4 ۲) 8 ۳) 16 ۴) 32

حل : گزینه ۲ صحیح می‌باشد.

$$\mathrm{var}(x)=E(x^2)-(E(x))^2 \Rightarrow 4=E(x^2)-(2)^2 \to 4=E(x^2)-4 \Rightarrow \boxed{E(x^2)=8}$$

مثال۲: تابع چگالی احتمال x به صورت $f(x)=\begin{cases}\frac{2x}{c^2} & 0\le x\le c\\ 0 & \text{سایرنقاط}\end{cases}$ است. c چه مقدار باشد تا این که $\sigma_x^2=2$ گردد.

۱) $c=2$ ۲) $c=4$ ۳) $c=6$ ۴) $c=9$

حل : گزینه ۳ صحیح می‌باشد.

$$\sigma_x^2 = 2 \Rightarrow E\left(x^2\right) - \left(E(x)\right)^2 = \int_0^c x^2 .\frac{2x}{c^2} - \left(\int_0^c x\frac{2x}{c^2}dx\right)^2 = 2 \rightarrow \left[\frac{2x^4}{4c^2}\right]_0^c - \left(\left[\frac{2x^3}{3c^2}\right]_0^c\right)^2 = 2$$

$$\rightarrow \frac{c^2}{2} - \left(\frac{2}{3}c\right)^2 = 2 \rightarrow \frac{c^2}{2} - \frac{4c^2}{9} = 2 \rightarrow \frac{c^2}{18} = 2 \Rightarrow c^2 = 36 \Rightarrow \boxed{c = 6}$$

مثال۳: در صورتی‌که $x = -1 , 0 , 1$ برای $p(x) = \frac{|x|+1}{5}$ تابع احتمال متغیر تصادفی ناپیوسته x باشد. آن‌گاه امید ریاضی و واریانس x به ترتیب از راست به چپ برابر چیست؟

۱) 0 و $\frac{4}{5}$ ۲) 1 و $\frac{1}{5}$ ۳) 0 و $\frac{2}{5}$ ۴) $\frac{4}{5}$ و $\frac{4}{25}$

حل : گزینه ۱ صحیح می‌باشد.

x	-1	0	1	
$p(x)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\Sigma p(x)=1$

$$E(x) = \Sigma x_i \times p\left(x_i\right) = (-1)\left(\frac{2}{5}\right) + (0)\left(\frac{1}{5}\right) + (1)\left(\frac{2}{5}\right) = 0 \Rightarrow \boxed{E(x) = 0}$$

$$\sigma^2(x) = E(x)^2 - \left(E(x)\right)^2 = \Sigma x_i^2 \times p\left(x_i\right) - \left(\Sigma x_i \times p\left(x_i\right)\right)^2$$

$$= \left[(-1)^2\left(\frac{2}{5}\right) + (0)^2\left(\frac{1}{5}\right) + (1)^2\left(\frac{2}{5}\right)\right] - 0^2 = \frac{4}{5} - 0 = \frac{4}{5} \Rightarrow \boxed{\sigma^2(x) = \frac{4}{5}}$$

● خواص واریانس

اگر a و b ثابت باشند:

$$\boxed{\begin{array}{ll} \sigma^2(\pm ax \pm b) = (\pm a)^2 \sigma^2(x) & \Rightarrow \sigma(\pm ax \pm b) = |\pm a|\sigma(x) \\ \sigma^2(b) = 0 & \Rightarrow \sigma(b) = 0 \end{array}}$$

مثال۱: متغیرتصادفی x دارای میانگین 5 و واریانس 9 می‌باشد. میانگین و واریانس $\frac{x-5}{3}$ به ترتیب (از راست یا چپ) کدام است؟

۱) 0 و 1 ۲) 1 و 0 ۳) 5 و 3 ۴) 5 و 9

حل : گزینه ۱ صحیح می‌باشد.

$$E\left(\frac{x-5}{3}\right) = \frac{E(x)}{3} - \frac{5}{3} = \frac{5}{3} - \frac{5}{3} = 0$$

$$\text{var}\left(\frac{x-5}{3}\right) = \frac{\text{var}(x)}{9} = \frac{9}{9} = 1$$

○ توزیع احتمال توأم ، چگالی احتمال توأم (مشترک)

اگر x و y دو متغیر تصادفی باشند، احتمال پیشامد همزمان این دو، توزیع احتمال توأم x و y نام دارد و آن را با f(x,y) یا $f_{x,y}(x,y)$ نشان می‌دهند. و دارای خواص زیر است:

$$f(x,y) \geq 0 \qquad ;\forall (x,y)$$

الف) گسسته
$$\begin{cases} \sum_{\forall x} \sum_{\forall y} f(x,y)=1 \\ P\{(x,y)\in A\} = \sum\sum_{\forall (x,y)\in A} f(x,y) = f_{x,y}(x,y) \end{cases}$$

ب) پیوسته
$$\begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\,dx\,dy=1 \\ P\{(x,y)\in A\} = \int\int_A f(x,y)\,dx\,dy = f_{x,y}(x,y) \end{cases}$$

مثال: دو متغیر تصادفی ناپیوسته (گسسته) X و Y با قانون توزیع (تابع احتمال‌ها) توأم $P(X=x, Y=y)=f(x,y)$ در نظر گرفته می‌شود، کدام یک از روابط زیر برای توزیع فوق صادق است؟(اقتصاد ۷۱)

۱) $\sum_x f(x,y)=1$ ۲) $\sum_x\sum_y f(x,y)=1$ ۳) $\sum_y f(x,y)=1$ ۴) هیچ‌کدام

حل : گزینه ۲ صحیح می‌باشد.

گزینه ۱، $\sum_x f(x,y)=f(y)$ و گزینه ۳، $\sum_y f(x,y)=f(x)$ می‌باشد.

○ توزیع حاشیه ای یا کناره‌ای *y , x*

تابع توأم یا تابع چگالی احتمال توأم یکی از دو متغیر است، وقتی که تمام ناحیه تغییرات متغیر دیگر، در نظر گرفته شود.

اگر x و y گسسته باشند:

$$\begin{cases} \text{x توزیع حاشیه‌ای} = f(x) = \sum_{\forall y} f(x,y) \\ \text{y توزیع حاشیه‌ای} = f(y) = \sum_{\forall x} f(x,y) \end{cases}$$

اگر x، y پیوسته باشند:

$$\begin{cases} \text{x توزیع حاشیه‌ای} = f(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy \\ \text{y توزیع حاشیه‌ای} = f(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx \end{cases}$$

نتیجه ۱:

$$\boxed{\text{مستقل } X,Y \Leftrightarrow f(x,y)=f(x)\times f(y)}$$

مثال ۱: تابع احتمال زیر را در نظر بگیرید.

X \ Y	1	2
-1	$\frac{1}{4}$	$\frac{1}{4}$
0	$\frac{1}{4}$	$\frac{1}{4}$

کدام گزینه صحیح است؟

۱) دارای ارتباط مثبت هستند

۲) دارای ارتباط غیر خطی هستند

۳) دارای ارتباط منفی هستند

۴) دو متغیر مستقل هستند

حل : گزینه ۴ صحیح می‌باشد.

با توجه به جدول زیر x , y مستقل هستند.

X \ Y	1	2	f(x)
-1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$
f(y)	$\frac{2}{4}$	$\frac{2}{4}$	1

$$f(-1,1)=f_x(-1)f_y(1) \Rightarrow \frac{1}{4}=\frac{2}{4}\times\frac{2}{4}$$

$$f(-1,2)=f_x(-1)f_y(2) \Rightarrow \frac{1}{4}=\frac{2}{4}\times\frac{2}{4}$$

$$f(0,1)=f(0)f(1) \Rightarrow \frac{1}{4}=\frac{2}{4}\times\frac{2}{4}$$

$$f(0,2)=f(0)f(2) \Rightarrow \frac{1}{4}=\frac{2}{4}\times\frac{2}{4}$$

مثال ۲: با توجه به تابع احتمال توزیع توأم X و Y آیا X , Y مستقل‌اند؟

$$P_{XY}(x,y)=\begin{cases}\frac{1}{3} & x,y=(-1,0),(0,1),(1,0)\\ 0 & \text{سایر}\end{cases}$$

با توجه به جدول زیر x,y مستقل نیستند.

y \ x	-1	0	1	f(y)
0	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
1	0	$\frac{1}{3}$	0	$\frac{1}{3}$
f(x)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

$$f(x)=\sum_{\forall y_i} f(x,y_i)$$

$$f(y)=\sum_{\forall x_i} f(x_i,y)$$

توجه کنید که برای استقلال y,x برقراری تساوی $f(x,y)=f_x(x)f_y(y)$ برای تمام xها و yها لازم می‌باشد.

مثال ۳: با توجه به تابع چگالی توأم روبرو آیا x,y مستقل‌اند؟

$$f(x,y)=e^{-(x+y)}\ ,\ 0\le x\ ,\ 0\le y$$

$$f(x)=\int_0^{+\infty} f(x,y)dy=e^{-x}$$

$$f(y)=\int_0^{+\infty} f(x,y)dx=e^{-y} \qquad \Rightarrow \quad e^{-(x+y)}=e^{-x}.e^{-y}$$

با توجه به f(x),f(y) بدست آمده دیده می‌شود که f(x,y)=f(x)f(y) ، $e^{-(x+y)}=e^{-x}.e^{-y}$ در نتیجه x,y مستقل‌اند.

هرگاه حدود x , y به هم مرتبط نباشند و تابع چگالی توأم x , y را بتوان به صورت دو تابع مجزا از x , y تجزیه کرد، x , y مستقل‌اند.

مانند مثال بالا

❍ توزیع احتمال شرطی

توزیع احتمال شرطی متغیر تصادفی x در صورتیکه Y=y باشد، چنین است:	$f_{x\|y}\ (x\mid y)=\dfrac{f(x,y)}{f_Y(y)}\quad ;f_Y(y)>0$
توزیع احتمال شرطی متغیر تصادفی y در صورتیکه X= x باشد، چنین است:	$f_{y\|x}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}\quad ;f_X(x)>0$

مثال ۱: توزیع احتمال مشترک دو متغیر تصادفی ناپیوسته y و x به شرح جدول زیر است:

x \ y	-3	2	4
1	0.1	0.2	0.2
3	0.3	0.1	0.1

به نظر شما $P(x=3\mid y=-3)$ برابر است با؟ (اقتصاد ۸۳)

۱) $\frac{3}{10}$ ۲) $\frac{2}{5}$ ۳) $\frac{3}{5}$ ۴) $\frac{3}{4}$

حل : گزینه ۴ صحیح می‌باشد.

$$P(x=3\mid y=-3)=\frac{P(x=3\ ,\ y=-3)}{P(y=-3)}=\frac{0.3}{0.4}=\frac{3}{4}$$

x \ y	−3	2	4	P(x)
1	0.1	0.2	0.2	0.5
3	0.3	0.1	0.1	0.5
P(y)	0.4	0.3	0.3	1

مثال ۲: اگر در فواصل $0\le y\le 1$ و $0\le x\le 2$ تابع توزیع مشترک این دو متغیر $f(x,y)=\frac{1}{3}(x+y)$ باشد، $f(y\mid x)$ در همین فواصل برابر است با:

۱) $\frac{2+3y}{2}$ ۲) $\frac{1}{3}(x+y^2)$ ۳) $\frac{x+y}{x+\frac{1}{2}}$ ۴) $\frac{x+2y}{3}$

حل : گزینه ۳ صحیح می‌باشد.

$$f(y\mid x)=\frac{f(x\ ,\ y)}{f(x)}=\frac{\frac{x+y}{3}}{\frac{2x+1}{6}}=\frac{2x+2y}{2x+1}=\frac{2(x+y)}{2\left(x+\frac{1}{2}\right)}=\frac{x+y}{x+\frac{1}{2}}$$

$$f(x)=\varphi(x)=\int_0^1\frac{x+y}{3}dy=\frac{xy+\frac{y^2}{2}}{3}\Bigg]_0^1=\frac{2x+1}{6}$$

○ امید شرطی

امید شرطی Y به شرط X=x طبق رابطه زیر تعریف میشود:

$$E\left(Y \mid X=x\right)=\sum_{\forall y} y \frac{f(x,y)}{f_x(x)}$$ گسسته؛ x,y

پیوسته ؛ x,y

$$E\left(Y \mid X=x\right)=\int_{-\infty}^{+\infty} y \frac{f(x,y)}{f_x(x)}\, dy$$

و نیز همین‌طور امید شرطیX به شرط Y=y :

$$E(X \mid Y=y)=\sum_{\forall x} x \frac{f(x,y)}{f_y(y)}$$ گسسته ؛ x,y

$$E(X \mid Y=y) \int_{-\infty}^{+\infty} x \frac{f(x,y)}{f_y(y)}\, dx$$ پیوسته ؛ x,y

مثال۱: توزیع احتمال مشترک دو متغیر تصادفی X و Y توسط جدول زیر بیان شده است. میانگین شرطی Y بر حسب X=6 کدام است؟ (اقتصاد ۸۱)

X \ Y	2	3	4	5	f(x)
4	0.10	0.05	0.03	0.02	0.20
6	0.05	0.20	0.20	0.05	0.50
8	0.05	0.10	0.10	0.05	0.30
f(y)	0. 2	0.35	0.33	0.12	1

۱) 4.2

۲) 3.8

۳) 3.5

۴) 3

حل : گزینه ۳ صحیح می‌باشد.

$$E\left(Y \mid X=6\right)=\frac{\sum y_i \times f\left(X=6, y_i\right)}{f(X=6)}=\frac{(2\times 0.05)+(3\times 0.20)+(4\times 0.20)+(5\times 0.05)}{0.05+0.20+0.20+0.05}=\frac{1.75}{0.5}=3.5$$

مثال۲: توزیع احتمال مشترک دو متغیره X و Y به صورت جدول زیر است:

x \ y	0	1	2
0	0.20	0.15	0.05
1	0.05	0.20	0.05
2	0.05	0.05	0.20

$E(Y \mid X=1)$ برابر است با:

۱) 2 ۲) 1 ۳) 0.5 ۴) 1.5

حل : گزینه ۲ صحیح می‌باشد.

X \ Y	0	1	2	f (x)
0	0.20	0.15	0.05	0.4
1	0.05	0.20	0.05	0.3
2	0.05	0.05	0.20	0.3
f (y)	0.3	0.4	0.3	1

$$E\left(Y \mid X=1\right)=\frac{\sum y_i f\left(X=1\,,\,y_i\right)}{f\left(X=1\right)}=\frac{0\times P\left(X=1,Y=0\right)+1\times P\left(X=1,Y=1\right)+2\times P\left(X=1,Y=2\right)}{0.05+0.2+0.05}$$

$$=\frac{0\times0.05+1\times0.20+2\times0.05}{0.3}=\frac{0.3}{0.3}=1$$

نکته : اگر x, y مستقل از هم باشند، آنگاه:

$$\boxed{x \text{ و } y \text{ مستقل‌اند} \Leftrightarrow \begin{cases} E\left[x \mid y\right]=E[x] \\ E\left[y \mid x\right]=E[y] \end{cases} \qquad \begin{cases} f\left[x \mid y\right]=f[x] \\ f\left[y \mid x\right]=f[y] \end{cases}}$$

$$\boxed{x \text{ و } y \text{ مستقل‌اند} \Rightarrow E(xy)=E(x).E(y)}$$

$$\boxed{E\left[xy\right]=\begin{cases} \sum\sum x\,y f\left(x,y\right) & \text{در حالت گسسته} \\ \iint xy\, f\left(x,y\right)dxdy & \text{در حالت پیوسته} \end{cases}}$$

توجه: امید ریاضی تابعی از دو متغیر تصادفی x و y به صورت k(x,y) که دارای تابع توزیع توأم $F_{x,y}\left(x,y\right)$ است به فرم زیر نشان داده میشود:

$$\boxed{\begin{array}{ll} E\left[k\left(x,y\right)\right]=\sum\limits_{x}\sum\limits_{y} k\left(x,y\right)f\left(x,y\right) & \text{گسسته} \\ E\left[k\left(x,y\right)\right]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} k\left(x,y\right)f\left(x,y\right)dx\,dy & \text{پیوسته} \end{array}}$$

○کوواریانس (هم‌پراش، Cov (x, y) ، $\sigma_{x,y}$، Sp_{xy})

کوواریانس را امیدریاضی تغییرات دو متغیر برحسب میانگینشان تعریف می‌کنیم.

کوواریانس معیاری عددی است که **نوع و جهت رابطهٔ خطی** بین دو متغیر تصادفی را نشان می‌دهد.

$$\boxed{\begin{array}{l} Cov\left(x\,,\,y\right)=E\left(x-E\left(x\right)\right)\left(y-E\left(y\right)\right) \\ Cov\left(x\,,\,y\right)=E\left(x\,y\right)-E\left(x\right)E\left(y\right) \end{array}}$$

توجه:

	رابطهٔ y , x	
با افزایش (کاهش) x, y افزایش (کاهش) یابد	رابطه مستقیم	$\sigma_{xy} = Cov(x,y) > 0$
با افزایش (کاهش) x , y کاهش (افزایش) یابد.	رابطه معکوس	$\sigma_{xy} = Cov(x , y) < 0$
افزایش (کاهش) یک متغیر هیچ تأثیری در دیگری نداشته باشد.	ناهمبسته مستقل ارتباط خطی ندارند	$\sigma_{xy} = Cov(x,y) = 0$

نکته :

$$(x , y)\ \text{مستقل} \Rightarrow \begin{cases} Cov(x , y) = 0 \\ E(xy) - E(x)E(y) = 0 \end{cases}$$

اما این رابطه عکس ندارد یعنی:

$$\begin{cases} Cov(x , y) = 0 \\ E(xy) - E(x)E(y) = 0 \end{cases} \not\Rightarrow (x , y)\ \text{مستقل}$$

● خواص کوواریانس

۱) اگر x , y , z متغیرهای تصادفی و a، b، c، d، e، f اعداد ثابتی باشند، آن‌گاه روابط زیر برقرار است.

1) $Cov(x , y) = Cov(y , x) = \sigma_{xy} = \sigma_{yx}$

2) $Cov(x , x) = var(x) = \sigma_x^2$

3) $Cov(x , a) = Cov(a , x) = 0$

4) $Cov(x + b, y + d) = Cov(x , y)$

5) $Cov(ax , by) = ab\, Cov(x , y)$

6) $Cov(ax + b , cy + d) = ac\, Cov(x , y)$

7) $Cov(ax + by + c , dx + ey + f) = ad\, var(x) + be\, var(y) + ae\, Cov(x , y) + bd\, Cov(x , y)$

8) $Cov(x + y , z) = Cov(x , z) + Cov(y , z)$

9) $Cov(x_1 + x_2 + \ldots + x_k , z) = Cov(x_1 , z) + Cov(x_2 , z) + \ldots + cov(x_k , z)$

10) $Var(ax \pm by + c) = a^2 var(x) + b^2 Var(y) \pm 2ab Cov(x , y)$

11) $Var(ax \pm by \pm cz + d) = a^2 Var(x) + b^2 Var(y) + c^2 Var(Z) \pm 2ab Cov(x,y) \pm 2ac Cov(x,z) \pm 2bc\, Cov(y,z)$

اگر متغیرهای $x_1 , x_2 , \ldots , x_n$ مستقل باشند آن‌گاه:

12) $Var(ax \pm by \pm cZ + d) = a^2 Var(x) + b^2 Var(y) + c^2 Var(Z)$

۲)
$0<Cov(x, y)$: x , y رابطهٔ مستقیم و مثبت دارند

$Cov(x, y) < 0$: x , y رابطهٔ معکوس و منفی دارند

$Cov(x, y) = 0$: x , y ناهمبسته
- x, y از هم مستقل هستند
- x , y اصلاً رابطه خطی با هم نداشته بلکه رابطهٔ غیرخطی دارند.

۳) $\boxed{-\infty < Cov(x, y) < +\infty}$

۴) قبل از محاسبه کوواریانس بهتر است ابتدا مستقل بودن x و y را چک کنیم چون اگر مستقل باشند، $Cov(x, y) = 0$ است.

مثال ۱: با توجه به جدول زیر دربارهٔ استقلال x , y و همچنین $Cov(x, y)$ نظر دهید.

جدول (1) ←

y \ x	0	1	f(y)
0	0.15	0.35	0.5
−1	0.15	0.35	0.5
f(x)	0.3	0.7	1

$$E(x) = \Sigma x.f(x) = (0)(0.3) + (1)(0.7) = 0.7$$

$$E(y) = \Sigma y.f(y) = (0)(0.5) + (-1)(0.5) = -0.5$$

$$E(xy) = \Sigma xy f(x,y) = (0)(0)(0.15) + (0)(1)(0.35) + (-1)(0)(0.15) + (-1)(1)(0.35) = -0.35$$

$$\Rightarrow Cov(x,y) = E(xy) - E(x)E(y) = -0.35 - (0.7)(-0.5) = 0 \rightarrow Cov(x,y) = 0$$

نتیجه: با کمی توجه به جدول (1) در می‌یابید که x , y مستقل هستند چون $f(x,y) = f(x).f(y)$ بنابراین بدون محاسبهٔ کوواریانس، می‌توانستیم نتیجه بگیریم که: $Cov(x,y) = 0$

جدول (2) ←

y \ x	−1	0	1	f(y)
0	0	$\frac{1}{3}$	0	$\frac{1}{3}$
1	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
f(x)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

$$E(x) = \Sigma x.f(x) = (-1)\left(\frac{1}{3}\right) + (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right) = 0$$

$$E(y) = \Sigma y.f(y) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{2}{3}\right) = \frac{2}{3}$$

$$E(xy)=\Sigma xy.f(x,y)=(0)(-1)(0)+(0)(0)\left(\frac{1}{3}\right)+(0)(1)(0)+(1)(-1)\left(\frac{1}{3}\right)+(1)(0)(0)+(1)(1)\left(\frac{1}{3}\right)=-\frac{1}{3}+\frac{1}{3}=0$$

$$\Rightarrow\quad Cov(x,y)=E(xy)-E(x)E(y)=0-(0)\left(\frac{2}{3}\right)=0\Rightarrow cov(x,y)=0$$

نتیجه: با کمی توجه به جدول (2) در می‌یابید که y,x مستقل نیستند چون $f(x,y)\neq f(x).f(y)$ ولی با محاسبهٔ کوواریانس به این نتیجه رسیدیم که $Cov(x,y)=0$ ، بنابراین x , y ارتباط خطی با یکدیگر ندارند.

مثال۲: مقدار کوواریانس تابع احتمال توام زیر چقدر است؟

۱) صفر   ۲) 1   ۳) -0.30   ۴) -0.06

x \ y	0	1
−1	0.30	0.30
0	0.30	0.10

حل : گزینه ۴ صحیح می‌باشد.

ابتدا مستقل بودن x , y را چک می‌کنیم، اگر x و y مستقل باشند $Cov(x,y)=0$ خواهد شد.

چون $f(x,y)\neq f(x).f(y)$، بنابراین x , y مستقل نیستند پس باید کوواریانس را محاسبه کنیم.

x \ y	0	1	f(x)
−1	0.30	0.30	0.60
0	0.30	0.10	0.40
f(y)	0.60	0.40	1

$$E(x)=\Sigma x.f(x)=(-1)(0.60)+(0)(0.40)=-0.60$$

$$E(y)=\Sigma y.f(x)=(0)(0.6)+(1)(0.4)=0.4$$

$$E(xy)=\Sigma xy.f(x,y)=(-1)(0)(0.30)+(-1)(1)(0.30)+(0)(0)(0.30)+(0)(1)(0.10)=-0.30$$

$$\Rightarrow\quad cov(x,y)=E(xy)-E(x)E(y)=-0.30-(-0.60)(0.4)=-0.30+0.24=-0.06$$

مثال۳: اگر متغیرهای تصادفی x , y , z دارای کمیت‌های $\mu_x=2$ و $\mu_y=-3$ و $\mu_Z=4$ و واریانس $\sigma_x^2=1$ ، $\sigma_y^2=2$ ، $\sigma_Z^2=2$ و کوواریانس‌های $Cov(x,y)=1$ و $Cov(x,z)=1$ و $Cov(y,z)=1$ باشند، میانگین و واریانس $w=3x-y+2z$ کدام است؟

۱) 15 , 14   ۲) 17 , 21   ۳) 17 , 18   ۴) 18 , 18

حل : گزینه ۲ صحیح می‌باشد.

$$E(w)=E(3x-y+2z)=E(3x)+E(-y)+E(2z)$$

$$=3E(x)-E(y)+2E(Z)=3\mu_x-\mu_y+2\mu_z=3(2)-(-3)+2(4)$$

$$\Rightarrow\quad E(w)=6+3+8=17\quad\Rightarrow\quad \boxed{E(w)=17}$$

$$Var(w)=Var(3x-y+2z)=9Var(x)+Var(y)+4Var(z)-6Cov(x,y)+12cov(x,z)-4cov(y,z)$$

$$=9(1)+2+4(2)-6(1)+12(1)-4(1)=21$$

مثال۴: در صورت صفر بودن کواریانس x , y کدام گزینه صادق است؟

۱) آن دو متغیر مستقل‌اند.

۲) دارای رابطه غیرخطی می‌باشند.

۳) موارد ۱ و ۲ هر دو صحیح است.

۴) یا دارای رابطه غیرخطی هستند یا هیچ رابطه‌ای با همدیگر ندارند.

حل : گزینه ۴ صحیح می‌باشد.

$$\text{if } Cov(x, y) = 0 \Rightarrow \begin{cases} x, y \text{ دارای رابطه غیرخطی هستند} \\ \text{یا} \\ x, y \text{ هیچ رابطه با همدیگر ندارند (مستقل)} \end{cases}$$

مثال۵: دو متغیر مستقل: (اقتصاد ۷۵)

۱) استقلال خطی ندارد.

۲) کوواریانس آن‌ها مخالف صفر است.

۳) مانع الجمع‌اند.

۴) ناهمبسته‌اند.

حل : گزینه ۴ صحیح می‌باشد.

$$\text{مستقل (ناوابسته)} \quad y, x \Rightarrow \begin{cases} Cov(x, y) = 0 \\ (x, y) \text{ ناهمبسته‌اند.} \end{cases}$$

مثال۶: کدام‌یک از موارد زیر برای $v(x - y)$ صحیح است؟

۱) $v(x) + v(y) - 2Cov(x, y)$

۲) $v(x) - v(y) - Cov(x, y)$

۳) $v(x) + v(y) + 2Cov(x, y)$

۴) $v(x) + v(y) - Cov(x, y)$

حل : گزینه ۱ صحیح می‌باشد.

$$v(x - y) = v(x) + v(y) - 2Cov(x, y)$$

مثال۷: در رابطه با دو متغیر تصادفی x و y کدام‌یک از گزینه‌های زیر غلط است؟

۱) اگر $Cov(x, y) = 0$ باشد، دو متغیر تصادفی y , x از لحاظ آماری مستقل از هم می‌باشند.

۲) اگر $P(x \cap y) = P(x).P(y)$ باشد، y , x از لحاظ آماری مستقل از هم می‌باشند.

۳) اگر $Cov(y, x) = 0$ باشد، وابستگی خطی بین y , x وجود ندارد.

۴) $Cov(x, y)$ برابر $E(xy) - E(x)E(y)$ است.

حل : بنابراین گزینه ۱ صحیح نمی‌باشد.

مثال۸: چنانچه $w = a + bx$ و $z = c + dy$، $cov(z , w)$ برحسب x , y عبارت است از:

۱) $(a + b)(c + d)\,Cov(x , y)$ ۲) $bd\,Cov(x, y)$

۳) $(ab + bd)\,Cov(x , y)$ ۴) $Cov(x , y)$

حل : گزینه ۲ صحیح می‌باشد.

بنابر قاعده پخشی:

$$Cov(z,w) = cov(c + dy , a + bx) = \underset{0}{Cov}\cancel{(c,a)} + \underset{0}{Cov}\cancel{(c,bx)} + \underset{0}{Cov}\cancel{(dy,a)} + cov(dy,bx) = bd\,Cov(x,y)$$

مثال۹: اگر x و y دو متغیر تصادفی مستقل باشند و $E(x) = 3$ و $E(y) = 2$ باشند، کدام‌یک از عبارت‌های زیر صحیح است؟(اقتصاد ۸۰)

۱) $Cov(x , y) = 0$ ۲) $E(x + y) = 5$ ۳) $E(xy) = 6$ ۴) هر سه صحیح است.

حل : گزینه ۴ صحیح می‌باشد.

$$x,y \text{ مستقل} \Rightarrow Cov(x,y) = 0$$

$$E(x + y) = E(x) + E(y) = 3 + 2 = 5$$

$$E(xy) = E(x)E(y) = (3)(2) = 6$$

توجه: اگر x و y مستقل نباشند، فقط گزینه ۲ صحیح است.

مثال۱۰: اگر $\sigma_x^2 = \frac{1}{2}$ و $\sigma_y^2 = \frac{2}{3}$ و $\sigma_{x+y}^2 = \frac{5}{6}$ باشد، مقدار کوواریانس کدام است؟ (حسابداری ۷۷)

۱) $\frac{-1}{2}$ ۲) $\frac{-1}{6}$ ۳) $\frac{1}{6}$ ۴) $\frac{1}{3}$

حل : گزینه ۲ صحیح می‌باشد.

$$\sigma_{(x+y)}^2 = \sigma_x^2 + \sigma_y^2 + 2Cov(x,y)$$

$$\frac{5}{6} = \frac{1}{2} + \frac{2}{3} + 2\,Cov(x,y) \Rightarrow 2Cov(x,y) = -\frac{2}{6} \Rightarrow \boxed{Cov(x,y) = -\frac{1}{6}}$$

مثال۱۱: در مورد دو متغیر تصادفی X و Y و صحت رابطه امید ریاضی $E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}$ می‌توان گفت:

۱) وقتی صادق است که X و Y مستقل باشند.

۲) وقتی صادق است که X و Y کوواریانس صفر داشته باشند

۳) وقتی صادق است که ناهمبسته باشند.

۴) هیچ وقت صادق نیست

حل : گزینه ۴ صحیح می‌باشد.

اگر x , y از هم مستقل باشند، x , $\frac{1}{y}$ نیز مستقل می‌باشند. ← $E\left(\frac{x}{y}\right) = E(x)E\left(\frac{1}{y}\right)$

اما هیچ‌گاه $E\left(\frac{1}{y}\right) = \frac{1}{E(y)}$ نمی‌باشد بنابراین تحت هیچ شرایطی $E\left(\frac{x}{y}\right) = \frac{E(x)}{E(y)}$ صادق نیست.

○ همبستگی

تحلیل همبستگی ابزاری است که به وسیلهٔ آن می‌توان درجه‌ای که یک متغیر به متغیری دیگر، **از نظر خطی**، مرتبط است اندازه‌گیری و برای تعیین میزان ارتباط دو متغیر استفاده می‌شود.

در همبستگی دربارهٔ دو معیار **ضریب همبستگی** و **ضریب تعیین** بحث می‌شود.

○ ضریب همبستگی (Correlation Coefficient)

ضریب همبستگی، درجه وابستگی بین دو متغیر را اندازه‌گیری می‌کند.

اگر به جای x و y مقادیر $\frac{x}{\sigma_x}$ و $\frac{y}{\sigma_y}$ را در فرمول کوواریانس به کار بریم، به معیاری بنام ضریب همبستگی می‌رسیم.

$$\boxed{Cov\left(\frac{x}{\sigma_x}, \frac{y}{\sigma_y}\right) = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = r_{x,y} = \rho_{x,y}}$$

نکته : ضریب همبستگی برای داده‌های آماری به صورت زیر محاسبه می‌شود:

$$\boxed{\frac{cov(x,y)}{\sigma_x \times \sigma_y} = \frac{\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n}}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n} \times \frac{\sum(y_i - \bar{y})^2}{n}}} = \frac{\frac{\sum x_i y_i}{n} - \left(\frac{\sum x_i}{n}\right)\left(\frac{\sum y_i}{n}\right)}{\sqrt{\left(\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2\right)\left(\frac{\sum y_i^2}{n} - \left(\frac{\sum y_i}{n}\right)^2\right)}}}$$

توجه کنید: معیار $Cov(x,y)$ به واحد اندازه‌گیری دو متغیر x و y بستگی دارد اما معیار $Cov\left(\frac{x}{\sigma_x}, \frac{y}{\sigma_y}\right)$ که همان ضریب همبستگی می‌باشد به واحد اندازه‌گیری دو متغیرx و y بستگی ندارد.

مثال : ضریب همبستگی جامعه دو متغیر وابسته y , x برابر است با:

x_i	7	10	4	11
y_i	14	20	8	22

۱) 2　　۲) 1　　۳) $\frac{1}{2}$　　۴) -1

حل : گزینه ۲ صحیح می‌باشد.

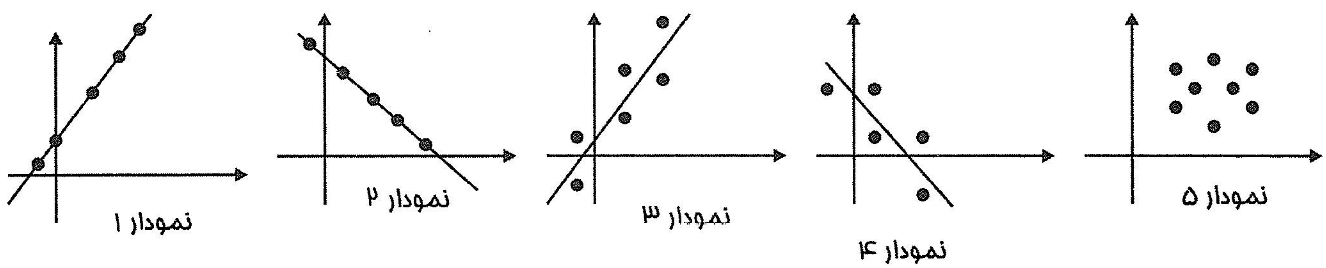
x_i	y_i	x_i^2	y_i^2	$x_i y_i$
7	14	49	196	98
10	20	100	400	200
4	8	16	64	32
11	22	121	484	242
$\sum x_i = 32$	$\sum y_i = 64$	$\sum x_i^2 = 286$	$\sum y_i^2 = 1144$	$\sum x_i y_i = 572$

$$r_{x,y} = \rho_{x,y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \times \sigma_y} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \times \frac{\sum y}{n}}{\sqrt{\left(\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right)\left(\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2\right)}}$$

$$= \frac{\frac{572}{4} - \frac{32}{4} \times \frac{64}{4}}{\sqrt{\left(\frac{286}{4} - \left(\frac{32}{4}\right)^2\right)\left(\frac{1144}{4} - \left(\frac{64}{4}\right)^2\right)}} = \frac{143 - 8 \times 16}{\sqrt{(71.5 - 64)(286 - 256)}} = \frac{15}{\sqrt{225}} = 1$$

• خواص ضریب همبستگی

۱ ـ همواره رابطهٔ $-1 \le \rho \le 1$ برقرار است.



$\rho = 1$	$\rho = -1$	$0 < \rho < 1$	$-1 < \rho < 0$	$\rho = 0$
بین x و y رابطهٔ مستقیم و کامل	بین x و y رابطهٔ معکوس و کامل	بین x و y رابطه مستقیم و ناقص	بین x و y رابطه معکوس و ناقص	x و y رابطه خطی ندارند.

توجه:

x , y مستقل $\Rightarrow \rho = 0$

$\rho = 0 \Rightarrow$
- y , x مستقل
- y,x رابطهٔ خطی ندارند

۲ ـ هر تغییری روی متغیر x و y تغییری در ρ ایجاد نمی‌کند.

$$\rho_{ax\pm b,\, cy\pm d} = \frac{\operatorname{Cov}(ax \pm b,\, cy \pm d)}{\sigma_{(ax\pm b)}\sigma_{(cy\pm d)}} = \frac{\operatorname{Cov}(x,y)}{\sigma_x \sigma_y}$$

- $= \rho_{x,y}$: اگر c و a هم علامت باشند
- $= -\rho_{x,y}$: اگر c و a مختلف العلامه باشند.

$\rho_{x,-x} = -1$

$\rho_{x,x} = 1$

$\rho_{x,a} = 0$

$\rho_{x,y} = \rho_{y,x}$

مثال ۱: به شرط صفربودن کوواریانس دو متغیر تصادفی x و y :

۱) ضریب همبستگی این دو متغیر صفراست.

۲) ضریب همبستگی این دو متغیر مثبت است.

۳) ضریب همبستگی این دو متغیر منفی است.

۴) این دو متغیر مستقل‌اند.

حل : گزینه ۳ صحیح می‌باشد.

$$Cov(x,y)=0 \;\Rightarrow\; \rho_{x,y}=\frac{Cov(x,y)}{\sigma_x.\sigma_y}=0 \;\Rightarrow\; \rho_{x,y}=0$$

مثال ۲: فرض کنیم تابع چگالی احتمال‌های مشترک دو متغیر x و y نرمال باشد، در این صورت اگر کوواریانس x و y صفر باشد؟

۱) ضریب همبستگی x و y مثبت است.

۲) x وy می‌توانند مستقل نباشند.

۳) x و y مستقل از هم هستند.

۴) ضریب همبستگی x و y منفی است.

حل : گزینه ۳ صحیح می‌باشد.

تنها و تنها در توزیع نرمال هرگاه یکی از شرایط * اتفاق بیفتد قطعاً y , x مستقل بوده‌اند اما در شرایط دیگر همان وضعیتی است که قبلاً صحبت کردیم.

$$*\begin{cases}\rho_{xy}=0\\ Cov(x,y)=0\\ E(xy)=E(x)E(y)\end{cases}$$

مثال ۳: اگر $Cov(x,y)=10$ و $\sigma_x=5$ و $\sigma_y=3$ باشد، ضریب همبستگی کدام است؟

۱) $-\frac{1}{2}$

۲) $\frac{1}{2}$

۳) $\frac{2}{3}$

۴) هیچکدام

حل : گزینه ۳ صحیح می‌باشد.

$$\rho_{x,y}=\frac{Cov(x,y)}{\sigma_x.\sigma_y}=\frac{10}{(5)(3)}=\frac{10}{15}=\frac{2}{3}$$

مثال ۴: ضریب همبستگی $-5x+\frac{2}{3}$ و $3x$ برابر است با:

۱) -1

۲) -0.82

۳) 0.82

۴) +1

حل : گزینه ۱ صحیح می‌باشد.

با توجه به خواص ضریب همبستگی داریم:

$$\rho_{3x,-5x+\frac{2}{3}}=\rho_{x,-x}=-1$$

❍ ضریب تعیین (ضریب تشخیص، R^2، r^2)

- ضریب تعیین مهمترین معیاری است که با آن می‌توان رابطهٔ بین دومتغیر x و y را توضیح داد.
- ضریب تعیین، بیان‌کنندهٔ درصد تغییرات تابعی یعنی y به وسیله تغییرات متغیر مستقل x می‌باشد.
- ضریب تعیین معلوم می‌کند که چند درصد از تغییرات y ناشی از تغییرات x است.
- ضریب تعیین R^2 با معلوم بودن ضریب همبستگی $\rho(r)$ برابر است با: $\boxed{R^2=r^2}$

برای تعیین میزان ارتباط و همبستگی میان متغیر مستقل Xو متغیر وابسته Y، به ازاء هر مقدار x_i متعلق به X مقدار متناظر آن y_i متعلق به Y را به‌دست می‌آوریم. در این حالت برای $i = 1, 2, \ldots, n$ نقاط $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ به‌دست می‌آید.

در صورتیکه این نقاط را روی صفحه مختصات مشخص کنیم، از بین همه خطوطی که می‌توانند از بین نقاط عبور کنند فقط یک خط وجود دارد که از بیشتر نقاط از جمله نقطه میانگین $(\bar{X}, \bar{Y})$ عبور کرده و فاصله آن نسبت به بقیه نقاط می‌نیمم می‌باشد، این خط همان خط برازش یا خط رگرسیون به شکل $y = ax + b$ می‌باشد، که در آن y به عنوان متغیر وابسته تابعی از متغیر مستقل x می‌باشد.

نکته : با استفاده از خط رگرسیون می‌توان به ازاء هر مقدار مشخص x مقدار y را پیش‌بینی نمود.

• تشکیل معادله خط رگرسیون $y = ax + b$

برای به‌دست آوردن معادلهٔ خط رگرسیون، بایستی ضرایب معادله خط یعنی a (شیب خط) و ثابت b (عرض از مبدأ) را به شرح زیر محاسبه کنیم.

ابتدا شیب خط (a) را محاسبه کرده، سپس از آنجائی‌که معادله خط رگرسیون همیشه از نقطه $(\bar{x}, \bar{y})$ یعنی نقطه میانگین عبور می‌کند به راحتی می‌توانیم مقدار b (عرض از مبدأ) را از رابطهٔ $\bar{y} = a\bar{x} + b$ بعد از محاسبه a، به‌دست آوریم.

۱) محاسبه شیب خط (a)

برای محاسبه مقدار a (شیب خط) باتوجه به داده‌های مسئله می‌توانیم از روابط زیر استفاده کنیم:

$$\text{شیب خط} = a = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \rho_{x,y} \times \frac{\sigma_y}{\sigma x} = \frac{SP_{xy}}{SS_x} \qquad (I)$$

در صورتیکه در رابطه (I) فرمول‌های آماری $\operatorname{cov}(x, y)$, σ_x^2 را در نظر بگیریم به رابطه (II) می‌رسیم.

$$\text{شیب خط} = a = \frac{\dfrac{\sum xy}{n} - \dfrac{\sum x}{n} \times \dfrac{\sum y}{n}}{\dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2} = \frac{\dfrac{\sum (x - \bar{x})(y - \bar{y})}{n}}{\dfrac{\sum (x - \bar{x})^2}{n}} \qquad (II)$$

نکته ۱: دقت کنید، که در داده‌های آماری روابط زیر صادق هستند:

$$SP_{xy} = \sum (x - \bar{x})(y - \bar{y}) \quad ; \quad SS_x = \sum (x - \bar{x})^2$$

نکته ۲: با توجه به داده‌های مسئله می‌توان از یکی از روابط (I) یا (II) برای محاسبه شیب خط استفاده کرد.

نکته ۳: معادله رگرسیون خطی x بر حسب y به صورت $x = a'y + b'$ می‌باشد که a' شیب معادله و b' عرض از مبدأ است.

$$a' = \frac{\operatorname{Cov}(x, y)}{\sigma_y^2} \quad ; \quad b' = \bar{x} - a'\bar{y}$$

نکته۴: اگر معادله رگرسیون خطی y برحسب x به صورت $y = a + bx$ و معادله رگرسیون خطی x برحسب y به صورت $x = a'y + b'$ باشد رابطه زیر برقرار است:

$$\boxed{r^2 = aa' \Rightarrow \rho = r = \sqrt{aa'}}$$

مثال : در یک جامعه نرمال دو بعدی معادلات رگرسیون y برحسب x و x برحسب y به صورت زیر به دست آمده است. ضریب تعیین $\left(r^2\right)$ کدام است؟

$$\hat{y}_x = 4.85 - 3.2x \text{ ، } \hat{x}_y = 8.32 - 0.28y$$

حل:

$$r^2 = aa' = (-3.2)(-0.28) = 0.896$$

۲) محاسبه عرض از مبدأ (b):

بعد از محاسبه شیب خط (a) باتوجه به آن‌که معادلهٔ خط رگرسیون همیشه از نقطه $(\bar{x}, \bar{y})$ یعنی نقطه میانگین عبور می‌کند، به راحتی می‌توانیم مقدار b (عرض از مبدأ) را از رابطه:

$$\boxed{\bar{y} = a\bar{x} + b}$$

با معلوم بودن مقدار a (شیب خط) و $\bar{x} = \frac{\sum x}{n}$ و $\bar{y} = \frac{\sum y}{n}$ به‌دست می‌آوریم.

نکته : در بعضی مراجع معادله خط رگرسیون به صورت $y = bx + a$ ارائه می‌گردد که در این حالت b (شیب خط) و a (عرض از مبدأ) می‌باشد.

مثال ۱ : اگر $SP_{xy} = 20$ و $SS_x = 20$ و $SS_y = 20$ و $\bar{X} = 5$ و $\bar{y} = 4$ باشد. معادله خط رگرسیون برابر است با: (مدیریت ۷۹)

۱) $y = -x + 1$　　۲) $y = x - 1$　　۳) $y = x + 1$　　۴) $y = \frac{x}{2} + 1$

حل : گزینه ۲ صحیح می‌باشد.

$$\begin{cases} a = \frac{SP_{xy}}{SS_x} = \frac{20}{20} = 1 \quad ① \\ \bar{y} = a\bar{x} + b \to b = 4 - 1 \times 5 = -1 \quad ② \\ y = ax + b \xrightarrow{①,②} y = x - 1 \end{cases}$$

راه تستی: با توجه به این نکته که نقطه $(\bar{x}, \bar{y})$ همیشه از معادله خط رگرسیون برآورد شده می‌گذرد. با قرار دادن مقدار $\bar{x}, \bar{y}$ در معادلات رگرسیون گزینه‌ها به راحتی گزینه صحیح را می‌یابیم در واقع گزینه‌ای که $\bar{x}, \bar{y}$ در معادله آن صدق کند، جواب است.

مثال ۲ : اگر شیب معادله رگرسیون -10 باشد، $\sum x = 100$ و $\bar{X} = 20$ و $\sum y = 20$ باشد، ثابت معادله کدام است؟

۱) 106 ۲) 110 ۳) 220 ۴) 204

گزینه ۴ صحیح می‌باشد.

$$\left\{\begin{array}{l} ① \ \bar{X} = \dfrac{\sum x}{n} \Rightarrow 20 = \dfrac{100}{n} \Rightarrow n = 5 \\ ② \ \bar{Y} = \dfrac{\sum y}{n} = \dfrac{20}{5} = 4 \\ ③ \ a = \text{شیب خط} = -10 \end{array}\right. \xrightarrow{①,②,③} \left\{\begin{array}{l} \bar{y} = a\bar{x} + b \\ 4 = -10 \times 20 + b \rightarrow b = 204 \end{array}\right.$$

مثال ۳ : با استفاده از اطلاعات زیر معادله رگرسیون کدام است؟

$$\sum y = 50 , \sum x = 75 , n = 25 , \sum y^2 = 228 , \sum xy = 30 , \sum x^2 = 625$$

۱) $y = 2.9 - 0.3x$ ۲) $y = 2.9 - 0.15x$ ۳) $y = 5.8 - 0.3x$ ۴) $y = 8.7 - 0.6x$

حل : گزینه ۱ صحیح می‌باشد.

$$\bar{x} = \frac{\sum x}{n} = 3 , \bar{y} = \frac{\sum y}{n} = 2$$

$$(\mathrm{I})\ a = \frac{\dfrac{\sum xy}{n} - \dfrac{\sum x}{n} \times \dfrac{\sum y}{n}}{\dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2} = -0.3$$

$$\xrightarrow{(\mathrm{I}),(\mathrm{II})} \boxed{\begin{array}{l} y = ax + b \\ y = -0.3x + 2.9 \end{array}}$$

$$(\mathrm{II})\ \bar{y} = a\bar{x} + b \Rightarrow 2 = -0.3 \times 3 + b \rightarrow b = 2.9$$

مثال ۴ : شیب خط رگرسیون $y = a + bx$ چقدر است؟

x	0	1	3	4	5
y	0	2	5	9	11

۱) 0.31 ۲) 1.21 ۳) 1.8 ۴) 2.2

حل : گزینه ۴ صحیح می‌باشد.

x^2	x	y	xy
0	0	0	0
1	1	2	2
9	3	5	15
16	4	9	36
25	5	11	55
$\sum x^2 = 51$	$\sum x = 13$	$\sum y = 27$	$\sum xy = 108$

$$b = \text{شیب خط} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \times \frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \frac{\frac{108}{5} - \frac{13}{5} \times \frac{27}{5}}{\frac{51}{5} - \left(\frac{13}{5}\right)^2} = 2.2$$

مثال ۵ : فرض کنید $\text{cov}(x, y) = 12$ و $n = 10$ و $\sum x = \sum y = 50$ و $\sigma_x = 4$ و $\sigma_y = 3$ است. معادله رگرسیون y برحسب x کدام است؟

۱) $y = 1.5 - 0.3\,x$ ۲) $y = 1.25 + 0.75\,x$ ۳) $y = 1.5 + 0.4\,x$ ۴) $y = 3 + 2.2\,x$

گزینه ۲ صحیح می‌باشد.

$$\begin{cases} \bar{x} = \frac{\sum x}{n} = 5 , \bar{y} = \frac{\sum y}{n} = 5 \\ (\mathrm{I})\, a = \text{شیب خط} = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{12}{16} = \frac{3}{4} = 0.75 \qquad \xrightarrow{(\mathrm{I}),(\mathrm{II})} \begin{array}{l} y = ax + b \\ y = 0.75\,x + 1.25 \end{array} \\ (\mathrm{II})\, \bar{y} = a\bar{x} + b \rightarrow 5 = 0.75 \times 5 + b \rightarrow b = 1.25 \end{cases}$$

نکته۱: علامت شیب خط معادله رگرسیون $y = ax + b$ یعنی علامت a همان علامت کوواریانس $(\text{cov}(x, y))$ و ضریب همبستگی $(\rho_{x,y})$ می‌باشد، در نتیجه با داشتن معادله خط می‌توانیم از روی علامت شیب خط جهت ارتباط y , x را به‌دست آوریم.

مثال۶: اگر از معادله خط رگرسیون برآوردی به صورت $\hat{y} = 2.4 - 0.6x$ به دست آمده و ضریب تعیین 0.49 باشد، ضریب همبستگی کدام است؟ (اقتصاد ۸۲)

۱) $+0.7$ ۲) -0.49 ۳) ± 0.7 ۴) -0.7

حل : گزینه ۴ صحیح می‌باشد.

از آن‌جایی که علامت شیب خط رگرسیون منفی است (0.6-) و با توجه به نکته فوق علامت شیب خط رگرسیون و ضریب همبستگی یکسان می‌باشد. بنابراین:

$$r^2 = 0.49 \Rightarrow r = \sqrt{r^2} = \begin{cases} -0.7 & \text{قابل قبول} \\ +0.7 \end{cases}$$