

# جزوه آمار قسمت آخر:

انواع توزیع احتمال و رابطه بین

متغیرها + نمونه سوال

(ب) توزیع پواسون Poison distribution

توزیع احتمال متغیرهای پیوسته :

(الف) توزیع نرمال یا توزیع استاندارد Normal distribution

(ب) توزیع دانشجویی Student's distribution

**توزیع نرمال : Normal distribution**

مهمترین توزیع پیوسته، توزیع نرمال است و این توزیع یکی از کاربردی ترین توزیع های آماری بوده و بیشتر در خصوص متغیر های کمی پیوسته مطرح می باشد.

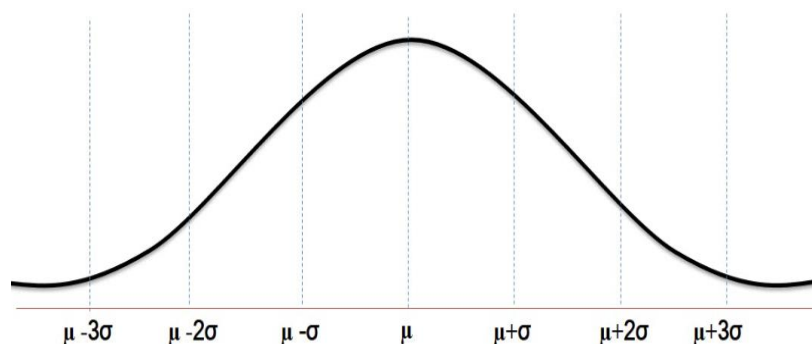
- بسیاری از پدیده های طبیعی دارای این نوع توزیع هستند.

- شکل حدی بسیاری از توزیع های دیگر نرمال است.

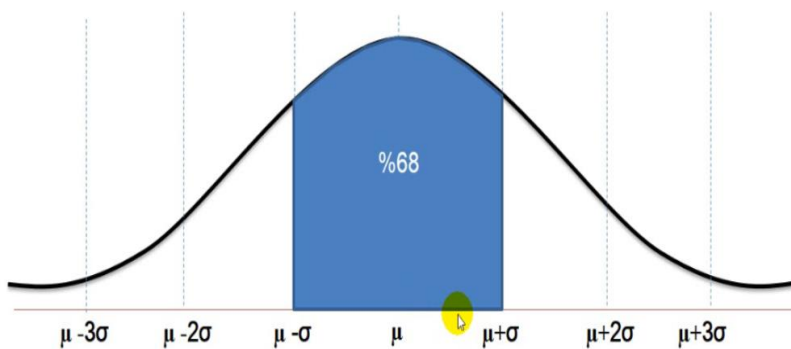
متغیر تصادفی پیوسته  $X$  با میانگین  $\mu$  و انحراف معیار  $\sigma$  دارای توزیع نرمال است اگر تابع چگالی آن به صورت زیر باشد:

$$X \approx N(\mu, \sigma^2)$$

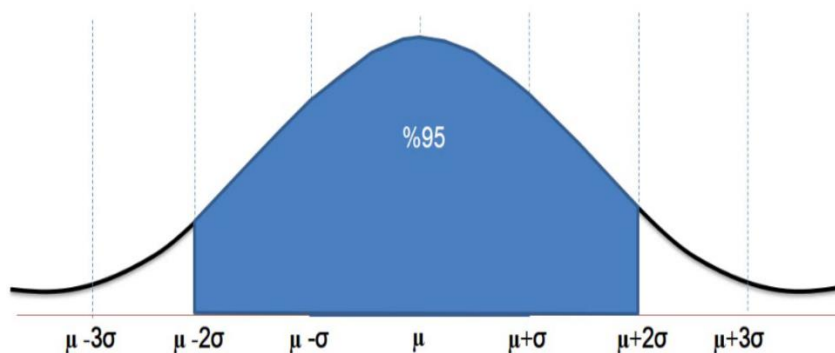
همان گونه که در فرمول توزیع نرمال مشخص است پارامترهای انحراف معیار و میانگین متغیر بوده و سایر عناصر ثابت اند. بر مبنای میانگین ها و واریانس های مختلف می توانیم بی نهایت توزیع نرمال داشته باشیم. با توجه به قرینه بودن منحنی توزیع نرمال و مشخص نمودن انحراف معیار و میانگین می توان از قاعده تجربی استفاده کنیم.



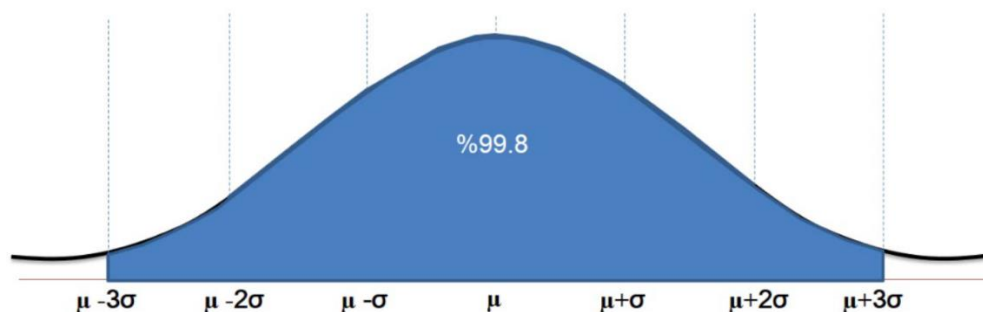
طبق این قاعده اگر فاصله های ۱، ۲ و ۳ انحراف معیار از میانگین را در طرفین میانگین روی منحنی مشخص کنیم می توان عنوان کرد که ۶۸٪ داده ها در فاصله یک انحراف معیار از میانگین ( $\mu \pm 1\sigma$ ) از طرفین واقع می شوند.



به همین ترتیب می توان عنوان کرد که حدود ۹۵٪ داده ها در فاصله دو انحراف معیار از میانگین ( $\mu \pm 2\sigma$ ) از طرفین واقع می شوند.



اگر فاصله را بیشتر کنیم یعنی از  $\mu - 3\sigma$  تا  $\mu + 3\sigma$  حدود ۹۹/۸٪ داده ها را پوشش خواهد داد. به عبارتی که داده هایی که خارج از  $\mu \pm 3\sigma$  باشند تعدادشان خیلی اندک و حدود ۰/۲ درصد است.



### خواص توزیع نرمال :

- ۱- این توزیع دارای منحنی متقارن به شکل زنگوله ای و متقارن می باشد (منحنی نسبت به خط  $X = \mu$  متقارن است) بدین معنا که اگر از وسط آن را تا کنیم سمت چپ و راست آن بر هم منطبق می باشد.
- ۲- میانگین، میانه و مد توزیع نرمال با هم برابرند.
- ۳- منحنی نرمال تنها دارای یک نقطه ماکزیمم در نقطه  $X = \mu$  است.
- ۴- منحنی توزیع نرمال با دو پارامتر میانگین و واریانس مشخص می شود.
- ۵- سطح زیر منحنی بالای محور  $X$ ها برابر با ۱ است.
- ۶- به ازای تمامی  $X$ ها مقدار  $f(x)$  بزرگتر و مساوی صفر است.

۷- با وجود اینکه دامنه تغییرات متغیر تصادفی  $X$  از  $-\infty$  تا  $+\infty$  است ولی تقریباً تمام مقادیر  $X$  در بازه  $\mu \pm 3\sigma$  یعنی بازه ای با طول سه برابر انحراف معیار در دو طرف میانگین قرار می گیرد و احتمال اینکه  $X$  مقداری خارج از این بازه داشته باشد بینهایت کم است.

### توزیع نرمال استاندارد یا توزیع $Z$ :

اگر متغیر تصادفی دلخواه  $X$  دارای توزیع نرمال با میانگین  $\mu$  و واریانس  $\sigma^2$  باشد، آنگاه متغیر تصادفی  $Z = \frac{X-\mu}{\sigma}$  دارای توزیع نرمال با میانگین ۰ و واریانس ۱ است که به آن **نرمال استاندارد** می گویند. لذا به منظور

تسهیل در محاسبه احتمالات مربوط به متغیر تصادفی نرمال آن را تبدیل به متغیر استاندارد می کنیم که میانگین آن صفر و انحراف معیار آن برابر یک است و آن را  $Z$  می نامیم. بنابراین داریم:

$$X \approx N(\mu = 0, \sigma = 1) \quad Z = \frac{X-\mu}{\sigma}$$

$Z$  را متغیر تصادفی نرمال استاندارد و توزیع آن را توزیع نرمال استاندارد گویند.

### توزیع دو جمله ای (Binomial distribution):

این توزیع برای پیش آمدهایی استفاده می شود که دارای دو حالت وقوع و عدم وقوع (موفقیت و شکست) هستند. اگر احتمال وقوع و عدم وقوع پیش آمدی را به ترتیب با  $p$  و  $q$  نشان دهیم و  $p+q=1$  باشد؛ چنانچه بخواهیم احتمال وقوع  $X$  بار موفقیت یا پیروزی از این پیش آمد را در  $n$  مرتبه آزمایش محاسبه کنیم؛ از رابطه ی زیر استفاده می کنیم:

$$P(X=x) = C_x^n p^x q^{n-x} \quad p + q = 1 \quad q = 1-p$$

$$C_x^n = \frac{n!}{x!(n-x)!}$$

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$n$  : تعداد افراد نمونه،  $X$  : تعداد پیشامد های موفق،  $P$  : احتمال وقوع پیشامد مورد نظر،  $q$  : احتمال وقوع پیشامد دیگر (بعدی).

نکته : در توزیع دوجمله ای اگر مقدار  $n$  و  $p$  تغییر کند، شکل توزیع احتمال دو جمله ای نیز تغییر می کند. مثلاً اگر  $n$  یعنی تعداد آزمایش بزرگ شود شکل توزیع دوجمله ای به سمت منحنی نرمال زنگوله ای شکل شباهت پیدا می کند. از طرف دیگر هر چه احتمال موفقیت یا  $p$  به  $0.5$  نزدیک تر شود توزیع قرینه تر خواهد شد.

**مثال ۱:** یک نمونه ۴ تایی بذر از یک توده ۵۰۰۰ هزارتایی بذر انتخاب می شود اگر فرض شود که ۲۰ درصد بذرهای این توده قوه نامید نداشته باشند احتمال اینکه نمونه شامل دقیقاً ۲ بذر بدون قوه نامید باشد را مقایسه کنید.

$$n=4 \quad p = \frac{20}{100} = 0/20$$

$$q=1-0/2=0/8 \quad x=2 \quad P(X=2)=?$$

$$P(x) = C_x^n p^x q^{n-x}$$

$$\begin{aligned} P(X=2) &= \frac{4!}{2!(4-2)!} \times (0/2)^2 \times (0/8)^{4-2} = \frac{4 \times 3 \times 2!}{2! \times 2 \times 1} \times (0/04) \times (0/64) \\ &= 6(0/04)(0/64) = 0/1536 \end{aligned}$$

مثال ۲: احتمال اینکه در ۸ مرتبه پرتاب یک تاس دقیقاً ۳ مرتبه عدد ۵ بیاید را محاسبه کنید.

$$n=8 \quad x=3 \quad P(X=3)=? \quad p = \frac{1}{6} \quad S = \{1, 2, 3, 4, 5, 6\}$$

$$p + q = 1 \Rightarrow \frac{1}{6} + q = 1 \Rightarrow 1 - \frac{1}{6} = \frac{5}{6}$$

$$\begin{aligned} P(X=3) &= \frac{n!}{X!(n-X)!} p^X q^{n-X} = \frac{8!}{3!(8-3)!} \times \left(\frac{1}{6}\right)^3 \times \left(\frac{5}{6}\right)^5 \\ &= 56 \times \left(\frac{1}{216}\right) \times \frac{3125}{7776} = 0/1035 \end{aligned}$$

مثال ۳: از ۸۰۰ خانواده ۵ فرزنده، در چند خانواده انتظار دارید که الف) ۳ پسر ب) ۵ دختر ج) ۲ یا ۳ پسر داشته باشند؟

الف- احتمال ۳ پسر در یک خانواده ی ۵ فرزنده:

$$\begin{aligned} P(X=3) &= \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{5!}{3!(5-3)!} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{5-3} \\ &= \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} \times \frac{1}{8} \times \frac{1}{4} = \frac{10}{32} \end{aligned}$$

$$\text{احتمال ۳ پسر در ۸۰۰ خانواده ی ۵ فرزنده} = \frac{10}{32} (800) = 250$$

ب - احتمال ۵ دختر در یک خانواده ی ۵ فرزنده:

$$\begin{aligned} P(X=5) &= \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{5!}{5!(5-5)!} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{5-5} = 1 \times \frac{1}{32} \times 1 \\ &= \frac{1}{32} \end{aligned}$$

$$\text{احتمال ۵ دختر در ۸۰۰ خانواده ی ۵ فرزنده} = \frac{1}{32} (800) = 25$$

ج - احتمال ۲ یا ۳ پسر در یک خانواده ی ۵ فرزنده:

$$\text{احتمال ۳ پسر در یک خانواده ی ۵ فرزنده (از قبل محاسبه شده)} = \frac{10}{32}$$

د- احتمال ۲ پسر در یک خانواده ی ۵ فرزنده:

$$\begin{aligned} P(X=2) &= \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{5!}{2!(5-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{5-2} \\ &= \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 3 \times 2 \times 1} \times \frac{1}{4} \times \frac{1}{8} = \frac{10}{32} \end{aligned}$$

$$\text{احتمال ۲ یا ۳ پسر در یک خانواده ی ۵ فرزنده} = \frac{10}{32} + \frac{10}{32} = \frac{20}{32}$$

$$500 = 800 \times \frac{20}{32} \text{ احتمال } 2 \text{ یا } 3 \text{ پسر در یک خانواده ی } 5 \text{ فرزند}$$

مثال ۴: در جعبه ای ۱۰۰ لامپ داریم که همه یک شکل هستند اما ۹۰ تا سالم و ۱۰ تا سالم داریم. ۵ لامپ به طور تصادفی از جعبه خارج می شود الف- احتمال اینکه دقیقاً دو لامپ سوخته باشند را محاسبه کنید.

$$P = \text{احتمال موفقیت} = \text{احتمال لامپ های سوخته} = \frac{10}{100} = 0/10$$

$$q = \text{احتمال عدم موفقیت} = \text{احتمال لامپ های سالم} = \frac{90}{100} = 0/90$$

$$\begin{aligned} P(X=2) &= C_2^5 p^2 q^{5-2} = \frac{5!}{2!(5-2)!} p^2 q^3 = \frac{5!}{2!(5-2)!} (0/1)^2 (0/9)^{5-2} \\ &= \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 3 \times 2 \times 1} \times 0/01 \times 0/729 = 0/0729 \end{aligned}$$

ب- در صورتی که موفقیت داشتن ۴ لامپ سوخته باشد احتمال آن را محاسبه کنید.

$$\begin{aligned} P(X=4) &= C_4^5 p^4 q^{5-4} = \frac{5!}{4!(5-4)!} p^4 q^1 = 5 (0/1)^4 (0/9)^1 \\ &= 5 \times 0/01 \times 0/9 = 0/0045 \end{aligned}$$

تمرین ۱: احتمال آنکه در ۱۰ بار پرتاب سکه ۶ بار شیر مشاهده شود چقدر خواهد بود؟

$$P(X=6) = \binom{10}{6} (0.5)^6 \times (0.5)^{10-6} = 210 \times \left(\frac{1}{2}\right)^{10} = 0.205$$

میانگین و انحراف معیار توزیع دو جمله ای :

$$\sigma = \sqrt{n \cdot p \cdot q} \quad \text{انحراف معیار:}$$

$$\text{Var}(X) = \sigma^2 = n \cdot p \cdot q \quad \text{واریانس:}$$

$$E(X) = \mu = n \cdot p \quad \text{میانگین:}$$

مثال ۵: احتمال این که یک قلم کالای تولید شده توسط کارخانه ای معیوب باشد ۰.۰۲ می باشد محموله ای شامل

۱۰ هزار قلم کالا روزانه انبار شده است میانگین و انحراف معیار این توزیع چقدر می باشد ؟

$$P = \frac{2}{100} \quad p + q = 1 \quad q = 1 - p$$

$$q = 1 - 0/02 = 0/98 \quad n = 10000$$

$$\mu = n \cdot p = 10000 \times 0/02 = 200$$

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{10000 \times 0/02 \times 0/98} = 13.85 \approx 14$$

$$\sigma^2 = n \times p \times q = 10000 \times 0/02 \times 0/98 = 196$$

بسط توزیع دو جمله ای :

$$= p^n + C_1^n p^{n-1}q + C_2^n p^{n-2}q^2 + \dots + C_x^n p^x q^{n-x} + C_1^n p q^{n-1} + q^n$$

معادله بالا از سه قسمت تشکیل شده است:

- $p$  و توان آن که از مقدار  $n$  شروع شده از جمله های بعدی به ترتیب از توان آن کاسته می شود از  $n-1$

تا 1

- $q$  و توان آن که از مقدار  $(n-1)$  شروع می شود برعکس  $p$  در جمله های بعدی به توان آن اضافه می

شود.

- مجموع توان های  $p$  و  $q$  باید برابر با  $n$  باشد.

- ضریب هر جمله که یک عدد صحیح می باشد و به دو روش می توان آن را محاسبه کرد:

1 - استفاده از فرمول های زیر:

$$1) \frac{\text{ضریب جمله ما قبل } \times \text{توان در حال کاهش (توان } p\text{)}}{\text{1+توان در حال افزایش (توان } q\text{)}}$$

$$2) \frac{\text{ضریب جمله } \times \text{توان در حال افزایش (توان } p\text{)}}{\text{شماره جمله}}$$

2 - استفاده از فرمول ترکیب :

$$C_x^n = \frac{n!}{x!(n-x)!}$$

مثال ۱ : توزیع دو جمله ای  $(p+q)^5$  را بسط نمائید.

$$(p+q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$$

$$C_1^5 = \frac{5!}{1!(5-1)!} = \rightarrow \frac{5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1} = 5$$

$$(2) \frac{1 \times 5}{0+1} = 5 \quad \text{ضریب جمله ۱}$$

$$C_2^5 = \frac{5!}{2!(5-2)!} = \rightarrow \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 3 \times 2 \times 1} = \frac{20}{2} = 10$$

$$(2) \frac{5 \times 4}{1+1} = \frac{20}{2} = 10 \quad \text{ضریب جمله ۲}$$

$$C_3^5 = \frac{5!}{3!(5-3)!} = \rightarrow \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} = \frac{20}{2} = 10$$

$$(2) \frac{10 \times 3}{2+1} = \frac{30}{3} = 10 \quad \text{ضریب جمله ۳}$$

$$C_4^5 = \frac{5!}{4!(5-4)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1} = \frac{5}{1} = 5$$

$$(2) \frac{10 \times 2}{3+1} = \frac{20}{4} = 5 \quad \text{ضریب جمله ۵}$$

$$(p+q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4p^1q^3 + q^4$$

$$= \frac{4 \times 3}{1+1} = \frac{12}{2} = 6 \gg \frac{6 \times 2}{2+1} = \frac{12}{3} = 4$$

مثال ۲: در بسط دو جمله ای  $(p+q)^6$  عبارت بعد از  $15p^4q^2$  را محاسبه کنید.

$$(p+q)^6 = \dots + 15p^4q^2 + 20p^3q^3$$

$$\text{ضریب جمله } \frac{15 \times 4}{2+1} = 20$$

مثال ۳: با استفاده از فرمول باینومیال حاصل عبارت  $(a^2 - 3b)^4$  را بدست آورید.

$$(a^2 - 3b)^4 = (p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4p^1q^3 + q^4 = (a^2)^4 + 4(a^2)^3(-3b) + 6(a^2)^2(-3b)^2 + 4(a^2)(-3b)^3 + (-3b)^4 = a^8 - 12a^6b + 54a^4b^2 - 108a^2b^3 + 81b^4$$

تمرین ۱: اگر  $p = -2x$  و  $q = 3y$  باشد حاصل توزیع دو جمله ای  $(P+q)^6$  را بدست آورید.

$$(p+q)^6 = p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + q^6$$

$$(-2x)^6 + 6(-2x)^5(3y) + 15(-2x)^4(3y)^2 + 20(-2x)^3(3y)^3 + 15(-2x)^2(3y)^4 + (3y)^6 = (-2x)^6 - (12x)^5(3y) + (30x)^4(3y)^2 - (40x)^3(3y)^3 + (30x)^2(3y)^4 + (3y)^6$$

تمرین ۲: تیراندازی با احتمال  $0/7$  هدف خود را مورد اصابت قرار می دهد. چه قدر احتمال دارد که از سه شلیک، دو شلیک به هدف بخورد؟

$$P(X = 2) = C_2^3 p^2 q^1 = \frac{3!}{2!(3-2)!} p^2 q = 3 (0/7)^2 (0/3)^1 = 3 \times 0/49 \times 0/3 = 0/49$$

تمرین ۴: اگر  $p = x$  و  $q = y^2$  باشد با استفاده از فرمول توزیع دو جمله ای حاصل عبارت  $(p+q)^5$  را محاسبه نمایید.

$$(p+q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$$

$$(x + y^2)^5 = x^5 + 5x^4y^2 + 10x^3(y^2)^2 + 10x^2(y^2)^3 + 5x(y^2)^4 + (y^2)^5$$

$$= x^5 + 5x^4y^2 + 10x^3y^4 + 10x^2y^6 + 5xy^8 + y^{10}$$

### توزیع پواسون یا پویسان :

وقتی که در توزیع دو جمله ای  $n$  بزرگ باشد محاسبه احتمال وقوع پیشامدها مشکل خواهد بود. اگر مقدار  $(np)\lambda$  کوچک تر از ۵ باشد، پیش آمد نادر تلقی می شود و برای محاسبه میزان احتمال وقوع پیش آمدهای نادر که در آنها  $n$  بزرگ و بیش از ۵۰ عضو داشته باشد از توزیع پواسن استفاده می شود که یک حد از توزیع دو جمله ای می باشد:

پس دو شرط لازم برای استفاده از توزیع پواسون عبارتنداز:

$$\lambda < 5 \quad (1)$$

$$N > 50 \quad (2)$$

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{\lambda^x}{x! e^\lambda}$$

پایه لگاریتم طبیعی  $e = 2/71828$

$X$ : تعداد پیشامدهای ناسازگار در فاصله زمانی یا مکانی.

e: عدد نپر (Napier's) (پایه لگاریتم طبیعی) که برابر با ۲/۷۱۸۲۸ است.

$\lambda$ : میانگین تعداد پیشامدهای ناسازگار در فاصله زمانی یا مکانی.

$$\lambda = n.p$$

✓ واریانس و میانگین در توزیع پواسون برابر است با مقدار  $\lambda$ :

$$\sigma^2 = \lambda$$

$$\lambda = \mu$$

مثال ۱: در کارخانه ای ۵۰۰ ماشین کار می کنند احتمال اینکه در یک روز معین از سال ۵ ماشین به علت نقص فنی از کار بیافتد چقدر می باشد؟

$$n=500 \quad p = \frac{1}{365} \quad x = 5$$

$$\lambda = n.p \quad 500 \times \frac{1}{365} = 1/36 \approx 1/4$$

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(1/4)^5 (2/72)^{-1/4}}{5!} = \frac{(1/4)^5}{5! (2/72)^{1/4}} = \frac{5/37}{120 \times 4/058} = \frac{5/37}{487/05} = 0/1010$$

مثال ۲: میزان مرگ و میر ناشی از یک بیماری خاص برابر با ۵ مرگ در هر ۱۰۰۰ مورد ابتلا به آن بیماری می باشد در یک گروه شامل ۳۶۰ مورد از این بیماری احتمال اینکه ۳ مرگ اتفاق بیافتد را بدست آورید.

$$n=360 \quad p=0/005 \rightarrow \frac{5}{1000}$$

$$\lambda = n.p \rightarrow 360 \times \frac{5}{1000} = 1/8$$

$$P(X=3) = \frac{\lambda^x}{x! e^\lambda} = \frac{(1/8)^3}{3! \times (2/72)^{1/8}} = \frac{5/83}{6 \times 6/056} = 0/160$$

مثال ۳: در یک بانک به طور متوسط ۷ نفر حساب خود را در یک سال (۳۶۵ روز) مسدود می کنند فرض کنید تعداد کسانی که در یک بازه زمانی خاص حساب خود را مسدود می کنند دارای توزیع پواسون است.

الف) احتمال اینکه در یک دوره ۴ ماه هیچ کس حساب خود را نبندد؟  $p(x=0)$

$$\lambda: \quad \frac{12}{4} \quad \frac{7}{\lambda} \Rightarrow \lambda = \frac{28}{12}$$

$$P(x=0) = \frac{\left(\frac{28}{12}\right)^0 (2/72)^{-28/12}}{0!} = \frac{1}{1 \times (2/72)^{28/12}} = 0/100$$

ب) احتمال اینکه در یک سال حداقل ۳ نفر حساب خود را ببندد را محاسبه کنید؟  $p(x \geq 3)$

$$P(x \geq 3) = 1 - P(x \leq 2)$$

$$1 - [P(x=2) + P(x=1) + P(x=0)] = 1 - \left[ \frac{e^{-\lambda} \lambda^2}{2!} + \frac{e^{-\lambda} \lambda^1}{1!} + \frac{e^{-\lambda} \lambda^0}{0!} \right] = 1 - \left[ \frac{\left(\frac{28}{12}\right)^2}{2 \times (2/72)^{28/12}} + \right.$$

$$\left. \frac{\left(\frac{28}{12}\right)^1}{1 \times (2/72)^{28/12}} + \frac{1}{1 \times (2/72)^{28/12}} \right] =$$

$$1 - \left[ \frac{5/43}{2 \times 10/29} + \frac{2/33}{10/29} + \frac{1}{10/29} \right] = 1 - [0/26 + 0/23 + 0/10] = 0/41$$

تمرین ۱: به طور متوسط روزی ۲ نفر به یک شرکت بیمه مراجعه می کنند؟ احتمال اینکه در یک روز:

الف) دقیقا یک نفر به شرکت بیمه مراجعه چقدر می باشد؟

$$P(x = 1) = \frac{(2)^1 (2.72)^{-2}}{1!} = \frac{2}{(2.72)^2} = 0/27$$

ب) کسی به شرکت بیمه مراجعه نکند چقدر است؟

$$P(x = 0) = \frac{(2)^0 (2.72)^{-2}}{0!} = \frac{1}{(2.72)^2} = 0/13$$

ج) دقیقا ۳ نفر به شرکت بیمه مراجعه کنند؟

$$P(x = 3) = \frac{(2)^3 (2.72)^{-2}}{3!} = \frac{8}{6 \times (2.72)^2} = 0/18$$

خطای استاندارد میانگین نمونه ها :

انحراف معیار توزیع میانگین نمونه ها را خطای معیار یا خطای استاندارد گویند. اگر انحراف معیار جامعه معلوم باشد :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

اگر انحراف معیار جامعه معلوم نباشد (نمونه داشته باشیم):

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

واریانس نمونه را طبق فرمول های کاربردی  $s^2_{\bar{x}} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$  یا تئوری  $s^2_{\bar{x}} = \frac{(x-\bar{x})^2}{n-1}$  بدست آورده و سپس

از آن جذر گرفت و انحراف معیار نمونه بدست می آید:

میانگین واریانس نمونه :

$$s^2_{\bar{x}} = \frac{s^2 x}{n}$$

مثال ۱ :

خطای معیار اعداد ۳ و ۴ و ۷ و ۱۰ را محاسبه نمایید.

$$s^2_{\bar{x}} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{174 - \frac{576}{4}}{4-1} = \frac{30}{3} = 10$$

$$\sum x^2 = 174 \quad (\sum x)^2 = (24)^2 = 576$$

$$s_{\bar{x}} = \frac{\sqrt{10}}{\sqrt{4}} = \frac{3.16}{2} = 1.58 \text{ خطای معیار}$$

یعنی برای نمونه هایی با اندازه  $n=4$  توزیع میانگین نمونه ها دارای میانگین ۶ خواهد بود و انحراف معیار این توزیع برابر ۱,۵۸

### رابطه بین متغیرها:

در جامعه صفاتی وجود دارند که به نحوی با دیگر خصوصیات ارتباط دارند. به عنوان مثال دمای هوا با میزان بارندگی، عملکرد گندم با مقدار آب آبیاری و عملکرد گندم با میزان کوددهی ارتباط دارد. در این مثال ها متغیرهای موردنظر مربوط به دو پدیده مختلف هستند. همین طور وزن انسان بالغ با قد او، محیط دایره با شعاع آن و فشار یک توده گاز با درجه حرارت آن مرتبط است. در این مثال ها متغیرهای موردنظر صفات، حالات و جنبه های مختلف یک فرد یا یک پدیده را نشان می دهند و هر دو تحت تاثیر عوامل مختلفی قرار می گیرند و ارتباط دو طرفه دارند که در مورد مثال های قبلی صادق نیست. چون عملکرد گندم و یا درصد جوانه زدن نه تنها اثری بر آبیاری ندارند، بلکه با آنها تحت تاثیر عوامل مشترک نیز نمی باشند.

ارتباط بین متغیرها را می توان به صورت معادله ریاضی نشان داد، در این صورت باید یکی از متغیرها را به عنوان  $X$  و دیگری را به عنوان  $Y$  نامگذاری نمود. معمولاً  $X$  را متغیر ثابت یا مستقل و  $Y$  را به عنوان متغیر تابع (وابسته) در نظر می گیرند. متغیر  $Y$  متغیری است تصادفی که دارای توزیع فراوانی نرمال می باشد و وقوع مقادیر مختلف آن تحت تاثیر عوامل ناشناخته، عوامل کنترل نشده و خطای نمونه برداری (sampling error) نیز می باشد. متغیر  $X$  دارای مقادیر ثابت است که از قبل توسط محقق تعیین می شوند و در مورد آن هیچ فرضی در مورد تصادفی بودن و توزیع نرمال نداشتن وجود ندارد. به طور کلی برای تعیین وجود و یا عدم وجود رابطه بین متغیرها از روش های رگرسیون و همبستگی استفاده می شود. از روش همبستگی برای تعیین رابطه بین دو متغیر تصادفی استفاده می شود و از روش رگرسیون برای پی بردن به وجود یا عدم وجود رابطه بین یک متغیر وابسته و یک متغیر ثابت و همچنین برای تعیین میزان تغییر یک متغیر به ازای هر واحد تغییر از متغیر دیگر (چه ثابت و چه تصادفی) از روش رگرسیون استفاده می شود.

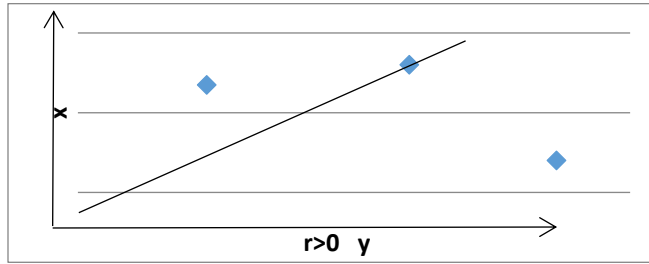
همبستگی یا کورلاسیون ( $r = \text{correlation}$ )

ضریب همبستگی شدت وابستگی و ارتباط دو متغیر تصادفی را نشان می دهد و یکی از متغیرها تابع دیگری نیست و تنها ممکن است با هم رابطه داشته باشند. مثل رابطه طول خوشه با عملکرد تعداد دانه در خوشه. هر چه ضریب همبستگی به ۱ یا -۱ نزدیکتر باشد، شدت رابطه خطی بین متغیرهای موردنظر شدیدتر است.

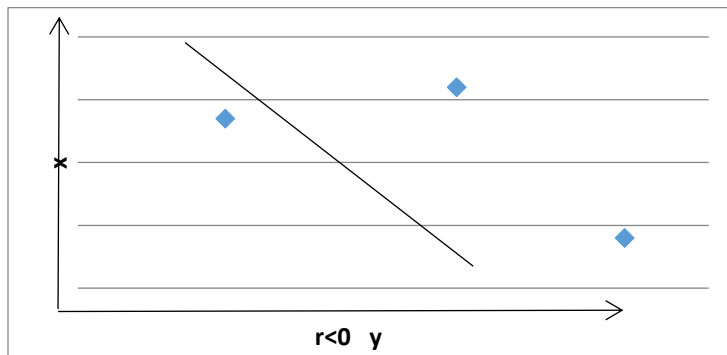
### ویژگی ضریب همبستگی:

۱. مقدار آن بین -۱ و +۱ تغییر می کند.

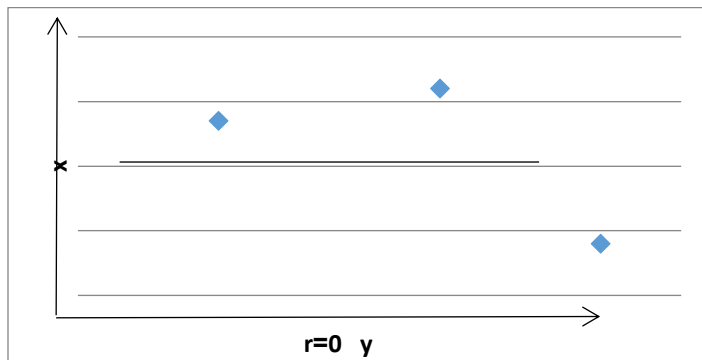
۲. اگر  $r > 0$  باشد (نزدیک به ۱ باشد)، جهت تغییرات هر دو متغیر یکسان است که در این حالت بین دو متغیر  $X$  و  $Y$  رابطه مستقیم برقرار است.



۳. اگر  $r < 0$  باشد (نزدیک به -۱ باشد) آنگاه بین  $X$  و  $Y$  رابطه معکوس وجود دارد یعنی اگر  $X$  اضافه شود  $Y$  کم می شود و بالعکس.



۴. اگر  $r = 0$  باشد آنگاه بین  $X, Y$  همبستگی وجود ندارد.



$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

چون

$$SP_{xy} = \sum XY - \frac{\sum X \sum Y}{N}$$

$$SS_y = (\sum Y^2 - \frac{(\sum Y)^2}{N}) \quad SS_x = (\sum X^2 - \frac{(\sum X)^2}{N})$$

$$r = \frac{SP_{XY}}{\sqrt{SS_x \cdot SS_y}}$$

نکته ۱: ضریب همبستگی X بر روی Y ( $r_{x,y}$ ) برابر است با ضریب همبستگی Y بر روی X ( $r_{y,x}$ ).

نکته ۲: اگر میانگین یکی از متغیرها و یا هردو به صورت اعشاری باشد برای محاسبه ضریب همبستگی بهتر

است از رابطه زیر استفاده شود:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

مثال ۱:

در جدول زیر تغییرات دو متغیر X و Y نشان داده شده است ضریب همبستگی X, Y را حساب کنید.

X	Y
۱	۱
۲	۳
۴	۷
۷	۱۳

x	۱	۲	۴	۷	$\sum x = 14$
y	۱	۳	۷	۱۳	$\sum y = 24$
XY	۱	۶	۲۸	۹۱	$\sum xy = 126$
X <sup>2</sup>	۱	۴	۱۶	۴۹	$\sum x^2 = 70$
Y <sup>2</sup>	۱	۹	۴۹	۱۶۹	$\sum y^2 = 228$

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

$$r = \frac{126 - \frac{14 \times 24}{4}}{\sqrt{(70 - \frac{(14)^2}{4})(228 - \frac{(24)^2}{4})}} = \frac{126 - 84}{\sqrt{(70 - 49)(228 - 144)}} = \frac{42}{\sqrt{21 \times 84}}$$

$$= \frac{42}{42} = 1$$

مثال ۲:

اگر  $SP_{xy} = 90$  و  $SS_x = 250$  و  $SS_y = 40$  باشد مقدار r را محاسبه کنید.

$$r = \frac{SP_{XY}}{\sqrt{SS_x \cdot SS_y}} = \frac{90}{\sqrt{250 \times 40}} = 0/9$$

مثال ۳:

اگر  $SS_x = 5$  و  $SS_y = 20$  و  $r = 0/8$  باشد،  $SP_{xy}$  را محاسبه کنید؟

$$r = \frac{SP_{XY}}{\sqrt{SS_x \cdot SS_y}} \Rightarrow 0/8 = \frac{SP_{XY}}{\sqrt{5 \times 20}} \Rightarrow SP_{XY} = 8$$

## ضریب تبیین (تشخیص):

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

ضریب تشخیص یا تبیین نشان دهنده قسمتی از تغییرات کل است که از طریق رابطه خطی موجود بین X و Y توجیه می شود و اکثراً بر حسب درصد بیان می شود.

ضریب رگرسیون یا تابعیت یا وایزی ( $\beta$ ):

یکی از روش هایی است که با آن می توان رابطه بین دو متغیر را محاسبه نمود و رابطه علت و معلولی بین تغییرات یک متغیر تصادفی و یک متغیر ثابت را تعیین کرد. در بحث رگرسیون شدت تغییرات یک متغیر (وابسته) به ازاء تغییرات متغیر ثابت سنجیده می شود. بنابراین ضریب رگرسیون مشخص می کند که به ازای یک واحد تغییر در متغیر مستقل یا X چه مقدار تغییر در متغیر وابسته یا Y ایجاد می شود.

نکته: ضریب رگرسیون X بر روی Y ( $b_{x,y}$ ) برابر با ضریب رگرسیون Y بر روی X ( $b_{y,x}$ ) نیست. در اینجا حالت

عکس مفهوم ندارد  $b_{y,x} \neq b_{x,y}$

رگرسیون خطی ساده (Simple Linear Regression):

اگر برای شناسایی و پیش بینی متغیر وابسته فقط از یک متغیر مستقل استفاده شود، مدل را رگرسیون خطی ساده می گویند. فرم مدل رگرسیون خطی ساده به صورت زیر است:

$$Y = a + bx + e \quad \text{یا} \quad Y = \beta_0 + \beta_1 X + \varepsilon$$

X: متغیر مستقل  $a(\beta_0)$ : عرض از مبدا  $b(\beta_1)$ : شیب خط

شیب خط در حالت رگرسیون خطی ساده، نشان می دهد که میزان حساسیت متغیر وابسته به متغیر مستقل چقدر است. به این معنی که با افزایش یک واحد به مقدار متغیر مستقل چه میزان متغیر وابسته تغییر خواهد کرد. عرض از مبدا نیز بیانگر مقداری از متغیر وابسته است که به ازاء مقدار متغیر مستقل برابر با صفر محاسبه می شود.

## برآورد پارامترهای رگرسیون خطی ساده

$$\bar{y} = a + b\bar{x} \Rightarrow a = \bar{y} - b\bar{x} \quad \text{رابطه (۱)}$$

مثال ۱: در آزمایشی عملکرد گندم (Y) را در حضور مقادیر متفاوتی از کود مصرفی X اندازه گیری کرده و معادله خطی زیر بدست آمده است:

$$y = 1/08 + 8/77X$$

رابطه بین Y, X را تعیین کنید؟

این معادله نشان می دهد که به ازای افزایش هر کیلوگرم کود، عملکرد گندم به مقدار  $8/77$  کیلوگرم افزایش می یابد یعنی  $\beta_1$  برابر است با مقدار تغییر در  $y$  به ازای هر واحد تغییر در  $x$ .

مثال ۲: رابطه خطی بین درجه حرارت برحسب فارنت هایت ( $x$ ) و تعداد ضربان نبض در دقیقه ( $y$ ) در انسان را براساس جدول زیر محاسبه نمایید.

بیمار	$x$	$y$	$x^2$	$y^2$	$\sum xy$
A	۱۰۰	۹۶	۱۰۰۰۰	۹۲۱۶	۹۶۰۰
B	۸۸	۹۰	۷۷۴۴	۸۱۰۰	۷۹۲۰
C	۸۵	۷۶	۷۲۲۵	۵۷۷۶	۶۴۶۰
D	۷۶	۷۳	۵۷۷۶	۵۳۲۹	۵۵۴۸
E	۷۴	۷۰	۵۴۷۶	۴۹۰۰	۵۱۸۰
	۴۲۳	۴۰۵	۳۶۲۲۱	۳۳۳۲۱	۳۴۷۰۸

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{SS_x} = \frac{sp_{xy}}{SS_x} = \frac{34708 - \frac{423 \times 405}{5}}{(36221 - \frac{(423)^2}{5})} = 1/0225$$

$$\bar{y} = \frac{405}{5} = 81 \quad \bar{x} = \frac{423}{5} = 84.6$$

$$a = \bar{y} - b\bar{x} = 81 - 1/0225 \times 84.6 = -5/505$$

$$Y = a + bX \Rightarrow Y = -5/505 + 1/022X$$

تفسیر: یعنی به ازای افزایش ۱ درجه حرارت بدن تعداد ضربان نبض در دقیقه به  $1/022$  افزایش می یابد. حال اگر  $x=83$  باشد آنگاه تعداد نبض در دقیقه به ۷۹ می رسد:

$$y = -5/505 + 1/022 \times 83 = 79/36$$

مثال ۳: جدول زیر تعداد دانه ( $y$ ) و طول خوشه  $x$  مربوط به ۵ بوته را نشان می دهد.

الف: معادله خط رگرسیون  $y$  به حسب  $x$  را بنویسید.

ب: به ازای افزایش هر ۴ سانتی متر طول خوشه چه مقدار به تعداد دانه اضافه یا از آن کسر می شود.

$x$	$y$	$x^2$	$y^2$	$\sum xy$
۱۰	۲۵	۱۰۰	۴۰۰	۲۵۰
۱۲	۳۰	۱۴۴	۹۰۰	۳۶۰
۱۴	۴۵	۱۹۶	۲۰۲۵	۶۳۰
۱۶	۴۵	۲۵۶	۲۰۲۵	۷۲۰
۱۸	۵۵	۳۲۴	۳۰۲۵	۹۹۰
۷۰	۲۰۰	۱۰۲۰	۸۳۷۵	۲۹۰۰

الف:

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{2530 - \frac{70 \times 200}{5}}{1020 - \frac{(70)^2}{2}} = \frac{150}{40} = 3/75$$

$$\bar{y} = \frac{200}{5} = 40 \quad \bar{x} = \frac{70}{5} = 14$$

$$a = \bar{y} - b\bar{x} = 40 - \frac{3}{75} \times 14 = -12.5$$

$$Y = a + bX \Rightarrow Y = -12/5 + 3/75X$$

ب:

$$Y = bX \Rightarrow Y = 3/75 \times 4 = 15$$

به ازای افزایش هر ۴ سانتی متر طول خوشه تعداد، ۱۵ دانه به تعداد دانه ها اضافه می شود.

مسئله ۱: با توجه به اطلاعات زیر در خصوص دو متغیر  $X, Y$  ضریب رگرسیون و عرض از مبدا را حساب کنید .

$$\bar{x} = 3 \quad \bar{y} = 24 \quad \sum x^2 = 55 \quad \sum xy = 448 \quad n = 5$$

$$\bar{x} = \frac{\sum x}{n} = 3 = \frac{\sum x}{5} = 15 \quad \bar{y} = \frac{\sum y}{n} = 24 = \frac{\sum y}{5} = 120$$

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{\sum 48 - \frac{15 \times 120}{5}}{55 - \frac{(15)^2}{5}} = \frac{88}{10} = 8.8$$

$$\bar{a} = \bar{y} - b\bar{x} \rightarrow a = 24 - 8.8 \times 3 = 2.4 = -2.4$$

$$Y = a + bx \rightarrow y = -2.4 + 8.8x$$

تجزیه واریانس رگرسیون و مراحل آن :

- محاسبه مجموع مربعات رگرسیون یا ( $SSR$ ) :

$$SSR = b^2(x - \bar{x})^2 = b^2 \cdot SS_x$$

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$$

- محاسبه مجموع مربعات کل ( $SST$ )

$$SST = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

- محاسبه مجموع مربعات خطا ( $SSE$  یا  $SSD$ )

$$SSD = SSE = SST - SSR$$

جدول تجزیه واریانس رگرسیون

$F_c$	MS	SS	df	منبع تغییرات
$\frac{MSR}{MSD}$	MSR	SSR	۱	رگرسیون
MSD	$MSD = S^2_{y.x}$	SSD	n-2	انحراف از رگرسیون
		SST	n-1	کل

نکته ۱: درجه آزادی رگرسیون همواره برابر با یک است بنابراین  $SSR = MSR$  خواهد بود.

نکته ۲: در تجزیه رگرسیون  $MSD = S^2_{y.x}$  می باشد یعنی واریانس در سطح نمونه با  $S^2_{y.x}$  برآورد شده و برابر میانگین مربعات انحراف از رگرسیون است.

### آزمون های فرض

الف- آزمون برابری یا عدم برابری ضریب رگرسیون با صفر یا  $H_0 : \beta = 0$

1) استفاده از آزمون t

برای آزمون فرض  $H_0 : \beta = 0$  از توزیع فراوانی t استفاده می شود:

$$t_c = \frac{b}{Sb}$$

که در فرمول فوق b ضریب رگرسیون و Sb انحراف معیار ضریب رگرسیون و  $t_c$  مقدار t محاسباتی می باشد که طبق فرمول زیر محاسبه می شود:

$$Sb = \sqrt{\frac{s^2_{Y.X}}{\sum X^2 - \frac{(\sum X)^2}{n}}} = \sqrt{\frac{s^2_{Y.X}}{\sum (X - \bar{X})^2}} = \sqrt{\frac{MSD}{SSX}}$$

t محاسبه شده با t جدول یا  $t_{\alpha}$  ب درجه آزادی برابر با درجه آزادی خطای رگرسیون یعنی  $df = (n-2)$  مقایسه می شود. اگر t محاسبه شده بزرگتر از t جدول باشد فرض صفر مبنی بر اینکه ضریب رگرسیون برابر با صفر است رد می شود و لذا نتیجه گرفته می شود که رابطه خطی بین متغیر X و Y وجود دارد.

2) استفاده از آزمون F

فرض  $H_0 : \beta = 0$  از طریق مقایسه واریانس های رگرسیون (MSR) و انحراف از رگرسیون ( $MSD = MSE$ ) نیز آزمون می شود. نسبت این دو واریانس دارای توزیع F است. اگر F محاسباتی از F جدول با درجه آزادی صورت (درجه آزادی رگرسیون = ۱) و درجه آزادی مخرج (درجه آزادی انحراف از رگرسیون =  $n-2$ ) و سطح آلفای مورد نظر بزرگ تر باشد فرض صفر مبنی بر اینکه ضریب رگرسیون برابر با صفر می باشد رد می گردد و استنباط می شود که اکثر تنوع و پراکندگی مقادیر Y در اطراف خط رگرسیون خلاصه می شود. به عبارت دیگر خط رگرسیون تنوع مشاهدات را توجیه می کند.

$$t^2 = F \quad t = \frac{b}{Sb} = \frac{b}{\sqrt{\frac{s^2_{Y.X}}{\sum (X - \bar{X})^2}}} \Rightarrow t^2 = \frac{b^2}{\frac{s^2_{Y.X}}{\sum (X - \bar{X})^2}} = \frac{b^2 \sum (X - \bar{X})^2}{s^2_{Y.X}}$$

$$= \frac{MSR}{MSD}$$

ب- آزمون برابری ضریب رگرسیون با یک مقدار ثابت و معین یا  $H_0 : \beta = \beta_0$

در این خصوص نیز t به صورت زیر محاسبه می شود:

$$t_c = \frac{\beta_1 - \beta}{sb}$$

مثال ۱:

داده های آماری مربوط به میزان رشد یک گیاه تحت مقادیر مختلف آب در جدول زیر ارائه شده است.  
الف) معادله خط رگرسیون را بدست آورید.

ب) فرض صفر مبنی بر  $H_0: \beta = 0$  را آزمون نمایید.

ج) فرض صفر مبنی بر  $H_0: \beta = 0.2$  را آزمون نمایید و نتیجه را تفسیر نمایید.

x	y	X <sup>2</sup>	Y <sup>2</sup>	xy	
68	65	4624	4225	4420	
63	66	3969	4356	4158	
67	68	4489	4624	4556	
64	65	4096	4225	4160	
68	69	4624	4761	4692	
62	66	3844	4356	4092	
70	68	4900	4624	4760	
66	65	4356	4225	4290	
67	67	4489	4489	4489	
69	68	4761	4624	4692	
71	70	5041	4900	4970	
68	71	4624	5041	4828	
<b>∑</b>	<b>803</b>	<b>808</b>	<b>53817</b>	<b>54450</b>	<b>54107</b>

$$\bar{Y} = 67.33 \quad \bar{X} = 66.92$$

$$\beta_1 = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} = \frac{54107 - \frac{803 \times 808}{12}}{53817 - \frac{(803)^2}{12}} = \frac{54107 - 54068}{53817 - 53734} \Rightarrow b = \frac{39}{83} = 0.469$$

$$\bar{y} = \beta_0 + \beta_1 \bar{X} \Rightarrow \beta_0 = \bar{y} - \beta_1 \bar{X} = 67.33 - 0.469 \times 66.92 = 35.94$$

$$Y = \beta_0 + \beta_1 X = 35.94 + 0.469X$$

ب- برای حل این قسمت باید جدول تجزیه واریانس تشکیل شود و براساس مقدار  $F$  نتیجه گیری شود.

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

$$SS_X = \sum X^2 - \frac{(\sum X)^2}{n} = 53817 - \frac{(803)^2}{12} = 53817 - 53734 = 83$$

$$SS_R = b^2 SS_X = (0.469)^2 (83) = 17.70$$

$$SS_T = \sum y^2 - \frac{(\sum y)^2}{n} = 54450 - \frac{(808)^2}{12} = 54450 - 54405.3 = 44.7$$

$$SS_T = SS_R + SS_D \Rightarrow SS_T - SS_R = SS_D \Rightarrow 44.7 - 17.70 = 27$$

منبع تغییرات	df	SS	MS	$F_c$
رگرسیون	۱	17/70	17/70	$\frac{17.70}{2.70} = 6.55$
انحراف از رگرسیون	۱۰	27	2/7	
کل	۱۱	44/70		

چون  $F_c$  بزرگتر از  $F_t$  است فرض صفر مبنی بر اینکه ضریب رگرسیون برابر صفر می باشد رد می شود یعنی بین  $x, y$  رابطه خطی وجود دارد. به عبارت دیگر خط رگرسیون تنوع مشاهده شده بین مقادیر  $Y$  را به خوبی توجیه می نماید و اکثر نقاط در اطراف این خط متمرکز هستند.

$$F_t \Big|_{\alpha=0.05}^{10} = 4.96$$

فرض مزبور با استفاده از توزیع  $t$  نیز آزمون می شود:

$$t_c = \frac{\text{ضریب رگرسیون}}{\text{انحراف معیار رگرسیون}} = \frac{b}{Sb}$$

$$Sb = \sqrt{\frac{s^2 Y.X}{\sum X^2 - \frac{(\sum X)^2}{n}}} = \sqrt{\frac{MSD}{SSX}} = \sqrt{\frac{2.674}{83}} = 0.180$$

نکته: واریانس  $Y$  برابر با میانگین مربعات خطا یا انحراف از رگرسیون می باشد که بایستی از جدول تجزیه واریانس استخراج شود ( $s^2_{yx} = MSD \text{ or } MSE$ ).

$$t_c = \frac{b}{sb} = \frac{0.469}{0.180} = 2.61 \quad t_t \Big|_{\alpha:0.5}^{df=n-2=12-2=10} = 2.23$$

چون  $t_c$  بزرگتر از  $t_t$  با درجه آزادی ۱۰ و سطح اطمینان ۵ درصد است بنابراین فرض صفر رد می شود و استنباط می شود که با احتمال ۹۵٪ بین دو متغیر رابطه خطی وجود دارد.

ج:

$$H_0: \beta = 0.2 \text{ or } \beta_1 = B$$

$$H_1: \beta \neq 0.2 \text{ or } \beta_1 \neq B$$

$$t_c = \frac{\beta_1 - B}{sb} = \frac{0.469 - 0.2}{0.180} = 1.5$$

$$t_t \Big|_{\alpha:0.5}^{df:10} = 2.23$$

چون  $t_c$  کوچک تر از  $t_t$  است پس فرض صفر رد نمی شود یعنی ضریب رگرسیون با ۰,۲ اختلاف معنی داری ندارد پس رابطه خطی برقرار می باشد.

نکته: وقتی که ضریب رگرسیون یا  $b$  برابر با صفر باشد یعنی بین دو متغیر رابطه خطی وجود ندارد.

مثال ۲: جدول زیر درصد میوه آفت زده ( $X$ ) را برای اندازه های مختلف میوه ( $Y$ ) نشان می دهد. الف) مطلوب است معادله ی خط رگرسیون؟

ب) فرض صفر مبتنی بر  $\beta=0$  را از طریق تجزیه ی واریانس و آزمون  $t$  حساب کنید.

ج) آیا می توان با ۹۹٪ اطمینان مقدار  $\beta$  را برابر ۰.۹ دانست؟

x	y	xy	X <sup>2</sup>	y <sup>2</sup>
39	65	2535	1521	4225
43	78	3354	1849	6084
21	52	1092	441	2704
64	82	5248	4096	6724
57	92	5244	3249	8464
47	89	4183	2209	7921
28	73	2044	784	5329
75	98	7350	5625	9604
34	56	1904	1156	3136
52	75	3900	2704	5625
$\Sigma$	460	36854	23634	59816

(الف)

$$\beta_1 = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{36854 - \frac{460 \times 760}{10}}{23634 - \frac{(460)^2}{10}} = \frac{36854 - 34960}{23634 - 21160} = \frac{1894}{2474} = 0.77$$

$$Y = \beta_0 + \beta_1 x \rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x} \rightarrow \beta_0 = 76 - 0.77(46) = 40/6$$

$$y = 40/6 + 0.77x$$

(ب)

$$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 23634 - \frac{(460)^2}{10} = 23634 - 21160 = 2474$$

$$SSR = b^2 (x - \bar{x})^2 = (0.77)^2 (2474) = 1466.83$$

$$SST = \frac{\Sigma y^2 - (\Sigma y)^2}{n} = \frac{59216 - (760)^2}{10} = 2056$$

$$SS_T = SS_R + SS_E \Rightarrow SS_T - SS_R = SS_E \Rightarrow 2056 - 1466.83 = 589.17$$

$$F_t |_{\frac{1}{8}} = 11/26$$

چون  $F_c$  از  $F_t$  بزرگتر است پس فرض صفر ( $H_0$ ) که مبتنی بر عدم وجود رابطه بین دو متغیر است رد می

شود یعنی دو متغیر با هم رابطه ی خطی دارند.

F <sub>c</sub>	MS	SS	df	S.O.V
$\frac{1466/83}{73/65} = 19/92$	1466/83	1466/83	1	رگرسیون
	73/65	589/17	8	خطا
		2056	9	کل

با استفاده از آزمون t:

$$tc = \frac{b}{Sb} \Rightarrow tc = \frac{0.77}{0.16} = 4.81$$

$$S_b^2 = \frac{s_{xy}^2}{SS_x} = \frac{73/65}{2474} = 0/0268 \Rightarrow Sb = \sqrt{0/0268} = 0/16$$

$$t_{t|8} = 3/36$$

چون  $T_c > T_t$  است فرض صفر رد می شود و بین دو متغیر رابطه ی خطی وجود دارد.

(ج)

$$H_0: \beta = 0/9 \quad H_1: \beta \neq 0/9$$

$$tc = \frac{\beta_1 - \beta}{Sb} = \frac{0/77 - 0/9}{0/163} = |-0/797| = 0/797$$

چون  $tc$  از  $t_t$  کوچکتر است بنابراین فرض صفر رد نمی گردد و از نظر آماری نمی توان  $0/77$  را از  $0/9$  متفاوت دانست.

نمونه سوالات امتحان پایان ترم:

۱- تاثیر درجه حرارت بر مقدار افزایش آنتی بیوتیک در یک طرح آماری مطالعه شده و نتایج زیر حاصل شده است.

الف) معادله خط رگرسیون را بدست آورده و مفهوم ضریب رگرسیون را بیان نمایید.

ب) جدول تجزیه واریانس را تشکیل داده و فرض صفر مبنی بر  $H_0: \beta = 0$  را آزمون نمایید و این فرض را بااستفاده از توزیع t در سطح احتمال  $\alpha = 1\%$  بررسی نمایید.

ج) چنانچه آزمایش در درجه حرارت ۱۰ و ۱۰۰ درجه سانتی گراد انجام شود مقدار آنتی بیوتیک مورد انتظار چقدر خواهد بود؟

د) فرض صفر مبنی بر  $\beta = 0.3$  را آزمون نمایید.

$$\alpha = 1\% \quad t_t = 4.032 \quad f_t = 16.26$$

(الف)

$$\beta_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{10530 - \frac{430 \times 190}{7}}{280000 - \frac{(420)^2}{7}} = \frac{-870}{2800} = -0/31$$

$$Y = \beta_0 + \beta_1 x \rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x} \rightarrow \beta_0 = 27/14 - 0/31(60) = 45/74$$

$$y = 45/74 - 0/31x$$

X درجه حرارت	آنتی بیوتیک y
۳۰	۳۹
۴۰	۳۱
۵۰	۳۰
۶۰	۲۷
۷۰	۲۲
۸۰	۲۳
۹۰	۱۸

مفهوم ضریب رگرسیون: در اثر افزایش یک واحد درجه حرارت مقدار آنتی بیوتیک به اندازه ی ۰/۳۱ واحد کاهش می یابد.

(ب)

$$SSR = \frac{(-870)^2}{2800} = 270.22$$

$$SST = \sum y^2 - \frac{(\sum y)^2}{n} = 54/48 - \frac{(190)^2}{7} = 290/9$$

$$SS_T = SS_R + SS_E \Rightarrow SS_T - SS_R = SS_E \Rightarrow 290/2 - 21/2 = 279/7$$

Fc	MS	SS	df	S.O.V
$\frac{269/7}{4/24} = 63/61$	270.22	270/22	۱	رگرسیون
	4/24	21/2	۵	انحراف از رگرسیون
		290/9	۶	کل

$$F_t \Big|_5^1 = 16/26$$

چون  $F_t$  کوچکتر از  $F_c$  است پس فرض صفر رد می شود یعنی اکثر نقاط در اطراف خط متمرکز است.

آزمون فرض صفر با آزمون  $t$ :

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

$$S_b = \sqrt{\frac{s^2 Y.X}{\sum X^2 - \frac{(\sum X)^2}{n}}} = \sqrt{\frac{MSD}{SSX}} = \sqrt{\frac{24/4}{2800}} = 0/038$$

$$t = \frac{b}{S_b} = \frac{(-0/31)}{0/038} = -8.16 \rightarrow |-16/05| = 8.16$$

$$t_t \Big|_5^1 = 4/032$$

$t$  محاسباتی از  $t$  جدول بزرگتر است بنابراین فرض صفر رد می شود.

(ج) چون  $X=100$  از دامنه داده ها فاصله دارد، بنابراین نمی توان مقدار آنتی بیوتیک را برآورد نمود اما برای  $10 =$

$X$  به شرح زیر محاسبه می شود:

$$Y = \beta_0 + \beta_1 x = 45/74 - 0/31x$$

$$Y = \beta_0 + \beta_1 x = 45/74 - 0/31(10) = 42/64$$

(د)

$$H_0: \beta = 0/3$$

$$H_1: \beta \neq 0/3$$

$$S_b = \sqrt{\frac{s^2 Y \cdot X}{\sum X^2 - \frac{(\sum X)^2}{n}}} = \sqrt{\frac{MSD}{SSX}} = \sqrt{\frac{24/4}{2800}} = 0/038$$

$$t_c = \frac{b-\beta}{S_b} = \frac{-0/31-0/3}{0/038} = 16/052$$

$$t_t|_5^1 = 4/032$$

چون  $t_c$  از  $t_t$  بزرگتر است بنابراین فرض صفر رد می گردد و از نظر آماری می توان  $-0/31$  را از  $0/3$  متفاوت دانست.

۲. در یک مزرعه ۲۵٪ بوته ها دیر رس ۱۵٪ پاکوتاه و ۱۰٪ دیر رس و پاکوتاه چنانچه بوته ای به تصادف انتخاب شود احتمال اینکه دیررس یا پاکوتاه باشد چند درصد است؟

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = 25 + 15 - 10 = 30$$

۳. متغیر تصادفی  $X$  برطبق قانون دو جمله ای با پارامترهای  $n=30$  و  $P=0.4$  توزیع می شود واریانس متغیر تصادفی  $X$  را بدست آورید.

$$V(X) = n \cdot p \cdot q$$

$$q = 1 - p \rightarrow 1 - 0.4 = 0.6$$

$$V(X) = n \times p \times q = 30 \times 0.4 \times 0.6 = 7.2$$

۴. در مطالعات هم بستگی دو متغیر بین بوته ها ( $X$ ) و ارتفاع آنها ( $Y$ ) اطلاعات زیر درست می باشد. ضریب رگرسیون و  $dF$  خطا یا انحراف از رگرسیون و عرض از مبدأ را به دست آورید.

$$N=5 \quad \sum xy = 448 \quad \sum n^2 = 55 \quad \bar{y} = 24 \quad \bar{x} = 3$$

ضریب رگرسیون:

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{448 - \frac{15 \times 120}{5}}{55 - \frac{(15)^2}{5}} = \frac{88}{10} = 8.8$$

$dF$  خطا:

$$dF = n - 2 = 5 - 2 = 3$$

عرض از مبدأ:

$$\bar{y} = \frac{\sum y}{n} = 24 = \frac{\sum y}{5} = 120 \quad \bar{y} = a + b\bar{x}$$

$$\bar{x} = \frac{\sum x}{n} = 3 = \frac{\sum x}{5} = 15$$

$$a = \bar{y} - b\bar{x} \Rightarrow 24 - 8.8(3) = -2.4$$

۵- اگر متغیر تصادفی  $X$  دارای تابع چگالی دوجمله ای  $p(x) = \binom{20}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{20-x}$  باشد؛ واریانس توزیع دو جمله ای را به دست آورید.

$$\text{واریانس} = npq = 20 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = 5$$

۶- تیراندازی با احتمال  $0.7$  هدف خود را مورد اصابت قرار می دهد. چه قدر احتمال دارد که از سه شلیک، دو شلیک به هدف بخورد؟

$$p(2) = \binom{3}{2} (0.7)^2 (0.3)^1 = 3(0.49)(0.3) = 0.44$$

احتمال این که از سه شلیک الف) کم تر از دو شلیک ب) حداقل یک شلیک ج) دو یا سه شلیک به هدف بخورد چه قدر است؟

۷- اگر احتمال قبولی یک دانشجو  $0.4$  باشد؛ احتمال این که از بین هر ۵ دانشجو: الف) صفر دانشجو ب) حداقل یک دانشجو قبول شود؛ چه قدر است؟  
احتمال قبولی صفر دانشجو:

$$p(0) = \binom{5}{0} (0.4)^0 (0.6)^5 = 0.07776$$

حداقل یک دانشجو قبول:

$$p(\text{حداقل یک دانشجو قبول}) = 1 - (\text{احتمال قبولی صفر دانشجو}) = 1 - 0.07776 = 0.92224$$

ج) احتمال قبولی یک دانشجو و نیز احتمال قبولی همه دانشجویان را هم برآورد کنید.

$$p(1) = \binom{5}{1} (0.4)^1 (0.6)^4 = 0.2592$$

$$p(5) = \binom{5}{5} (0.4)^5 (0.6)^0 = 0.1024$$

$$\delta^2 = npq = 5(0.4)(0.6) = 1.2$$

$$\mu = np = 5(0.4) = 2$$

۸- مطلوبست احتمال مشاهده حالات ۰ تا ۶ شیر در آزمایش ۶ بار یک سکه متعادل:

$$p(0) = \binom{6}{0} p^0 q^6 = 1 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 = 1 \left(\frac{1}{64}\right) = \frac{1}{64}$$

$$p(1) = \binom{6}{1} p^1 q^5 = 6 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^5 = 6 \left(\frac{1}{2}\right) \left(\frac{1}{32}\right) = \frac{6}{64}$$

$$p(2) = \binom{6}{2} p^2 q^4 = 15 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = 15 \left(\frac{1}{4}\right) \left(\frac{1}{16}\right) = \frac{15}{64}$$

$$p(3) = \binom{6}{3} p^3 q^3 = 20 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^3 = 20 \left(\frac{1}{8}\right) \left(\frac{1}{8}\right) = \frac{20}{64}$$

$$p(4) = \binom{6}{4} p^4 q^2 = 15 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 = 15 \left(\frac{1}{16}\right) \left(\frac{1}{4}\right) = \frac{15}{64} \quad \text{و} \quad \frac{6!}{4!2!} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 = 15 \left(\frac{1}{16}\right) \left(\frac{1}{4}\right) = \frac{15}{64}$$

$$p(5) = \binom{6}{5} p^5 q^1 = 6 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^1 = 6 \left(\frac{1}{32}\right) \left(\frac{1}{2}\right) = \frac{6}{64}$$

$$p(6) = \binom{6}{6} p^6 q^0 = 1 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^0 = 1 \left(\frac{1}{64}\right) = \frac{1}{64}$$

احتمال حداقل ۴ بار مشاهده ی شیر چه قدر است؟

$$p = \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{22}{64}$$

همان طور که مشاهده می شود؛ توان  $p$  از صفر شروع شده تا به  $n=6$  رسیده و توان  $q$  از ۶ شروع و به صفر رسیده است.

$$(p+q)^6 = p^6q^0 + p^5q^1 + p^4q^2 + p^3q^3 + p^2q^4 + p^1q^5 + p^0q^6$$

برای ضریب قبل از هر جمله، از ترکیب  $\binom{n}{x}$  استفاده می کنیم که در آن  $x$  برابر است با توان  $p$  در هر جمله

$$(p+q)^6 = 1p^6q^0 + 6p^5q^1 + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6p^1q^5 + 1p^0q^6$$

۹- از جعبه ای حاوی ۱۰ توپ قرمز، ۳۰ توپ سفید و ۲۰ توپ آبی و ۱۵ توپ نارنجی، تویی برمی داریم. پیدا کنید احتمال این که توپ انتخابی

(الف) نارنجی یا قرمز (ب) نه قرمز نه آبی (ج) نه آبی (د) قرمز، سفید و یا آبی باشد

$$\text{ب) } \frac{30}{75} + \frac{15}{75} = \frac{45}{75} = \frac{3}{5}$$

$$\text{الف) } \frac{15}{75} + \frac{10}{75} = \frac{25}{75} = \frac{1}{3}$$

$$\text{ه) } \frac{10}{75} + \frac{30}{75} + \frac{20}{75} = \frac{60}{75} = \frac{4}{5}$$

$$\text{ج) } \frac{10}{75} + \frac{30}{75} + \frac{15}{75} = \frac{55}{75} = \frac{11}{15}$$

۱۰- جعبه ای دارای ۳ توپ سفید و ۳ توپ سیاه است. دو توپ را یکی پس از دیگری و بدون جایگزاری از جعبه خارج می کنیم. احتمال این که

دو توپ انتخابی هم رنگ نباشند چه قدر است؟

$$\text{احتمال اولی سیاه دومی سفید } \left(\frac{3}{6}\right)\left(\frac{3}{5}\right) = \frac{9}{30}$$

$$\text{احتمال اولی سفید دومی سیاه } \left(\frac{3}{6}\right)\left(\frac{3}{5}\right) = \frac{9}{30}$$

$$\frac{9}{30} + \frac{9}{30} = \frac{18}{30} = \frac{3}{5}$$

۱۱- خانواده ای دارای ۳ فرزند است. احتمال آن که هر سه فرزند آنها پسر یا هر سه دختر و یا آن که دو پسر و یک دختر و یا دو دختر و یک پسر

باشند؛ چه قدر است؟

۷. پ د د

۵. پ پ د

۳. پ د پ

۱. پ پ پ

۸. د د د

۶. د پ د

۴. د د پ

۲. د پ پ

$$P = \frac{h}{h} = \frac{3}{8} \text{ دو پسر و یک دختر}$$

$$P = \frac{h}{h} = \frac{1}{8} \text{ سه پسر}$$

$$P = \frac{h}{h} = \frac{3}{8} \text{ دو دختر و یک پسر}$$

$$P = \frac{h}{h} = \frac{1}{8} \text{ سه دختر}$$

۱۲- اگر در پرتاب یک تاس متغیر تصادفی مورد نظر، ۲ برابر عدد ظاهر شده به اضافه ۱ باشد امید ریاضی، میانگین،

واریانس و انحراف معیار این متغیر را محاسبه کنید.

$$Y = 2X+1$$

متغیر X	2	4	6	8	10	12
$2X+1=Z$	5	9	13	17	21	25
P(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(x) = \mu = \sum Z \cdot P(X)$$

$$= 5\left(\frac{1}{6}\right) + 9\left(\frac{1}{6}\right) + 13\left(\frac{1}{6}\right) + 17\left(\frac{1}{6}\right) + 21\left(\frac{1}{6}\right) + 25\left(\frac{1}{6}\right) = \frac{90}{6}$$

$$= 15$$

۱۳- در پرتاب یک تاس، احتمال اینکه عدد ظاهر شده کمتر از 3 نباشد، چقدر است؟

$$S = \{1,2,3,4,5,6\}$$

$$E = \{1,2\}$$

$$P(E) = \frac{n(E)}{n(S)} = \frac{2}{6} = \frac{1}{3}$$

$$P(E') = 1 - P(E) = 1 - \frac{1}{3} = \frac{2}{3}$$

۱۴- در پرتاب دو سکه، احتمال اینکه حداقل یک بار شیر ظاهر شود، چقدر است؟

$$S = \{\text{ش ش، ش خ، خ ش، ش ش، ش ش، ش ش}\}$$

$$E = \{\text{ش ش، ش ش، ش ش}\}$$

$$P(E) = \frac{n(E)}{n(S)} = \frac{3}{4} = 0.75$$

۱۵- در پرتاب سه سکه، احتمال اینکه دو بار شیر ظاهر شود، چقدر است؟

$$S = \{\text{ش ش ش، ش ش ش، ش ش ش، ش ش ش، ش ش ش، ش ش ش، ش ش ش، ش ش ش}\}$$

$$E = \{\text{ش ش ش، ش ش ش}\}$$

$$P(E) = \frac{n(E)}{n(S)} = \frac{2}{8} = 0.25$$

۱۶- تاس سالمی را پرتاب می کنیم. اگر بدانیم عدد ظاهر شده فرد می باشد، احتمال اینکه عدد ظاهر شده کمتر از 4 باشد، چقدر است؟

$$S = \{1,2,3,4,5,6\}$$

$$E_1 = \{1,2,3\}$$

$$E_2 = \{1,3,5\}$$

$$P(E_1) = \frac{n(E_1)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

$$P(E_2) = \frac{n(E_2)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

$$E_1 \cap E_2 = \{1,3\} \Rightarrow P(E_1 \cap E_2) = \frac{2}{6} = \frac{1}{3}$$

$$P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

۱۷- یک تاس سالم دو بار پرتاب می شود. مطلوبست احتمال اینکه مجموع شماره هایی که در دو بار پرتاب ظاهر می شوند، کمتر از ۰ یا بیشتر از ۹ باشد؟

منابع مورد استفاده :

مفاهیم آمار و احتمالات، تألیف دکتر عبدالمجید رضایی، نشر مشهد.  
آمار پیشرفته و بیومتری، تألیف دکتر عباسعلی خواجه نوری، انتشارات دانشگاه تهران.  
آمار و احتمالات (کاربرد در کشاورزی)، تألیف عبدالمجید رضایی، سیدعلی محمد، انتشارات جهاد دانشگاهی، واحد اصفهان.  
آمار و احتمالات کاربردی، تألیف بهمن یزدی صمدی و همکاران (سیدعلی پیغمبری، و حسن امیری اوغان)، انتشارات دانشگاه تهران.  
طرحهای آزمایشی در علوم کشاورزی؛ تألیف دکتر محمد فارسی و همکاران، انتشارات به نشر کتاب.  
طرحهای آزمایشی در علوم کشاورزی؛ تألیف دکتر سیدعلی پیغمبری، انتشارات دانشگاه تهران.