آمار توصیفی

شامل **مجموعه ای از روش ها برای برنامه ریزی آزمایشها**َ **بدست آوردن اطلاعات و سازماندهی**َ **خلاصه سازی** **ارایه و تجزیه و تحلیل اطلاعات** و در نهایت **نتیجه گیری بر مبنای این اطلاعات** می باشد.

آمار در لغت به معنای شمارش است.

آمار توصیفی ، مجموعه روشهایی است که به خلاصه کردن ، طبقه بندی ، توصیف و تفسیر داده ها می پردازد.

هدف آمار توصیفی ، توصیف واقعیات موجود است که تنها خطای موجود در آن خطای اندازه گیری است.

مشخص کردن قد دانش آموران یک کلاس ، بررسی روند رشد سرانه ملی در سه سال گذشته نمونه ای از مفاهیم مورد بررسی آمار توصیفی هستند.

نمادهای مهم در آمار عبارت اند از :

شاخص مورد مطالعه	نمونه (آماره)	جامعه (پارامتر)
میانگین	$\bar{X}$	μ (مو)
نسبت	ρ	$\hat{P}$ (پی هد)
همبستگی	r_{xy}	ρ_{xy} (رو)
واریانس	S^2	δ^2 (سیگما به توان ۲)
انحراف معیار	S	δ (سیگما)

N	N	تعداد

X = نمره یا نمره ها

X_c= حد میانی نمرات

L = کرانه طبقات

F_i = فراوانی مطلق

F_C= فراوانی تجمعی

X = انحراف نمره از میانگین

X^2 = مجذور (مربع نمره) انحراف نمره از میانگین

SS= مجموع مجذورات انحراف نمره از میانگین

MS = میانگین مجموع مجذور انحراف نمره از میانگین

i= فاصله طبقاتی

حد واقعی اعداد

حدود واقعی : حدود واقعی نمرات بصورت کم کردن **0/5** نمره در اعداد صحیح ودر اعداد اعشاری ، نیم واحد یعنی **0/05** و مانند آن کسر میشود. یعنی : کرانه عدد **25** ⟵ **24/5** تا **25/5** و کرانه عدد **25/5** ⟵ **25/45** تا **25/55** می باشد.

مفهوم حدود واقعی مخصوصاً زمانی مفید است که اعداد گروه بندی یا طبقه بندی شوند.

مثال :پس از اجرای یک آزمون ریاضی مشاهده میشود که ۱۰ نفر نمره ۱۲ گرفته اند . این بدان معنی نیست که همه توانایی یکسان دارند ، بلکه دقیق نبودن وسیله اندازه گیری ممکن است موجب این امر شده باشد. به این خاطر نیاز به حدود واقعی می باشد : یعنی ۵-۱۲-۱۱/۵

.

توزیع فراوانی

عبارت است از سازمان دادن اندازه ها یا مشاهدات بصورت طبقات همراه با فراوانی هر طبقه .

توزیع فراوانی ، داده ها را بصورت خلاصه ومرتب ، به نحوی که تفسیر آنها آسان شود ، نمایش میدهند.

❖ مراحل ساخت جدول توزیع فراوانی :

۱- مرتب کردن اعداد از بزرگ به کوچک یا برعکس.

۲- مشخص کردن تعداد دفعاتی که هر عدد تکرار شده است (تعداد فراوانی) . زمانی که همه اعداد تک تک در جدول آورده شوند، جدول توزیع فراوانی منفرد یا طبقه بندی نشده گفته می شود. اما زمانی که نمره ها یا اعداد دارای دامنه ی گسترده ای هستند و تنظیم اعداد بصورت توزیع فراوانی طبقه بندی نشده وقتگیر و طاقت فرسا است. اعداد را طبقه بندی می کنیم و از جدول توزیع فراوانی طبقه بندی شده استفاده می کنیم.(در حال حاضر با وجود رایانه طبقه بندی اعداد غیر منطقی است.)

زیرا : در اثر طبقه بندی کردن اطلاعات ، برخی از اطلاعات از بین می رود ، ستون داده ها (طبقات) را در جدول فراوانی با X نشان می دهند. فراوانی مطلق (f) برابر است با مقدار دفعات تکرار هر داده در هر طبقه.

➢ مثال : در توزیع فراوانی درس آمار یک کلاس ، نمرات به شرح زیر می باشد . جدول فراوانی مربوط به توزیع را فراهم کنید ؟

۱۰-۱۲-۱۱-۱۰-۱۱-۱۲-۱۰-۱۲-۱۵-۱۰-۱۱-۱۱-۱۱-۱۱

X	F
15	1
12	3
11	5
10	4

نکته : با توجه به جدول فوق ، عدد ۵ در ستون f بیانگر اینست که عدد ۱۱ پنج بار تکرار شده است.اگر داده های ستون فراوانی (F) را باهم جمع کنیم ، تعداد کل داده ها بدست می آید.

$N = \sum F$ یعنی در مثال فوق N=13

توزیع فراوانی طبقه بندی شده

زمانی که تعداد اعداد یک توزیع و همچنین فاصله ی بین آنها خیلی زیاد باشد ، از توزیع فراوانی طبقه بندی شده استفاده می شود. زمانی که تفاضل بین بزرگترین و کوچکترین نمره یا عدد مساوی یا بزرگتر از ۲۰ باشد از توزیع فراوانی طبقه بندی شده استفاده می شود.
طبقات بایستی ناسازگار باشند ، یعنی یک عدد معین فقط در یک طبقه قرار داده شود

❖ نحوه ساخت توزیع فراوانی طبقه بندی شده

برای ساختن توزیع فراوانی طبقه بندی شده دو روش وجود دارد.

الف)روش اول دارای مراحل زیر است :

۱. *تعیین دامنه تغییرات* *برای داده های گسسته* $R = X_{Max} - X_{min}$

برای داده های پیوسته $R = X_{Max} - X_{min}+1$

۲. *تقسیم R به تعداد طبقات ، عدد بدست آمده فاصله طبقات و مقسوم علیه تعداد طبقات می باشد.*

۳. *نوشتن طبقات*

٤. *نوشتن فراوانی طبقات*

ب) روش دوم دارای مراحل زیر است :

۱- تعیین دامنه تغییرات

۲- تعیین تعداد طبقات با استفاده از قانون استرژ $\boldsymbol{K = 1 + 3/3\ \mathrm{Log}_N}$

۳- تعیین اندازه یا حجم هر طبقه (فاصله طبقات) $\boldsymbol{i = \frac{R}{K}}$

۴- نوشتن طبقات

۵- نوشتن فراوانی طبقات

- ✓ **نکته** : تعداد طبقات اختیاری است و معمولا بین ۱۰-۲۰ است. و اگر تعداد طبقات بزرگتر از ۲۰ باشد ، تهیه و تنظیم جدول نیاز به وقت و کار بیشتر دارد. اگر تعداد طبقات کوچکتر از ۱۰ باشد اندازه طبقات بزرگ می شود و اطلاعات بیشتری از دست می رود.

نماینده طبقات (نقاط وسط طبقات) : نماینده طبقات یا نقاط میانی را با X' نمایش می دهند و از طریق فرمول زیر بدست می آید : $X' = \frac{\text{حد بالای طبقه} + \text{حد پایین طبقه}}{2}$

توزیع فراوانی تراکمی :

اگر پژوهشگری علاقمند به دانستن تعداد افراد یا نمره هایی باشد که در پایین نمره یاعدد خاصی وجود دارند ، نیاز به توزیع فراوانی تراکمی دارد . فراوانی تراکمی با (cf) نشان داده می شود که از جمع کردن فراوانی های ساده هر طبقه یا طبقه بزرگتر بدست می آید.

- ✓ نکته ۱ : فراوانی تراکمی کوچکترین طبقه همیشه برابر با ساده یا مطلق آن است.
- ✓ نکته ۲ : فراوانی تراکمی بزرگترین طبقه همیشه برابر با مجموع داده ها$(\sum F)$ یا N می باشد.

محاسبه ی فراوانی نسبی :

$$\text{فراوانی نسبی} = \frac{\text{فراوانی مطلق}}{\text{جمع کل فروانی}}$$

درصد فراوانی مطلق و تراکمی :

$$P = \frac{\text{فراوانی مطلق هر طبقه}}{\text{تعداد کل فراوانی ها}} \times 100 \qquad P = \frac{F}{N} \times 100$$

$$cf\% = \frac{\text{فراوانی تراکمی هر طبقه}}{\text{تعداد کل فراوانی ها}} \times 100 \qquad CF\% = \frac{CF}{N} \times 100$$

- مثال:

داده های زیر نمرات درس آمار یک کلاس میباشند جدول فراوانی را با ۵ طبقه تشکیل دهید

۱۷ ۹ ۱۴ ۱۴ ۱۱ ۱۲ ۱۳ ۱۶ ۵ ۸ ۹ ۱۰ ۱۰ ۷ ۱۹ ۱۹ ۱۸ ۱۶ ۱۳ ۱۳ ۱۰

۱۰ ۱۳ ۱۳ ۱۲

نمودار ها

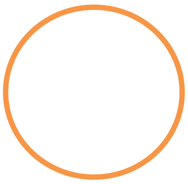
❖ **نمودار دایره ای** :نموداری است که با داده های اسمی و کیفی بکار می رود . برای محاسبه درجه و زاویه مرکزی متعلق به یک گروه از فرمول زیر استفاده می کنیم

$$\text{درجه}^{\circ} = \frac{n}{N} \times 360$$

تعداد داده های آن گروه=n

تعداد کل داده ها=N

➢ مثال : اگر داخل یک گروه ۱۲۰ نفری ۶۰ نفر قبول ۴۰، نفر تجدید و ۲۰ نفر مردود شده باشند، درجه مربوط به تجدیدی ها چند است ؟

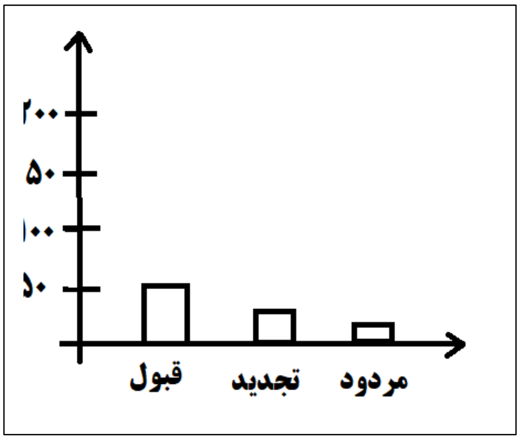


$$\text{درجه تجدیدی} = \frac{۴۰}{۱۲۰} \times ۳۶۰ = ۱۲۰$$

❖ **نمودار ستونی (میله)** : نموداری است که با داده های اسمی بکار می رود که در محور عمودی ، فراوانی و در محور افقی طبقات قرار می گیرند . در نمودار میله ای فاصله بین اعداد و نقاط یکسان و ثابت است. مثل تعداد فرزندان. ولی اگر برای متغیر های طبقه ای مانند قبولی – مردودی ، رنگ چشم و اعتقادات مذهبی نمودار رسم شود.

هیچ ضرورتی ندارد که فاصله ها در محور x ثابت می ماند. برای این متغیرها از نمودار دایره ای و مانند آن استفاده می کنند (نمودار پای).

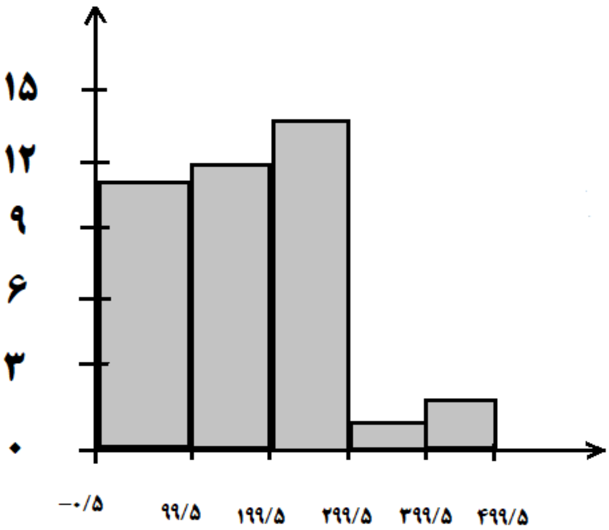
مثال :



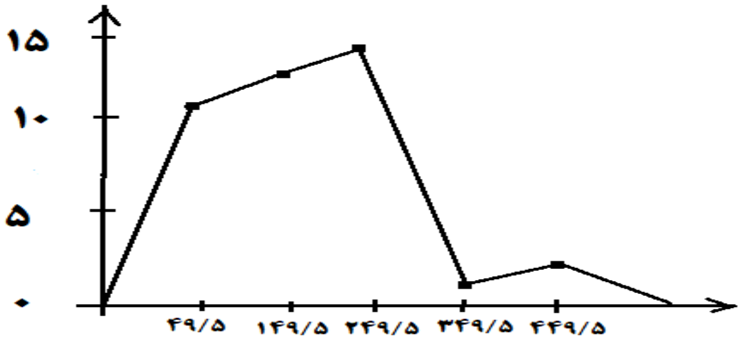
❖ **نمودار هیستوگرام** : با داده های فاصله ای و نسبتی بکار می رود.نموداری شبیه نمودار ستونی است ، ولی در آن ستونها به هم چسبیده است و در محور افقی (x) کرانه (حدود واقعی) طبقات و در محور عمودی (y) فراوانی مطلق قرار می گیرد.

✓ **نکته** : این نمودار اغلب در مواردی استفاده می شود که بخواهیم فقط یک نمودار واحد را نشان دهیم. در حالیکه در موقعیت های آزمایشی و شبه آزمایشی که مایلیم نمره های آزمون ... را با هم مقایسه کنیم ناگزیر هستیم از پلیگون استفاده

	i
٤٠٠ _ ٤٩٩	٢
٣٠٠ _ ٣٩٩	١
٢٠٠ _ ٢٩٩	١٤
١٠٠ _ ١٩٩	١٢

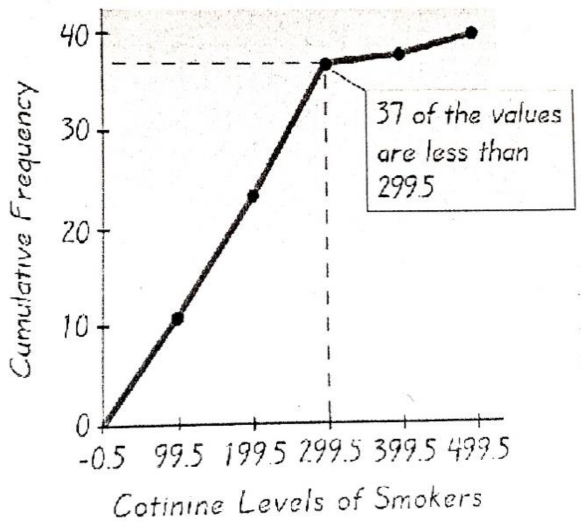


❖ **نمودار چند ضلعی (پلیگون)** : با داده های فاصله ای و نسبتی بکار می رود .
نموداری شبیه نمودار هیستوگرام است که در محور x چند نمره میانی و در محور y فراوانی مطلق قرار می گیرد.



✓ **نکته** : برای مقایسه ی چندگروه در یک محور مختصات اگر تعداد آزمودنی ها در گروه ها برابر باشد، می توان از نمودار ستونی یا هیستوگرام استفاده کرد ولی اگر تعداد افراد گروه ها برابر نباشد ، باید از این توزیع استفاده کرد.

❖ **نمودار چندضلعی تراکمی (اجایو)** : مانند نمودار چند ضلعی است که با داده های نسبتی و فاصله ای بکار می رود. در محور x کرانه طبقات و در محور y فراوانی تجمعی قرار می گیرد . اگر به جای فراوانی تراکمی درصد فراوانی تراکمی را بصورت هندسی نمایش دهیم حاصل اجایو انتخاب می شود.



شاخص های گرایش مرکزی :

❖ **نما (مد)**

بی ثبات ترین شاخص گرایش مرکزی است که با داده های اسمی بکار می رود و عددی است که دارای بیشترین فراوانی می باشد ،

- مثلا ۹-۵-۶-۶-۶-۶-۶-۶-۶-۶- ۷- ۶- ۵-۷-۶-۵-۱

مد ؟

برخی موارد توزیع ، دو نمایی یا چند نمایی می شود .

مثال : ۷-۶-۵-۵-۴-۲-۳-۲-۱

در اینجا ۲ و ۵ نما هستند که نما دراصل $۳/۵ = \frac{۲+۵}{۲}$

در جدول فراوانی ، نما عبارت است از عددی که F_i آن بزرگتر است

مد؟

F_i	x
۵	۵
۹	۴
۳	۳
۲	۲
۶	۱

✓ **نکته ۱** : برای محاسبه نما ابتدا باید اعداد را از کوچک به بزرگ مرتب کنیم. ✓ **نکته ۲** : در موقعیت هایی که دو عدد مجاور **x** فراوانی یکسان داشته باشند که بزرگتر از فراوانی سایر ارزش های **x** باشند ، نما را می توان بطور قرار دادی به عنوان میانگین دو ارزش مجاور **x** در نظر گرفت

۱ – ۲ – ۳ – ۳ – ۴ – ۴ – ۵ میانه = ؟

❖ **نکته ۳ :** در جایی که ارزشهای غیر مجاور فراوانی های بزرگتر از فراوانی های طبقه های مجاور داشته باشند. هر کدام از این ارزشها را می توان به عنوان نما در نظر گرفت. در چنین شرایطی توزیع دو نمایی نامیده می شود.

❖ **مثال:** مد؟

۱۴ – ۱۴ – ۱۴ – ۱۳ – ۱۳ – ۱۳ – ۱۳ – ۱۳ – ۱۲ – ۱۲ – ۱۲ – ۱۱ – ۱۱

۱۸ – ۱۷ – ۱۷ – ۱۶ – ۱۶ – ۱۵ – ۱۵ – ۱۵ – ۱۵ –

❖ **محاسبه ی نما در داده ای طبقه بندی شده**

از معادله ی زیر محاسبه می شود ، که در آن:

$$MO = L + i\left(\frac{d_1}{d_1+d_2}\right)$$

L : حد واقعی پایین طبقه ای که دارای بیشترین فراوانی است

i : طول یا فاصله طبقات

d_1 : تفاوت فراوانی طبقه نمادار از فراوانی طبقه ی قبل

d_2 : تفاوت فراوانی طبقه نمادار از فراوانی طبقه ی بعد

X	F
۵۴ - ۵۶	۱

➢ مثال :

۵۷ - ۵۹	۳
۶۰ - ۶۲	۶
۶۳ - ۶۵	۸
۶۶ - ۶۸	۲

$$d_1 = 8 - 6 = 2$$
$$d_2 = 8 - 2 = 6$$

طبقه

$$MO = 62/5 + 3\left(\frac{2}{2+6}\right) = 63/25$$

❖ **نکته** : چنانچه توزیع نرمال باشد ، نما از فرمول زیر محاسبه می شود :

$$\boxed{MO = ۳\,Md - ۲\bar{X}}$$

❖ **میانه** :

میانه جایگاهی در توزیع نمره هاست و توزیع نره ها را به دو قسمت مساوی تقسیم می کند ، یعنی جایی است که دقیقا ۵۰ درصد نمره ها بالای آن و ۵۰ درصد نمره ها زیر آن قرار می گیرند . میانه از نما با ثبات تر و از میانگین بی ثبات تر است و با داده های رتبه ای بکار می رود . زیرا ما ابتدا نمره ها را از کوچک به بزرگ مرتب می کنیم.

❖ **طریقه محاسبه میانه نمرات خام در اعداد گسسته**

ابتدا نمره هارا از کوچک به بزرگ مرتب می کنیم و سپس اگر تعداد اعداد فرد باشد ، میانه ، عدد وسط است.

➢

مثال : ۱۲ — ۹ — ۴ — ۳ — ۱ — ۶ — ۵

➢ **مثال** : در توابع اعداد ۳۱ — ۴ — ۵ — ۲۱ — ۳۰ — ۳۰ — ۱۷ — ۸ — ۱۲ میانه برابر است با

❖ طریقه محاسبه میانه اعداد خام در تعداد زوج

۱ – اعداد را از کوچک به بزرگ مرتب می کنیم .

۲ – دو عدد وسط را باهم جمع و تقسیم بر دو می کنیم.

مثال ۱۲ — ۹ — ۲ — ۸ — ۵ — ۳

❖ **نکته** : هنگامی که نمره یا عددی که توزیع را به دو قسمت تقسیم می کند تکراری است ، میانه از طریق محاسبه زیر بدست می آید .

الف : ابتدا حد پایین عدد تکراری که میانه یکی از آنهاست را می نویسیم.

ب : کسری را در نظر می گیریم که مخرج آن تعداد اعداد تکراری و صورت آن نشان دهنده تعداد اعداد تکراری است که در سمت چپ خط رسم کننده میانه قرار می گیرند.

ج : حاصل مراحل (الف) و (ب) را باهم جمع و میانه را بدست می آوریم . (میانه روی عدد ۴ می افتد ، بنابراین نیمی از آن بعلاوه ی یک مورد ۴ قبل از آن جمعا ۱/۵ تا از ۴ ها به حد پایین اضافه می شود) مانند :

میانه = ۶ — ۵ — ۴ — ۴ — ۳

میانه = ۶ — ۵ — ۴ — ۳ ⊥ ۳ — ۳ — ۲ — ۱

❖ محاسبه میانه در جدول اعداد طبقه بندی شده

۱ – ابتدا نمرات را از کوچک به بزرگ مرتب می کنیم.

۲ – سپس توزیع فراوانی و جدول را تشکیل و با فرمول نمرات طبقه بندی شده میانه را محاسبه می کنیم .

مثال :

۳ – ۵ – ۲ – ۴ – ۸ – ۴ – ۴ – ۴ – ۴ – ۲

$$Md = L + \frac{\frac{N}{2} - F_C}{F_i} i$$

L: حد پایین طبقه میانه دار

$\frac{N}{2}$ = مجموع فراوانی تقسیم بر دو

F_C= فراوانی تجمعی طبقه قبل از طبقه میانه

F_i = فراوانی مطلق طبقه میانه دار

i = فاصله طبقاتی

X	F_i	F_C
۸	۱	۱۰
۵	۱	۹
۴	۵	۸
۳	۱	۳
۲	۲	۲
	$\sum F = 10$	

$F_C = ۳$ ، $F_i = ۵$ ، $i = ۱$

➢ **مثال** :

$$Md = L + \frac{\frac{N}{2} - F_C}{F_i} \times i \Rightarrow Md = ۳/۵ + \frac{۵-۳}{۵} \times ۱ \Longrightarrow Md = ۳/۹$$

✓ **نکته** : در صورتی که داده در مقیاس فاصله ای یا نسبتی باشد ، بهترین شاخص ، گرایش مرکزی میانگین است. ولی اگر در توزیعی که نمره ای در کرانه (نمره خیلی بزرگ یا خیلی کوچک) باشد ، (توزیع دارای کجی باشد) میانه شاخص مناسب تری است بعنوان مثال در توزیع ۳۰۰ – ۱۵ – ۹ – ۷ – ۶ – ۵ میانه مناسب تر است.

❖ ویژگی ها

۱ – نسبت به اعداد بزرگ یا کوچک حساس نیست .بنابراین بهترین شاخص است که تمرکز اعداد را در وسط توزیع نشان می دهد. بعنوان مثال در توزیع های زیرکه دارای میانگین های متفاوتی هستند ، میانه برابر و مساوی ۲۰میباشد.

۱۰ – ۱۵ – ۲۰ – ۵۲ – ۶۳ و ۵ – ۷ – ۲۰ – ۲۴ – ۲۵

۲ – مورد استفاده میانه زمانی است که مقیاس اندازه گیری رتبه ای باشد ، هرچند که می تواند برای داده هایی با مقیاس فاصله ای و نسبی هم استفاده شود.

۳ – مجموع قدر مطلق انحراف های نمره ها از میانه کوچکتر یا مساوی مجموع قدر مطلق انحراف های نمره ها از هر عدد دیگری است. (بدون در نظر گرفتن علامت)

$$\sum|X - Md| \le \sum|X - C|$$

نمره ها	قدر مطلق انحرافات				
	میانه (۶)	۴	۵	۷	۹
۴	۲	۰	۱	۳	۵
۵	۱	۱	۰	۲	۴

۳ ۲ ۰	۱ ۰ ۲	۱ ۲ ۴	۲ ۳ ۵	۰ ۱ ۳	۶ ۷ ۹
۱۴	۸	۸	۱۱	۷	

میانگین

باثبات ترین شاخص گرایش مرکزی است که با داده های فاصله ای و نسبی بکار می رود و مرکز ثقل داده هاست.

❖ طریقه محاسبه میانگین اعداد خام

➢ مثال ۱۱ – ۹ – ۱۰

$$\bar{X} = \frac{\sum X}{N} \rightarrow \quad \text{میانگین} = \frac{\text{مجموع نمره ها}}{\text{تعداد نمره ها}}$$

$$\bar{X} = \frac{۱۰ + ۹ + ۱۱}{۳} = ۱۰ \qquad \bar{X} = ۱۰$$

❖ طریقه محاسبه میانگین اعداد طبقه بندی شده با $\mathbf{i} = ۱$

$$\bar{X} = \frac{\sum F.X}{n}$$

$\boldsymbol{X}$	$\boldsymbol{F_i}$

۵	۳
۴	۸
۳	۵
۲	۱۲

❖ **طریقه محاسبه میانگین اعداد طبقه بندی شده**

با i ≠ ۱

$$\bar{X} = \frac{\sum F \cdot X_C}{n}$$

F: فراوانی هر رده یا دسته

n: تعداد کل داده ها یا مجموع فراوانی های تمام دسته ها

X_C =**حد میانی طبقه یا نماینده طبقه**

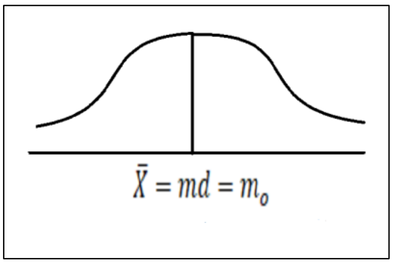
برای پیدا کردن حد میانی نمرات ، دو نمره طبقه را با هم جمع و تقسیم بر ۲ می نماییم.

❁**مثال:**

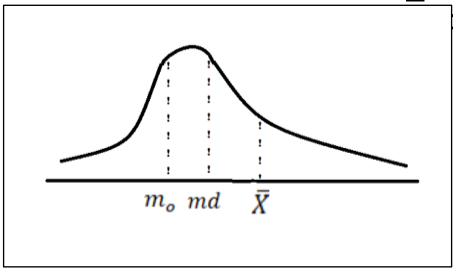
✓ **نکته** :

۱ – اگر میانگین و نما و میانه برابر باشند ، توزیع متقارن است و کجی آن صفر است.

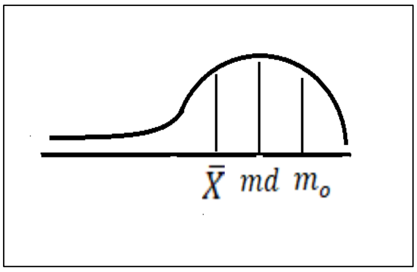
$$\bar{X} = md = m_o$$



۲ – اگر میانگین بزرگتر از میانه و میانه بزرگتر از نما باشد ، کجی مثبت است . یعنی اکثر نمرات پایین بوده و امتحان سخت است. $\bar{X} > md > m_o$



۳ – اگر میانگین کوچکتر از میانه و میانه کوچکتر از نما باشد ، کجی منفی است و اکثر نمرات بالاست و امتحان آسان بوده است. $\bar{X} < md < m_o$



۴ – کشیدگی ارتفاع (y) منحنی را نشان می دهد به برآمدگی یت فرورفتگی منحنی در مقایسه با منحنی طبیعی کشیدگی می گویند.

۵ – در صورتی که یک نمره ثابت (C) در کلیه نمرات یک توزیع ضرب ، تقسیم ، جمع و یا تفریق شود . در هر چهار حالت میانگین تغییر می کند . یعنی میانگین قبلی در آن عدد ثابت ضرب ، تقسیم ، جمع و یا تفریق میشود

جدید $\bar{X}_1 = \bar{X} - C$, $\bar{X}_1 = \bar{X} + C$, $\bar{X}_1 = \bar{X} : C$, $\bar{X}_1 = \bar{X} . C$

۶ – مجموع انحراف نمره ها از میانگین همیشه صفر است. $\sum(X - \bar{X}) = 0$ در واقع شاخصی از مکان مرکزی به معنای حداقل مجذورات است.(فرگوسن ، ناکاته ،۸۱: ۱۳۸۰).

۷ – مجموع مجذور انحرافات از میانگیت همیشه مینیمم است (کمتر از هر عدد دیگری است) .

$$\sum (X-\mu)^2 < \sum (X-a)^2$$

✓ **رابطه شاخص های گرایش به مرکز** :

$$MO = 3md - 2\bar{X} \qquad md = \frac{mo + 2\bar{X}}{3} \qquad \bar{X} = \frac{3md - m_O}{2}$$

میانگین میانگین ها :

در صورتی که حجم نمونه ها برابر باشد یعنی . $n_1, n_2, \dots, n_n$

$$\bar{X}_T = \frac{\sum \bar{X}_\iota}{N}$$

$: \bar{X}_1 = 5 \ , \ \bar{X}_2 = 10 \ , \ \bar{X}_3 = 15 \ , \ n_1 = 10 \ , \ n_2 = 10 \ , \ n_3 = 10$مثال

خواهیم داشت :

$$\bar{X}_T = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{N} = \frac{5 + 10 + 15}{3} = 10$$

در صورتی که حجم نمونه ها برابر نباشدبطریق زیر محاسبه میشود :

$$\bar{X}_T = \frac{\sum \overline{X_\iota}.n_i}{n_i} = \frac{\bar{X}_1 n_1 + \bar{X}_2 n_2 + \bar{X}_3 n_3}{n_1 + n_2 + n_3}$$

➢ **مثال** : $\bar{X}_1 = 5\ ,\ \bar{X}_2 = 10\ ,\ \bar{X}_3 = 15\ ,\ n_1 = 15\ ,\ n_2 = 10\quad n_3 = 20$

میانگین هارمونیک (همساز)

این نوع میانگین برای مواردی بکار می رود ، که مقیاس ترکیبی باشد ، مانن متر بر ثانیه و یا کیلومتر بر ساعت . این میانگین در عینک سازی و مطالعات شبکه های برقی بکار میرود.

$$HM = \frac{1}{\frac{1}{N}\left[\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right]}$$

✓ **مثال** : میانگین سرعت های ۵ و ۶ و ۷ و ۲ کیلومتر بر ساعت ۴ ماشین چند است؟

$$HM = \frac{1}{\frac{1}{4} \times \left[\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{2}\right]} = 3.96$$

میانگین هندسی

نوعی میانگین است که با G نشان داده می سوند و معمولا هر گاه x_1 ها از درصدها یا نسبت ها تشکیل شده باشند ، استفاده می شود ، در کارهای اقتصادی یا جمعیت شناسی بکار می رود. $G = \sqrt[n]{x_1 . x_2 . x_3 (x_n)}$

✓ **مثال** : میزان سود شرکت مهرداد در ۵ سال گذشته برحسب درصد به ترتیب ۲ و۳ و۴ و۴ و۳ شده است .

$$G = \sqrt[5]{3 \times 4 \times 4 \times 2 \times 3} = 3/01$$

رابطه ی بین سه میانگین فوق بصورت زیر است :

$$\bar{X} > G > HM$$

میانگین هارمونیک < میانگین هندسی < میانگین وزنی

- ✓ **نکته** : اگر بین داده ها عدد صفر یا منفی وجود داشته باشد نمی توان از این روش استفاده کرد.

شاخص های پراکندگی

فرض کنید گفته شود درجه حرارت دما در مکانی ۶۵ درجه فارنهایت است . اگر صرفا با نگا کردن به میانگین نتیجه بگیریم که این مکان برای زندگی مناسب است . خطا کرده ایم . زیرا این عدد مقدار متوسط ماه های زمستان و تابستان است. ما نیاز داریم تغییر پذیری درجات حرارت را بدانیم . یعنی درجه حرارت روزانه چقدر از متوسط ۶۵ درجه فارنهایت متفاوت است . از این رو باید شاخص های پراکندگی که نشان می دهد داده ها از میانگین چقدر دور و یا به آن نزدیک هستند را محاسبه کنیم.

شاخص های پراکندگی میزان پراکنده بودن نمرات حول و حوش مرکز داده ها را نشان مدهد که به ترتیب عبارت اند از :

دامنه تغییرات – انحراف متوسط . انحراف چارکی . واریانس و انحراف معیار .

❖ دامنه تغییرات (R)

دامنه تغیرات یک شاخص پراکندگی و در واقع بی ثبات ترین شاخص و حساس ترین شاخص پراکندگی است که با داده های فاصله ای بایستی بکار رود و تفضل بین کوچکترین و بزرگترین عدد در توزیع است (بدون در نظر گرفتن حدود واقعی اعداد) $R = max - \min$ $\quad R = max - \min + 1$

مثال : در توزیع نمرات ۳ – ۲۰ – ۱۴ – ۱۵ – ۹ – ۶ – ۵ دامنه تغییرات را محاسبه کنید :

$$R = 20 - 3 + 1 = 18$$

✓ **نکته** :

۱ – دامنه تغییرات برای نمونه های بزرگ شاخص بی ثباتی است .

۲ – واریانس نمونه برداری دامنه تغییرات برای نمونه های کوچک ، بزرگتر از واریانس نمونه برداری انحراف معیار نیست .ولی سریعا با زیاد شدن ***N*** افزایش می یابد.

۳ – دامنه تغییرات بجز در موارد خاص مستقل از حجم نمونه نیست.

٤ – دامنه تغییرات برای نمونه های کوچک مناسب است.

❖ انحراف چارکی

بیشتر در مواردی کاربرد دارد که نمره ها دارای مقیاس رتبه ای هستند و یا نمره ای در کرانه باشد. پراکندگی را در اطراف مرکز توزیع نشان می دهد. در داده های پرت به جای انحراف استاندارد می شود.

➢ مثال : برای نمرات ۶۰ – ۳۲ – ۱۵ – ۹ – ۶ – ۵ – ۳ انحراف چارکی مناسب است.

Q_3 *md* Q_1

$$SIRQ = \frac{Q_3 - Q_1}{۲} = \frac{\text{چارک سوم} - \text{چارک اول}}{۲}$$

$$Q = \frac{32 - 5}{2} = 13/5$$

برای محاسبه چارک متوسط اعداد طبقه بندی شده نیاز به محاسبه چارک اول و سوم داریم.شرایط استفاده از چارک متوسط (انحراف چارکی) مانند میانه است. بهترین مورد استفاده از انحراف چارکی هنگامی است که **چولگی شدید** است زیرا تحت تاثیر نمرات بالا و پایین قرار نمیگیرد.**این شاخص از دامنه تغییرات کوچکتر است.**

❖ محاسبه چارک ها در اعداد طبقه بندی شده

۱ – اعداد را از کوچک به بزرگ مرتب کنید

۲ – میانه را حساب کنید Q_2

○ سوال : میانه اعداد سمت چپ و راست را محاسبه کنید . $Q_۲$ و $Q_۱$

➢ مثال :۱۷ – ۱۴ – ۱۳ – ۱۲ – ۱۱ – ۹ – ۸ – ۶ $Q_3 = 13/5$ $Q_2 = 11/5$ $Q_1 = 8/5$

چارک یکم : یعنی نقطه ای که یک چهارم افراد زیر آن و سه چهارم بالای آن قرار دارند.

دهک یکم : نمره های یک دهم افراد زیر آن و ۰/۹ افراد از آن بزرگتر است .

صدک یکم : نمره های یک صدم افراد زیر آن و ۰/۹۹ افراد بالای آن قرار دارد.

✓ نکته ۲ : پس از محاسبه نمره چارک اول وسوم انحراف چارکی بدست می آید در صورتی که :

$Q_3 - Q_2 = Q_2 - Q_1$ متقارن

$Q_3 - Q_2 > Q_2 - Q_1$ کجی مثبت

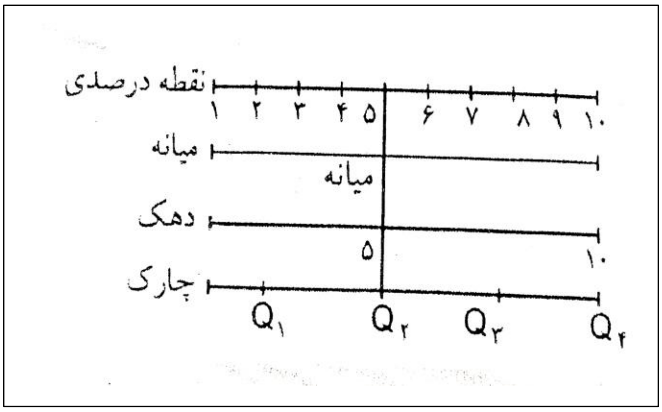
$Q_3 - Q_2 < Q_2 - Q_1$ کجی منفی

❖ **نقطه درصدی μ_n**

نقاط درصدی نمره را بر مقیاس صد نشان می دهد. یعنی نقاط ۵۰ درصدی برابر میانه است و نقطه ۲۵ درصدی برابر چارک اول است و نقطه ۷۵ درصدی برابر چارک سوم است.

$P_{0/75} = Q_3 \qquad P_{0/10} = D_1 \qquad P_{0/20} = D_2 \qquad P_{0/50} = md = Q_2 \qquad P_{0/25} = Q_1$

دهک هم مثل نقطه درصدی ، موقعیت نمره را در مقیاس ۱۰ تایی نشان می دهد . و چارک هم موقعیت نمره را در مقیاس ۲۵ تایی در چهار موقعیت نشان می دهد ، یعنی :



❖ فرمول محاسبه میانه و دیگر نقاط

$$md = L + \left(\frac{\frac{N}{2} - F_c}{F_i}\right) i \qquad Q_1 = L + \left(\frac{\frac{1N}{4} - F_c}{F_i}\right) i \qquad Q_2 = L + \left(\frac{\frac{2N}{4} - F_c}{F_i}\right) i$$

$$D_1 = L + \left(\frac{\frac{1N}{10} - F_c}{F_i}\right) i \qquad D_6 = L + \left(\frac{\frac{6N}{10} - F_c}{F_i}\right) i \qquad P_{40} = L + \left(\frac{\frac{40N}{100} - F_c}{F_i}\right) i$$

➢ **مثال** : چارک سوم ، چارک اول انحراف چارکی دهک دوم ، صدک شصت و هشتم را در داده های زیر محاسبه کنید.

$\boldsymbol{X}$	$\boldsymbol{F_i}$	$\boldsymbol{F_c}$
۲۵ – ۲۹	۱۷	۱۰۰
۲۰ – ۲۴	۲۹	۸۳
۱۵ – ۱۹	۲۱	۵۴
۱۰ – ۱۴	۱۸	۳۳
۹ – ۵	۱۵	۱۵
	$\sum F_i$ = ۱۰۰	

❖ رتبه درصدی

رتبه درصدی مثل نمره های استاندارد است که موقعیت نسبی فرد را در داخل گروه نشان می دهد یعنی با داشتن رتبه درصدی فرد می توانیم بگوییم که او از چند درصد گروه بهتر و یا بدتر عمل کرده است.

$$P_R = \left(\frac{F_c + \frac{F_i}{2}}{n}\right) \times 100$$

✓ **نکته :**

رتبه درصدی موقعیت فرد را در گروه و نقطه درصدی موقعیت نمره فرد را در داخل نمره ها ***(X)*** نشان میدهد.

- مثال : کسی که با نمره ۱۸ ، رتبه درصدی ۸۵ کسب کند ، یعنی نمره ۱۸ (نقطه درصدی) از ۸۵ درصد افراد (رتبه درصدی) بهتر و از ۱۵ درصد افراد بدتر عمل کرده است.

- **مثل** :اگر بخواهیم رتبه درصدی عدد ۲۲ را در جدول فوق مشخص کنیم ، خواهیم داشت :

✓ **نکته ۱ : در توزیع های متقارن اغلب از میانه به عنوان شاخص مرکزی و انحراف چارکی به عنوان شاخص پراکندگی استفاده می شود.**

✓ نکته ۲ :اگر انحراف چارکی اندازه ها برابر صفر باشد ، ۵۰ درصد اندازه هایی که در وسط قرار گرفته اند با هم برابرند در نتیجه چارک های یکم و دوم و سوم با هم برابرند و بر عکس.

- **پرسش** : کدام پراکندگی برای توزیع فراوانی زیر مناسب تر است.؟

حدود طبقات	۵۰ - ۲۰	۴۰ - ۲۰	۶۰ – ۴۰	۶۰ و بیشتر
فراوانی	۲۵	۳۵	۳۰	۱۰

الف) انحراف چارکی ب)انحراف معیار ج) ضریب تغییرات د) انحراف متوسط

- **انحراف متوسط *(MD)***

انحراف متوسط یک شاخص پراکندگی است که به آن میانگین قدر مطلق انحراف نمره از میانگین میگویند

$$MD = \frac{\sum|X-\bar{X}|}{N}$$

➢ **مثال** : برای اعداد ۵ – ۴ – ۳ – ۲ – ۱ خواهیم داشت.

X	$X-\bar{X}$	$\lvert X-\bar{X}\rvert$
5	2	2
4	1	1

3	0	0
2	-1	1
1	-2	**2**
	$\sum \lvert X - \bar{X} \rvert = 6$	

نکات مهم

۱ – با این شاخص عملیات جبری را نمی توان انجام داد .

۲ – در انحراف متوسط علایم اعداد و در انحراف چارکی کلیه اعداد مورد مطالعه قرار نمی گیرند.

۳ – تاثیر انحرافات بزرگ را در شرایطی که تعداد زیادی انحرافات کوچک در برابر تعداد کمی انحراف بزرگ باشد.

نشان می دهد (مهمترین کاستی)

۴ – انحراف متوسط اعداد ثابت صفر است.

واریانس (S^2)

یک شاخص پراکندگی است که میانگین مجموع مجذورات انحراف نمره از میانگین می باشد.

یعنی :

$$S^2 = \frac{\sum(X-\bar{X})^2}{n} = \frac{\sum(X)^2}{n} = \frac{ss}{n} = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n}$$

- **محاسبه واریانس از راه انحراف نمره از میانگین**

$$S^2 = \frac{\sum(X-\bar{X})^2}{n}$$

- **مثال** : برای اعداد ۵ – ۴ – ۳ – ۲ – ۱

X	$X-\bar{X}$	X^2
5	2	4
4	1	1
3	0	0
2	-1	1
1	-2	4
	$\sum\lvert X-\bar{X}\rvert = 0$	$\sum \mathbf{X^2} = \mathbf{10}$

- **محاسبه واریانس از راه نمرات خام**

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n}$$

- **نکته مهم** : هنگامی که واریانس را در نمونه محاسبه می کنیم در مخرج کسر n-1 که به آن درجه آزادی (dt)گفته می شود استفاده می کنیم.
- درجه آزادی : تعداد مشاهداتی است که می تواند آزادانه تغییر کند بنابر این درجه آزادی واریانس df=n-1 است.

- ✓ **مثال** : واریانس ۵ عدد متوالی در نمونه برابر ۲/۵ و واریانس ۵ عدد متوالی در جامعه برابر ۲ می باشد.

X	X^2
5	25
4	26
3	9
2	4
1	1
$\sum X = 15$	$\sum X^2 = 55$

❖ **انحراف معیار (انحراف استاندارد) s**

جذر واریانس را انحراف معیار می گویند ،

یعنی در مثال فوق :

- ✓ نکته **توان دوم انحراف معیار . واریانس نام دارد.**
- ✓ **انحراف معیار با ثبات ترین شاخص پراکندگی است.**

- ✓ هر چقدر انحراف استاندارد بیشتر باشد پراکندگی بیشتر است. شاید اساسی ترین فایده انحراف استاندارد این باشد که با استفاده از آن می توان مشخص کرد چه نسبتی از نمره ها در فاصله های مختلف نسبت به میانگین قرار گرفته است.
- ✓ نکته مهم : **زمانی که توزیع دارای کجی زیاد باشد از انحراف استاندارد باید با احتیاط استفاده شود . زمانی از** S **استفاده می شود.که میانگین بعنوان شاخص مرکزی مورد استفاده قرار میگیرد.کلیه شاخص های پراکندگی (دامنه تغییرات . انحراف متوسط . انحراف معیار) با مقیاس حداقل فاصله ای بکار میروند.**

ثبات پراکندگی به ترتیب : از بیشتر به کمتر

انحراف معیار ⟵ واریانس ⟵ انحراف متوسط ⟵ چارک متوسط ⟵ دامنه تغییرات

$$R \leftarrow Q \leftarrow MD \leftarrow S^2 \leftarrow S$$

$$S > MD > Q$$

- **نکته مهم :**
- **تاثیر چهار عمل اصلی در واریانس**

در شرایط **جمع و تفریق** یک عدد ثابت در نمرات یک توزیع **واریانس تغییر نمی کند.**

تفریق $S_1{}^2 = S^2$ جمع $S_1{}^2 = S^2$

ولی در شرایط **ضرب وتقسیم** واریانس قدیم **در توان دوم مجذور عدد ثابت ضرب یا تقسیم** می شود.

تقسیم $S_1{}^2 = S^2 : C$ ضرب $S_1{}^2 = S^2 . C$

- **مثال** : اگر واریانس یک توزیع ۲ باشد ، و عدد ۳ در در کلیه نمرات توزیع ضرب شود و واریانس جدید برابر خواهد بود با :

- **مثال**:اگر واریانس ۵ و عدد ثابت ۶ با کلیه نمرات جمع شود واریانس جدید برابر است با :

تاثیر چهار عمل اصلی در انحراف معیار

در شرایط جمع و تفریق مثل واریانس است : یعنی انحراف معیار تغییری نمی کند ، ولی در شرایط ضرب و تقسیم مثل میانگین است ، یعنی در همان عدد ثابت ضرب یا تقسیم میشود.

تفریق $S_1 = S$ جمع $S_1 = S$

تقسیم $S_1 = S:C$ ضرب $S_1 = S.C$

➢ **مثال** : اگر انحراف معیار یک توزیع ۶ و یک نمره ثابت ۲ در کلیه نمرات توزیع ضرب شود ، انحراف معیار جدید برابر است با :

اگر انحراف معیار یک توزیع ۶ و نمره ثابت ۳ با کلیه نمرات توزیع جمع شود انحراف معیار جدید برابر است با : $S = ۶$

(تغییر نمی کند)

❖ **نکات مهم**

۱ – میانه و نما در جمع و تفریق و ضرب و تقسیم یک عدد ثابت تابع میانگین بوده و مثل هم تغییر می کنند.

۲ – در جمع و تفریق ، واریانس – انحراف معیار – انحراف متوسط- دامنه تغییرات و انحراف چارکی مثل هم بوده و تغییر نمی کنند.

۳ – در شرایط ضرب و تقسیم ، انحراف معیار – انحراف متوسط- انحراف چارکی و دامنه تغییرات در همان عدد ضرب یا تقسیم می شوند.

۴ – واریانس در مجذور آن عدد ضرب یا تقسیم می شود.

۵ – واریانس اعداد ثابت صفر است.

۶ – اگر چند جامعه با هم ترکیب شوند ، میانگین و واریانس جامعه کل ترکیبی ،بزرگتر از میانگین و واریانس جوامع تشکیل دهنده خواهد بود . مگر آن میانگین جوامع برابر باشد که در آن صورت واریانس آنها هم برابر است .

۷ – واریانس تفاوت ها در نمونه های همبسته برابر است با :

$$S^2{}_{y_1} - S^2{}_{y_2} = S^2{}_{y_1} + S^2{}_{y_2} - 2\rho S_{\ y_1}\ S_{\ y_2}$$

۸ – واریانس تفاوت ها در نمونه های مستقل برابر است با :

$$S^2{}_{\bar{X}_1 - \bar{X}_2} = S_{\overline{x_1}}{}^2 + S_{\overline{x_2}}{}^2 = \frac{S_{\ 1}{}^2}{N_1} + \frac{S_{\ 2}{}^2}{N_2}$$

۹ – واریانس مجموع برابر است با :

$$S^2{}_{(x+y)} = S_x{}^2 + S_y{}^2 \pm 2Cov(X,Y) = (S_x{}^2 + S_y{}^2)2rS_x\ .S_y$$

❖ تصحیح شپرد

فرمول شپرد برای تصحیح انحراف معیار زمانی کاربرد دارد که فاصله طبقاتی بزرگ و تعداد طبقات کمتر از ۱۲ باشد.

$$S_C = \sqrt{S^{\,2} - \frac{i^2}{12}}$$

❖ **مثال** : اگر فاصله طبقاتی ۱۰ و انحراف معیار توزیع ۵ باشد ، انحراف معیار تصحیح شده برابر خواهد بود با :

❖ ضریب پراکندگی (V)

همان ضریب نسبی واریانس است که بر اساس آن پراکندگی ویژگی یک گروه را با پراکندگی ویژگی دیگر همان گروه مقایسه می کنند.

$$V = \frac{S}{\bar{\bar{X}}} \times 100 = \frac{\text{انحراف معیار}}{\text{میانگین}} \times 100$$

❖ **مثال** : کارخانه ای دو نوع لاستیک اتومبیل تولید می کند. برای نوع الف میانگین عمر ۱۰۰۰۰ کیلومتر ، با انحراف استاندارد ۲۰۰۰ کیلومتر و برای نوع ب میانگین عمر ۱۱۰۰۰ کیلومتر با انحراف استاندارد ۱۰۰۰ کیلومتر می باشد.کدام نوع لاستیک بهتر است ؟

❖ موارد استفاده ضریب پراکندگی

زمانی که دو یا چند جامعه در مقایسه باهم دارای **مشاهدات ناهمگون** از نظر واحد اندازه گیری باشند مانند یک جامعه بر حسب متر و یک جامعه برحسب اینچ و یا

جامعه دارای میانگین های متفاوتی باشند استفاده می شود. گاهی نیز مقیاس صفت مورد اندازه گیری در دو جامعه یکسان است ، ولی بزرگی مشاهدات آنها بطور قابل ملاحظه ای تفاوت دارد مانند مقایسه پراکندگی سود و زیان در صنایع دستی با صنایع سنگین.

❖ **نکته ۱ : ضریب تغییرات اعداد ثابت برابر صفر است.**

- **پرسش** : برای تعیین آنکه در ۳۰ روز گذشته به نسبت ، قیمت دلار از ثبات بیشتری برخوردار بوده است ، یا یورو استفاده از کدام شاخص آماری مناسب تر است ؟

الف) انحراف متوسط ب) ضریب پراکندگی ج) ضریب چولگی د) واریانس

گشتاوردهای پیرامون میانگین

گشتاور اول : $m_1 = \frac{\sum(X-\bar{X})}{N} = 0$

گشتاور دوم : $\mathrm{m_2} = \frac{\sum(\mathrm{X}-\bar{\mathrm{X}})^2}{\mathrm{n}} = \mathrm{S}^2$

گشتاور سوم : $\mathrm{m_3} = \frac{\sum(\mathrm{X}-\bar{\mathrm{X}})^3}{\mathrm{n}}$

گشتاورچهارم : $m_4 = \frac{\sum(X-\bar{X})^4}{n}$

گشتاور اول همیشه صفر است.

گشتاور دوم همان واریانس است.

گشتاور سوم برایر محاسبه چولگی (کجی) بکار می رود. یعنی :

$$\mathrm{SK} = \frac{\mathrm{m_3}}{\mathrm{m_3}\sqrt{\mathrm{m_3}}}$$

گشتاور چهارم برای محاسبه کشیدگی بکار میرود یعنی: $Kg = \frac{m_4}{(m_3)^2} - 3$

> ❖ نکته: **کشیدگی معیاری است بدون واحد که ارتفاع را نشان می دهد و رابطه معکوس با پراکندگی دارد.**

اگر k=0 باشد، کشیدگی هم اندازه و هم ارتفاع توزیع نرمال است.اگر $k > 0$ باشد از نرمال بزرگتر و پراکندگی کمتر و اگر $k < 0$ باشد، از رمال کوتاه تر و پراکندگی بیشتر است.

❖ **فرمول کجی پیرسون**

$$g_1 = \frac{\bar{X} - m_0}{s}$$

۱ - اگر $|sk| \leq 0/1$ باشد تقریبا کجی وجود ندارد و جامعه نرمال است.

۲ – اگر $0/1 \leq |sk| \leq 0/5$ باشد ، چولگی موجود اندک ولی غیر قابل اغماض است . در حقیقت جامعه از نظر تقارن اندکی با توزیع نرمال متفاوت است.

۳ – اگر $|sk| > 0/5$ باشد ، چولگی زیاد ولی غیر قابل اغماض است بعبارت دیگر جامعه از نظر قرینگی دارای تفاوت فاحشی با توزیع نرمال است.

> ❖ **نمره های استاندارد**

> نمونه های استاندارد موقعیت فرد را در گروه معین می کنند ، با داده های مقیاس فاصله ای کاربرد دارند . در تبدیل نمرات خام به نمرات استانداراد از فرمول مقابل استفاده می کنیم . :
>
> $$\boldsymbol{Z = \frac{X - \bar{X}}{s}}$$

❖ **مثال** : در امتحانی که میانگین ۳۸ و انحراف معیار آن ۳ باشد ، کسی که نمره ۴۴ گرفته است .دارای Z=? می باشد.

در تبدیل نمره Z به دیگر نمرات استاندارد از فرمول مقابل استفاده می کنیم :

$$X = \bar{X} - Z_{\frac{a}{2}} S_{\overline{X}}$$

❖ **مثال** : کسی که نمره او Z=2 باشد ، نمره t او برابر چند است

نمره های استاندارد	میانگین	انحراف استاندارد
Z	۰	۱
T	۵۰	۱۰
CEEB	۵۰۰	۱۰۰
نه گانه	۵	۲
هوش وکسلر	۱۰۰	۱۵
هوش بینه	۱۰۰	۱۶
AGCT ارتش آمریکا	۱۰۰	۲۰

سپس در تبدیل نمرات به نمرات استاندارد دیگر به روش زیر عمل می کنیم :

هوش وکسلر $= 100 + 15z$ $\qquad t = 50 + 10\,z \qquad CEEB = 500 + 100\,z$

نه گانه $= 5 + 2z$ $\qquad$ هوش بینه $= 100 + 16\,z \qquad AGCT = 100 + 20\,z$

❖ مثال :

۱ - در آزمون $CEEB$ فردی ۸۰۰ گرفته است. نمره Z او چقدر است ؟

۲ – فردی که در یک آزمون نمره نه گانه او ۳ شده است . نمره Z او چقدر است ؟

۳ – نمره IQ فردی ۱۴۰ شده است . نمره Z او چقدر است ؟

۴ – نمره t فردی ۳۰ شده است. نمره Z او چقدر است ؟

۵ – نمره Z فردی ۲ شده است. نمره $CEEB$ او چقدر است ؟

۶ – نمره Z فردی ۱/۵ می باشد.نمره نه گانه ی او چقدر است ؟

۷ – نمره Z فردی ۲- است. نمره AGCT او چقدر است ؟

۸ – نمره Z فردی ۱/۵- شده است. نمره t او چقدر است ؟

۹ – نمره $CEEB$ فردی ۳۵۰ شده است.نمره هوشبهر او چقدر است ؟

۱۰ – نمره نه گانه فردی ۸ شده است ، نمره t او چقدر است ؟

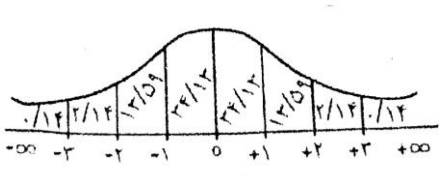
۱۱ – نمره IQ فردی ۴۵ شده است. نمره t او چقدر است ؟

❖ منحنی طبیعی و سطوح زیر منحنی

منحنی طبیعی (زنگوله ای یا گوس) یک توزیع طبیعی است

۱ – شکل آن به میانگین و انحراف استاندارد بستگی دارد.

۲ – میانگین آن صفر و انحراف استاندارد آن یک می باشد



۳ – در عمل یک منحنی طبیعی داریم و بالاترین ارتفاع

۴ – در منحنی طبیعی میانگین و میانه و نما باهم برابر هستند.

بطوریکه میبینیم مشخص میشود که ۳۴/۱۳ درصد افراد بین ۱+ تا $\bar{X}$ قرار می گیرند. همینطور ۱۳/۵۹ درصد افراد بین ۲+ تا ۱+ قرار میگیرند.و ۲/۱۴ درصد افراد بین ۳+تا ۲+ و ۰/۱۴ درصد افراد بین $+\infty$ تا ۳+ قرار می گیرند. این درصدها برای طرف چپ منحنی به دلیل تقارن هم صادق است. اگر بخواهیم رتب درصدی افراد را بر اساس منحنی نرمال محاسبه کنیم به روش زیر عمل می کنیم :

اگر داده (Z) یک عدد باشد توجه می کنیم ، مثبت است یا منفی . در صورت مثبت بودن با رتبه درصدی اولیه ۵۰ جمع می کنیم ،و اگر منفی بود از رتبه درصدی ۵۰ کم می شود.

رتبه درصدی $z = 1$ چقدر است ؟ ۵۰ +

۳۴/۱۳ → ۸۴/۱۳

رتبه درصدی z = -1 چقدر است ؟ $۵۰ -$

$۳۴/۱۳ \rightarrow ۱۵/۸۷$

درصورتی ک سطوح بین دو نمره z خواسته شود ، دیگر ۵۰ کاربردی ندارد و سطوح زیر منحنی محاسبه می شود.

❖ ***مثال*** : سطوح زیر منحنی بین $\pm$۱ چقدر است ؟ درصد $۳۴/۱۳ +$

$۳۴/۱۳ \rightarrow ۶۸/۲۶$

سطوح دیگر عبارتند از :

$\bar{X}$ تا $+ ۳ \rightarrow$ درصد ۴۹/۸۶

$\bar{X}$ تا $+ ۲ \rightarrow$ درصد ۴۷/۷۲

$\bar{X}$ تا $- ۳ \rightarrow$ درصد ۴۹/۸۶

$\bar{X}$ تا $- ۲ \rightarrow$ درصد ۴۷/۷۲

❖ سطوح بین مثبت و منفی

-۲ تا $+ ۲ \rightarrow$ درصد ۹۵/۴۴

-۳ تا $+ ۳ \rightarrow$ درصد ۹۹/۷۷

$+۱$ تا $- ۲ \rightarrow$ درصد ۴۹/۸۶

در صورتی که بصورت اعشاری باشد به روش زیر عمل می کنیم :

رتبه درصدی $z = 1/5$ چقدر است ؟

رتبه درصدی $z = -2/5$ چقدر است ؟

رتبه درصدی $z = 1/02$ چقدر است ؟

سطوح زیر منحنی بین $Z = 1/75$ تا $Z = -1/5$ چقدر است ؟

- **نکات مهم** : ۱ – استفاده از روش بالا حدود رتبه درصدی را نشان می دهد برای محاسبه دقیق باید از جدول مربوط که سطوح بین z های مختلف را ارایه کرده استفاده کرد. (پیوست)

۲ – انتقال از نمره های خام به نمره تراز z شکل نمره را تغییر نمی دهد .

۳ – اگر توزیع نمره های خام دارای کجی باشد توزیع z نیز دارای کجی است.

۴ – برخلاف رتبه های درصدی اختلاف در نمره های z اختلاف در نمره های خام را نشان می دهد.

۵ – نسبت اختلاف نمره ها در توزیع اصلی یا نمره های خام ، مساوی نسبت اختلاف بین نمره های z آنها است. بنابراین فاصله ی بین اندازه نمره های اصلی در تبدیل به نمره های z تغییر نم کند.

۶ – در نمره های نه گانه ۹ فاصله یا واحد وجود دارد که هر فاصله یا واحد آن مساوی نصف واحد انحراف معیار است. واحد میانی (پنجمین نمره نه گانه) نقطه میانی توزیع را نشان می دهد که در بر گیرنده ۲۰ درصد موارد است .

چهارمین ، سومین ، دومین و نخستین نمره نه گانه از مرکز توزیع به سمت کرانه ی انتهایی پایین توزیع به ترتیب ۱۷،۱۲،۷،۴ درصد از موارد را در بر میگیرد. (و بر عکس) نمره ی نه گانه مانند رتبه درصدی نقطه ای برروی مقیاس نیست.بلکه دربرگیرنده ی محدوده ثی وسیعی از توزیع است.به همین دلیل نسبت به اختلاف های قابل توجه در طول مقیاس حساس نیست

،داما سبت به اختلاف های موجود در کرانه های انتهایی توزیع بصورت گمراه کننده ای حساس است (آیزاک ، ۱۳۸۴ : ۱۱۲)

۷ – نمره های کلاسی مانند نمره های هوشی دارای توزیع به هنجار هستند وبیشتر برای دوره ی ابتدایی که مطالب از پیوستگی بیشتری برخوردرند مناسب است. (آیزاک ، ۱۳۸۴ : ۱۱۵)

❖ **همبستگی**

همبستگی بررسی رابطه ی بین دو متغیریا چند متغیر است. یعنی به دنبال این امر هستیم که آیا افزایش یا کاهش یک متغیر تاثیری بر متغیر های دیگر دارد.

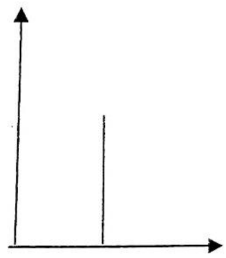
فنون توصیفی در اینگونه مطالعات (که در آنها ارتباط بین متغیرها بدون آنکه هیچ یک از آنها دستکاری یا کنترل شود ، مورد بررسی قرار میگیرد.) عبارتست از

میزان همخوانی (مانند ضریب فای یا ضریب مشابه دیگر) برای مواردی که دو متغیر اندازه گیری شده طبقه ای باشند . **ضریب رتبه ای اسپیرمن** برای مواقعی که دو متغیر مورد اندازه گیری در مقیاس ترتیبی بیان می شوند. و سرانجام ضریب **همبستگی پیرسون** برای مواقعی که دو متغیر مورد بحث دارای مقیاس فاصله ای باشند هر کدام از اینها نشان دهنده ی قدرت ارتباط بین دو متغیر است. (هومن ۱۳۸۲: ۱۲۷) .

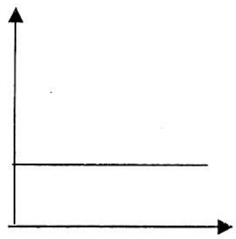
اگر هدف بررسی یک مغییر مثل x روی متغیر دیگر مثل y باشد بررسی همبستگی تک متغیری است . اصطلاحا به پژوهش های تک متغیری موقعیت های مصنوعی گفته می شود. اگر هدف بررسی چند متغیر مثل $x_1, x_2, \dots, x_n$ برروی یک متغیر دیگر یعنی y باشد بررسی همبستگی چند متغیری است.اگر هدف بررسی چند متغیر $x_1, x_2, \dots, x_n$ برروی چند $y_1, y_2, \dots, y_n$ همبستگی کانونی است. در این مطالعات متغیر مستقل را تمغیر پیش بینی کننده و متغیر وابسته را متغیر پیش بینی شونده گویند.

جهت همبستگی و شدت وابستگی

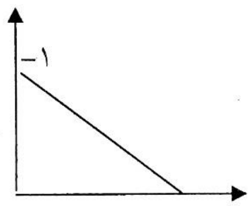
❖ **جهت همبستگی** : اگر افزایش یک متغیر با افزایش متغیر دیگر همرا ه شود یا کاهش یک متغیر با کاهش متغیر دیگر همراه شود همبستگی مستقیم است. ولی اگر افزایش یک متغیر با کاهش متغیر دیگر یا کاهش یک متغیر با افزایش متغیر دیگر همراه باشد. همبستگی منفی است. هنگامی که شیب پراکندگی (خط رگرسیون) افقی یا عمودی باشد. همبستگی صفر است.



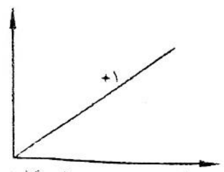
همبستگی صفر



همبستگی صفر



همبستگی کامل منفی

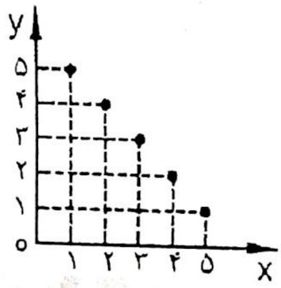


همبستگی مستقیم خطی کامل

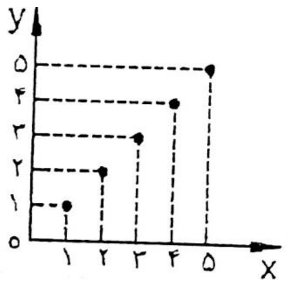
❖ **شدت همبستگی :**

دامنه همبستگی ۱+ و ۱- یعنی همبستگی ۱- معکوس و کامل ، همبستگی ۱+ مستقیم و کامل است. مقدار همبستگی صفر تا ۱ می باشد. (شیولسون، ۱۳۶۰ :۱۷۶). هر داده ای بین این دو حد ناقص است.یعنی همبستگی ۰/۹۹ ناقص است . هرچه همبستگی به قدر مطلق ۱ نزدیکتر باشد شدت آن بیشتر است. مثال : همبستگی ۰/۹۰- ازهمبستگی ۰/۷۰+ شدت بیشتری دارد.

✓ **نکته** : اگر ضریب همبستگی از ۰/۷۰+ بالاترر باشد میگوییم رابطه خطی است. اگر پایین تر باشد رابطه غیر خطی است و در صورت غیر خطی بودن رابطه بین دو متغیر از ضریب همبستگی ایتا η استفاده کنیم.



X	Y
٥	۱
٤	۲
۳	۳
۲	٤
۱	٥



نام	X	Y
حسن	٥	٥
علی	٤	٤
احمد	۳	۳
مهدی	۲	۲
محمد	۱	۱

- ✓ **نکته** ۱ : برای محاسبه ی همبستگی ابتدا باید نمودار پراکندگی را رسم کرد. اگر نقاطبه خط فرضی نزدیک باشند. همبستگی زیاد است.
- ✓ **نکته** ۲ : چهار عمل اصلی $(+, -\times, \div)$ یک نمره در کل نمرات توزیع هیچ تاثیری روی ضریب همبستگی ندارد مثلا :اگر همبستگی بین x وy (قد و وزن) ۰/۵ باشد همبستگی بین 2x و y همان ۰/۵ خواهد بود.

x	y
۱۵۰	۱۴۵
۱۶۰	۱۴۰
۱۷۰	۱۷۳
۱۶۰	۱۴۲
۱۵۰	۱۴۱

$r_{xy} = -۱$

x	y
۱۵۰	۱۱۲
۱۶۵	۱۰۸
۱۷۵	۱۰۳
۱۸۰	۱۰۱
۱۹۲	۹۹

$r_{xy} = -۱$

انواع ضریب همبستگی

الف : در حالتی که دو متغیر اسمی و یا یک اسمی و دیگری رتبه ای باشند : ضرایب عبارتند از کرامر ـ توافق c -لاندا – تاو گودمن و کراسکال.

همبستگی کرامر : (جدول از 2×2 بیشتر باشد و حداقل یکی از متغیرها چند ارزشی است) .رابطه متقارن نیست.

مانند بررسی این فرض : بین جنس دانشجویان و گرایشات سیاسی آنها رابطه وجود دارد جنس اسمی و گرایشات سیاسی هم متغیر اسمی.(بی طرف ، تندرو ، محافظه کار) معادله عبارت است از : $v = \sqrt{\frac{x^2}{n(L-1)}}$ که n تعداد حجم نمونه و L تعداد ردیف یا ستون هر کدام که تعداد کمتری دارند.

گرایش سیاسی / جنسیت	بی طرف	تندرو	محافظه کار	جمع
خانمها	۳۲	۲۲	۵	۵۹
آقایان	۵	۳۰	۱۰	۴۵
جمع	۳۷	۵۲	۱۵	۱۰۴

✓ **نکته** :هنگامی که تعداد کمتر سطر یا ستون برابر ۲ باشد. فرمول بالا بعلت اینکه مقدار L برابر یک می شود بصورت زیر که به همبستگی فی معروف است تبدیل میشود.

$$\varphi = \sqrt{\frac{x^2}{n}}$$

❖ **همبستگی c توافقی**: (جدول از 2×2 بیشتر باشد و حداقل یکی از متغیرها چند ارزشی است) معادله ی آن عبارت است از ؛ $c = \sqrt{\frac{x^2}{x^2 + n}}$

❖ **ضریب همبستگی لاندا :** این ضریب تفسیر روشن تری از همبستگی می دهد . شما تا چه اندازه می توانید از روی جنس افراد گرایشا سیاسی آنها را پیش بینی کنید و یا بر عکس. این ضریب امکان چنین پیش بینی را می دهد .

معادله آن عبارت است از : $\gamma = \frac{e_1 - e_2}{e_1}$ در این رابطه e_1 اشتباه گروه بندی در موقعیت اول و e_2 اشتباه گروه بندی در موقعیت دوم می باشد.

ب : ضریب همبستگی در حالتی که هر دو متغیر دارای مقیاس رتبه ای باشند : عبارت اند از گاما – تاو کندال b– تاو کندال c– ضریب d سامرز .

✓ **نکته** : همه ی این ضرایب قرینه هستند بدین معنا که با تغییر متغیر مستقل و وابسته در مقدار ضریب تییری ایجاد نمی شود.

ضریب گاما (G): رابطه ی بین دو متغیر دو مقوله ای ترتیبی را بدست میدهد مانند رابطه ی سطح تحصیلات مادران با نگرش آنان نسبت به تحصیلات دختران .

ضریب لامبدا (λ) : رابطه ی بین دو متغیر نا متقارن است و رابطهی ماقبل و مابعد را در یک توالی از رفتارها مطرح می کند ، به ما می گوید متغیر a تا چه حد متغیر b

را پیش بینی می کند . عکس آن ممکن نیست (سرمد و دیگران ، ۱۳۸۰ : ۲۲۳)
مقدار آن بین صفر و یک است.

- **ضریب کپا (*k*)** : اگر چند متغیر اسمی و رابطه متقارن باشد از این ضریب استفاده میشود. مثلا اگر k داور N شیئی را به m مقوله اسمی تبدیل کنند ، آمار کپا میزان توافق داورها در مورد اینگونه طبقه بندی را نشان میدهد. مقدار آن بین صفر ویک است.

- **ضریب هماهنگی (w) با توافق کندال (u) :** در پژوهش هایی که در آنها بیش از دو مجموعه رتبه وجود دارد و مایلیم بدانیم که رتبه هایی که توسط m داور به n فرد داده شدخه تا چه حد توافق وجود دارد ، از این دو شاخص استفاده می شود. وقتی داده ها بصورت زوج های همتا و نه بصورت رتبه جمع آوری شده باشد* روش ضریب توافق کندال (u) مناسب تر است (سرمد ودیگران ، ۱۳۸۰ : ۲۲۴) مقدار w بین صفر و یک هست مقدار u اگر k زوج باشد $\left(\frac{-}{k}-1\right)$ و اگر k فرد باشد $\left(\frac{-}{k}\right)$ حداکثر برابر یک است.

m = تعداد دوران

$$\omega = \frac{ss_r}{\frac{1}{12} - m^2(n^2-n)}$$

n : تعداد افراد رتبه بندی شده

ss_r : واریانس مجموع رتبه ها

آزمون تقریبی معنا دار بودن ω نیز با استفاده از مشخصه آماری زیر بدست می آید :

$$X^2{}_{ab} = m(n-1)\omega$$

✓ **نکته** : هر گاه پژوهشگر بخواهد اثر متغیر سوم را ثابت نگه دارد و میزان رابطه ی دو متغیر دیگر را مشخص کند. از **ضریب همبستگی رتبه ای تفکیکی کندال** باید استفاده کند (سرمد و دیگران ، ۱۳۸۰ :۲۲۴)

همبستگی گشتاوری پیرسون

زمانی استفاده می شود که :

۱ – هر دو متغیر توزیع نرمال داشته باشند

۲ – پراکندگی نمرات در هر دو متغیر یکسان باشد.

۳ – توزیع خطی باشد.

۴ – مقیاس متغیر ها حداقل فاصله ای باشد (یا نسبی) مهمترین مفروضه این است که هر دو متغیر دارای حداقل مقیاس فاصله ای باشند .

مثل : قد و وزن هوش و پیشرفت تحصیلی

فرمول مهم

$$r_{xy} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{S_x.S_y} = \frac{cov}{S_x.S_y} = \frac{Z_x.Z_y}{n}$$

ضریب همبستگی رتبه ای اسپیرمن

هنگامی که فقط دو دسته رتبه وجود داشته باشد . (تعداد رتبه هاکم) از ضریب اسپیرمن و یا تاو کندال می توان استفاده کرد.ضریب اسپیرمن از زوج های همتا استفاده می کند .زمانی کاربرد دارد که داده ها رتبه ای باشد. یکی از متغیرها یا هر دو متغیر و یا اینکه داد ها فاصله ای ، ولی مفروضه های آمار پارامتریک رعایت نشده باشد. مثل ترتیب تولد و رتبه در کلاس . وقتی رابطه غیر خطی باشد. داده های فاصله ای باید به رتبه تبدیل شوند .(شیولسون، ۱۳۸۰،۱۹۰) این ضریب برای رابطه های کاملا غیر خطی مناسب نیست.(شیولسون، ۱۳۸۰،۱۹۶)

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} \qquad D = R_x - R_y$$

X خلاقیت	Y پیشرفت تحصیلی	R_x	R_y	D	D^2
۱۲۰	۱۸				
۱۱۲	۱۹				
۱۱۸	۱۲				
۱۲۵	۱۷				
۱۰۱	۱۹/۵				

✓ **نکته** : زمانی که اعداد تکرار می شوند برای تعیین رتبه اعداد مشابه باید بشرح زیر عمل کرد :

X خلاقیت	Y پیشرفت تحصیلی	R_x	R_y	D	D^2
۱۸	۲۹			–	
۱۹	۴۲				
۱۷	۳۵				

۱۸	۳۵				
۲۵	۲۸				
۱۲	۲۱				
۲۳	۳۵				
۱۱	۶۶				
۱۸	۱۱				
۱۷	۱۴				

نکته : اگر $n \leq 30$ از جدول استفاده کنید .اگر $n \geq 30$ از فرمول زیر مقدار بحرانی را بیابید.اگر بین مقدار مثبت و منفی بود همبستگی وجود ندارد

$$r_s = \frac{\pm z}{\sqrt{n-1}}$$

.

در مثال قبل اگر فرض کنیم افراد نمونه ۶۵ نفر هستند بنابراین داریم :

$r_s = \frac{\pm}{\sqrt{۶۵-1}} = \pm ۰/۲۴۵$ اگر آماره آزمون ۴۰ $= r_s$ باشد چون خارج از دامنه ی $\pm ۰/۲۴۵$ است. فرض صفر رد می شود. یعنی بین رتبه ها همبستگی وجود دارد.

ضریب همبستگی دورشته ای نقطه ای

زمانی مورد استفاده قرار می گیرد که یک متغیر پیوسته و دیگری دو ارزشی واقعی باشدیا یک متغیر پیوسته از دیگری دو ارزشی و فرض نرمال بودن توزیع رعایت نشده باشد

مثال: : همبستگی بین نمره کل آزمون و یک سوال با جنسیت و نمره در یک آزمون که می تواند از یک کمتر باشد ، تابع p وq است. در حالت بیشینه مقدار آن از ۱ کمتر است .

$r = ۰/۷۹۷۸$

۱ – ضریب همبستگی دورشته ای : یک متغیر پیوسته و دیگری دو ارزشی ساختگی است. می تواند از یکی بیشتر باشد . توزیع نرمال و رابطه ی خطی است .

منظور از ساختگی یعنی مثلا ما نمرات ریاضی 15-15-9-8 را بصورت مردود – قبول طبقه بندی کنیم . (۱،۰)

X	X'
۱۵	۱
۱۵	۱
۹	۰
۸	۰

۲ – ضریب همبستگی فی : زمانی کاربرد دارد که هر دو متغیر دو ارزشی واقعی باشند . یا هر دو متغیر دو ارزشی و فرض نرمال بودن رعایت نشده باشد. و رابطه متقارن باشد . یعنی جای متغیر ها عوض شود نتیجه تغییر نمی کند.

مانند رابطه جنسیت و وضع سواد (بی سواد –باسواد) و یا همبستگی بین دو سوال (در کتاب خانم سرمد برااین مورد تتراکوریک پیشنهاد شده است .ص ۲۲) که البته بیشتر فی بکار میرود .

۳ – ضریب همبستگی تتراکوریک: زمانی کاربرد دارد که یک یا دو متغیر دو ارزشی ساختگی باشند .یا هر دو متغیر دو ارزشی و فرض نرمال بودن رعایت شده باشد مثل همبستگی بین پاسخ گروه قوی و ضعیف به سوال ۱ .

....

۴ – ضریب همبستگی کانونی : هنگامی که بخواهیم رابطه ببین دو یا چند متغیر را با دو یا چند متغیر دیگر بررسی کنیم .ازاین ضریب همبستگی استفاده می کنیم.

....

۵ – همبستگی تفکیکی (پاره ای) : ضریب همبستگی دو متغیری است .برای داده های نا پیوسته که اثر متغیر سوم و بعدی را می خواهیم روی هر دو متغیر حذف کنیم .بکار می رود .

$r_{XY}.Z_{fg}$	$r_{XY}.Z \quad F$	$r_{XY}.Z$
مرتبه سوم	مرتبه دوم	مرتبه اول

رابطه پیشرفت تحصیلی و خلاقیت با حذف اثر هوش بر روی هر دو متغیر (تفکیکی مرتبه اول)

۶ – ضریب نیمه تفکیکی :اثر متغیر سوم و بعدی را روی یکی از متغیر ها حذف می کنیم :

$r_{X(Y.Z)f}$	$r_{X(Y.Z)}$
مرتبه دوم	مرتبه اول

✓ نکته : تفسیر ضریب همبستگی نباید بر حسب درصد و نسبت باشد .مثلا $r_X = ۰/۷۰$ هفتاد درصد رابطه بین متغیرها را تبیین نمی کند و $r_X = ۰/۹۰$ دقیقا دو برابر $r_X = ۰/۴۵$ نیست.

روش	نماد	متغیر ۱	متغیر ۲	توضیح

گشتاوری پیرسون	R	پیوسته	پیوسته	با ثبات ترین روش است
تفاوت رتبه ها (rho)	ρ	رتبه ای	رتبه ای	بیشتر به جای همبستگی گشتاوری بکاربرده میشود وقتی که تعداد موارد کمتر از ۳۰ است.
تای کندال	τ	رتبه ای	رتبه ای	زمانی که تعداد موارد ۱۰ کمتر است بجای (rho) بکار برده می شود.
دورشته ای	r_{bis}	دو ارزشی ساختگی	پیوسته	مقدار این همبستگی گاهی اوقات از ۱ بیشتر می شود و خطای استاندارد آن بزرگتر از r است .ازاین روش در تجزیه و تحلیل سوال استفاده می شود .
دورشته ای گسترده	r_{wbis}	دو ارزشی ساختگی گسترده	پیوسته	زمانی بکار برده می شود که علاقمند به مطالعه ی افرادی هستید که در کرانه های انتهایی متغیر دو ارزشی قرار دارند
دورشته ای نقطه ای	r_{pbis}	دو ارزشی واقعی	پیوسته	ضریب همبستگی حاصل در این روش کوچکتر از r و بسیار کوچک تر از r_{pbis} است..
چهار خانه ای (تتراکوریک)	r_t	دو ارزشی ساختگی	دو ارزشی ساختگی	زمانی بکار برده می شود که هردو متغیر را بتوان در نقطه ی بحرانی معینی به دو قسمت تقسیم کرد
ضریب فنی	φ	دو ارزشی واقعی	دو ارزشی واقعی	برای محاسبه همبستگی درونی بین سوال ها در آزمون های چند گزینه

				ای یا دو گزینه ای بکار برده میشود.
ضریب توافقی	c	دو یا چند طبقه ای	دو یا چند طبقه ای	در برخی شرایط معین با r_t قابل مقایسه است ارتباطی نزدیکی با خی دارد .
شاخص رابطه نامتقارن سامرز	d	هر دو متغیر دو مقوله ای یا رتبه ای		اینکه کدام متغیر مستقل یا وابسته نامگذاری می شود در نوع محاسبه ی شاخص تاثیر دارد.

۱ – متغیر پیوسته :

به متغیری گفته می شود که پیوستار زیربنایی آن تمایل به توزیع طبیعی به هنجار دارد . نمونه هایی از این متغیر عبارت اند از :وزن توانایی یا پیشرفت تحصیلی که بوسیله آزمون های استاندارد اندازه گیری می شوند.

۲ – متغیر دو ارزشی ساختگی :

وقتی حاصل می شود که یک متغیر پیوسته بر اساس یک معیار قرار دادی که غالبا نزدیک مرکز داده هاست .به دو گروه تقسیم شود. به عنوان امثال تقسیم بندی دانش آموزان به موفق – ناموفق بالای متوسط پایین متوسط ، قبول – رد وصمیمی – بی تفاوت در یک مقیاس نگرش سنج.

۳ – متغیر دو ارزشی واقعی :

نقطه برش نسبتا روشنی دارد (البته نه ضرورتا مطلق) بطوریکه سرانجام داده های این متغیر به دو گروه تقسیم می شوند . نمونه هایی از این متغیر عبارتند از : مذکر – مونث ، مرده – زنده ، معلم – غیر معلم ، مردود – نامردود ، و سیگاری – غیر

سیگاری . متغیر های دیگری وجود دارند که برای محاسبه همبستگی می توان آنها را دو ارزشی واقعی تلقی کرد.نظیر کوررمگی- غیرکوررنگی ، الکلی – غیر الکلی و پاسخ های صحیح – غلط یک سوال معین در یک آزمون توزیع های زیربنایی دو ارزش های واقعی اگر دارای تفاوت های مطلق نباشند ، دونمایی و دونمایی یا بطور نسبی ناپیوسته هستند.

کوواریانس (واریانس مشترک) : کوواریانس تغییر پذیری مشترک بین دو متغیر گویند. (واریانس یک متغیر) مقدار و جهت رابطه ی بین دو متغیر را اندازه می گیرد . ولی دو محدودیت دارد. مقدار کوواریانس به مقدار تغییر پذیری نمره های x وy بستگی دارد . اگر پراکندگی زیاد کوواریانس نیز زیاد باشد در نتیجه نمی توان کوواریانس اندازه های رمختلف را باهم مقایسه کرد.تفاوت در مقدار کوواریانس ممکن است ، در اثر رابطه ی بین متغیر ها ، تفاوت در انحراف معیار ها و یا تفاوت در رابطه و انحراف معیار ها باشد.شاخصی که از تفاوت در اندازه مقیاس ها تاثیر نپذیرد همبستگی است. این عمل تصحیح از رابطه ی الف بدست می آید (شیولسون ، ۱۳۸۰ : ۱۷۵)

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

✦ پرسش : اگر کوواریانس x وy صفر باشد ،کدام عبارت درباره x وy صحیح است ؟

الف) رابطه ای وجود ندارد

ب) رابطه ای غیر خطی وجود دارد

ج) دو متغیر مستقل هستند

د) یا رابطه غیر خطی با استقلال وجود دارد

- پاسخ : گزینه (د) است . اگر کوواریانس صفر شد ،نمی توان حکم به استقلال دو متغیر داد چون ممکن است بین آنها رابطه ای غیر خطی وجود داشته باشد . که بوسیله ی

کوواریانس مشخص نمی شود . چون کوواریانس فقط نوع رابطه خطی دو متغیر را تعیین می کند . (آذر ، ۱۳۸۰ : ۱۳۲) .

- مثال ۱: اگر cov برابر ۵۵ و $S^2{}_y = ۶۴$ باشد ، همبستگی بین xو y چقدر است ؟

- مثال ۲: اگر $\sum(X - \bar{X}) = 5$ و $\sum\ (Y - \bar{Y}) = 8$ باشد و انحراف معیار X برابر ۱۰ و انحراف معیار y برابر ۱۶ باشد. ضریب همبستگی چقدر است ؟

- مثال۳ : اگر $Z_Y = 3/5$ و $Z_X = 2$ باشد و حجم نمونه برابر ۴۰ نفر ، ضریب همبستگی چقدر است ؟

- مثال ۴ : چنانچه میانگین وزن دانش آموزان (n=10) برابر ۵۰ و میانگین قد همین گروه 110 cm باشد و فردی قد 115 cmو وزن 58 kg داشته باشد میزان همبستگی بین قد و وزن او چقدر است ؟ $S_1 = 5$ وزن و $S_2 = ۳$قد

ضریب تعیین (درصد واریانس مشترک)

ضریب تعیین نشان دهنده میزان تاثیری است که متغیر x (مستقل) در متغییر y (وابسته) ایجاد می کند یا به عبارتی با محاسبه ی این ضریب می توان تعیین کرد چند درصد از کل واریانس ناشی از واریانس y است.

$$r^2{}_{xy} = \frac{cov^2}{S^2{}_x . S^2{}_y} = \frac{\text{واریانس مشترک}}{\text{کل واریانس}}$$

- **نکته** : ضریب تعیین هیچ وقت منفی نخواهد شد زیرا برای محاسبه ی آن ضریب همبستگی مجذور می شود سوال : اگر همبستگی بین نمره ریاضی و نمره هندسه ۰/۷ باشد ، درصد واریانس را بدست آورده و با شکل نشان دهید.

عواملی که برضریب همبستگی تاثیر می گذارند .:

۱ – اساس رابطه ای از جامعه ای به جامعه ی دیگر فرق می کند .مثلا در افراد بشر در سنین ۱۶-۱۰ سالگی بین سن تقویمی و توانایی فیزیولوژیکی همبستگی بالایی وجود دارد ولی بین این دو متغیر در سن ۲۶-۲۰ سالگی همبستگی وجود ندارد.

۲ - پراکندگی متغیرها در جوامع مختلف متفاوت است بدین معنی که هر چه تجانس بیشتر باشد (واریانس کمتر باشد) همبستگی کمتر است بعنوان مثال : در یک پیژوهش اگر همه باهوش باشند دامنه محدود ومقدار r کاهش می یابد.

۳ – همبستگی بین دو متغیر تحت تاثیر همبستگی آنها با متغیر سوم قرار دارد. بعنوان مثال :همبستگی بین فیزیک و ریاضی ممکن است به دلیل همبستگی این متغیرها باهوش باشد.

۴ – استفاده از گروه های انتهایی ضریب همبستگی را زیاد می کند ،نمره های انتهایی در حجم های کوچک بر همبستگی تاثیر زیاد می گذارد.

۵ – اگر n افزایش یابد احتمال معناداری r بیشتر است .

۶ – وقتی متغیرها نامرتبط باشند r کاهش می یابد.

۷ – رابطه غیر خطی امکان دارد ضریب همبستگی پیرسون را به صفر نزدیک کند.

۸ – محدودیت در دامنه تغییرات ضریب همبستگی را کاهش می دهد . برای وجود محدودیت در دامنه تغییر یابد واریانس ها یا انحراف معیار های متغیرهایی که ضریب همبستگی آنها محاسبه می گردد. مورد بررسی قرار گیرند.

کوچک بودن واریانس ها می تواند نشانه محدودیت در دامنه تغییرات باشد. (شیولسون ۱۳۸۰:۱۹۷:)

۹ – اگر واریانس کوچک باشد تفسیر ضریب همبستگی باید باید با احتیاط لازم صورت گیرد .

- ✓ **نکته** ۱ : آزمون معناداری برای همبستگی در دو متغیر مستقل آزمون **t** فریدلی است و اگر گروه ها مستقل است آزمون **z** فیشر و اگر ضریب همبستگی را در دو گروه وابسته مقایسه کنیم **t** استیودنت است. (فرگوسن تاکانه :۱۳۸۰)
- ✓ **نکته** ۲ : درجه آزادی همبستگی برابر است با : **d.f=n-2**

تحلیل همبستگی از نوع پیش بینی : (رگرسیون و پیش بینی)

زمانی که بین دو متغیر همبستگی وجود داشته باشد می توان از طریق رگرسیون مقدار یک متغیر y را از اروی یک متغیر دیگر x پیش بینی یا برآورد کرد. و هرجه همبستگی بین متغیرها بالاتر باشد ، به همان اندازه پیش بینی دقیق تر است . (معدل نمره دروس دانشگاهی را از روی نمره آزمون استعداد پیش بینی می کنیم) این پیش بینی ازر روی **خط رگرسیون** بدست می آید. این خط بهترین برازش از میان مجموعه نقاط نمودار پراکندگی است .

راه دیگر که به داوری ذهنی بستگی ندارد **اصل کمترین مجذورات** است که مجذور انحراف ما حول خط رگرسیون را کمینه می سازد .می توان معادله ی این خط را بدست آورد و سپس نمودار آن را رسم کرد.

خط رگرسیون در حقیقت یک میانگین متحرک یا خط کمترین مجذورات است. . رگرسیون در حقیقت معادله ی خط است . Y=a+xb

- ✓ **نکته** : رابطه بین متغیر پیش بینی شونده (املاک) **(Y)** و متغیر پیش بینی کننده **(X)** تابع علامت و شدت ضریب همبستگی است .اگر $r_{xy} = 1$ باشد رگرسیون صفر و اگر $r_{xy} = 0$ باشد ، همبستگی کامل است .

پیش بینی Y **از روی** x

- ○ **شیوه شهودی** :که از طریق ترسیم نمودار ممکن است تغییر و تفسیر آن تابع فرد مشاهده گر باشد.

شکل الف : چون خط افقی است ارتفاع شیب آن صفر است در این حالت پیش بینی دقیق امکان پذیر است . این شکل دو عامل مهم در پیشبینی را دارد :

۱ – میانگین متغیر Y یعنی $\overline{Y}$ و شیب خط فرضی در نمودار پراکندگی . ۲- نمدار نشان می دهد .اگر شیب خط (ضریب زاویه) صفرباشد.برای هر مقدار از X نمره Y برابر است با میانگین نمره های Y (شیولسون ، ۱۳۸۰ : ۲۱۹)

شکل ب : بین X و Y رابطه مثبت و کامل وجود دارد اگر X افزایش یابد مقدار Y به همان میزان افزایش می یابد.

شکل ج : بین X و Y رابطه منفی و کامل وجود دارد اگر X افزایش یابد مقدار Y به همان میزان کاهش می یابد.

$$\text{شیب و ضریب زاویه} = \frac{\text{تغییر در } Y}{\text{تغییر در } X} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

برای پیش بینی بطور کامل باید اطلاعات زیر را داشته باشیم :

الف : میانگین نمره های متغیر Y

ب : شیب خط فرضی که نمودار پراکندگی را توصیف می کند.

ج : موقعیت نسبی فرد در متغیر X (نمره مفروض) این مقدار باید به میانگین نمره های Y اضافه یا از آن کسر گردد .

$$Y = \bar{Y} + \frac{Y_2 - Y_1}{X_2 - X_1}\ (X - \bar{X})$$

پیش بینی با استفاده از رگرسیون خطی
Y=a+xb

a : عر ض از مبدا (مقدار Y وقتی X صفر است .)

B : شیب خط (مقدار متغیر در Y به ازای یک واحد تغییر در X)
$b = \frac{Y_2 - Y_1}{X_2 - X_1}$

ازاین معادله هنگامی استفاده می شود که همه نقاط یک توزیع مشترک ری یک خط مستقیم قرار می گیرند .

پیش بینی نمره های استاندارد (Z):

مقدماتی ترین روشی که در استفاده از ضریب همبستگی پیرسون برای پیش بینی بکار برده می شود نمره های استاندارد است . $Z_Y = (Z_X)(r_{XY})$

- ✓ **مهم** :اگر همبستگی کامل (۱) باشد نمره استاندارد پیش بینی نیز کامل در غیر اینصورت غیر کامل می باشد .بعنوان مثال اگر نمره استاندارد Z برابر ۲ و همبستگی کامل (۱) باشد نمره پیش بینی استاندارد نیز ۲ می باشد .
- ✓ **مهم :** هنگا می که همبستگی بین دو متغیر کم باشد نمره استانداردی که پیش بینی می کنیم نزدیک به میانگین خواهد بود.

- ✓ **نکته ۱** : زمانی که همبستگی بین دو متغیر پایین باشد ،نمرات پیش بینی شده نزدیک به میانگین نمره پیش بینی شوندهستند تا نمره واقعی به این پدیده رگرسیون می گویند .

- ✓ **نکته ی ۲**: میزان همبستگی بین دو متغیر حدود یا مقدار اتفاق رگرسیون را تعیین میکند
- ✓ **نکته ی ۳**: پدیده ی رگرسیون اولین بار به وسیله ی گالتن مورد استفاده قرار گرفت بر اساس مطالعات گالتن فرزندان والدین بلند قدر بلند قد هستند اما نه به اندازه ی والدین خود به همین ترتیب فرزندان والدین کوتاه قد کوتاه قد هستند اما نه به کوتاهیه والدین قد والدین خود .
- ✓ **نکته ی ۴**: خط رگرسیون خطی ست که خطاهای پیشبینی رار به حداقل می رساند (خط حداقل مجذور ها)یعنی این که مجموع مجذور فاصله Yها از خط رگرسیون کوچکتر از فاصله هر خط دیگری را محور Yها میباشد خط رگرسیون را خط برازنده نیز می نامند .
- ✓ **نکته ی ۵**: اختلاف بین نمره ی واقعیه Yو نمره ی پیش بینی شده ی Y' را خطای پیش بینی e می گویند

$E=Y-Y'$

نکات بسیار مهم

۱-رگرسیون زمانی اتفاق می افتاد که همبستگی کامل نباشد.یعنی نمرات از گروه های بالا و پایین جامعه انتخاب شده باشند (کرانه بالا و پایین).

۲-اگر هبستگی صفر باشد رگرسیون کامل و اگر یک باشد رگرسیون صفر است .

۳-بین شدت هبستگی و رگرسیون هبستگی معکوس وجود دارد .

۴-انحراف نمره از خط رگرسیون از انحراف نمره از هر خط دیگری کوچکتر است.

۵- چنانچه پراکندگی نقاط در اطراف خط رگرسیون به شکل بیضی باشد هبستگی بین دو متغیر در حد متوسط و مقدار ان نزدیک به ۰,۵ است (آیزاک ۱۷۲:۱۳۸۱)

معادله خط رگرسیون

برای پیش بینی یک متغیر ازروی متغیر دیگر از معادل خط رگرسیون استفاده می شود $Y'=a+bx$ که در ان :

B=شیب خط
Y=متغیری که می خواهیم پیش بینی کنیم
a=محل تلاقی خط رگرسیون با محور عرض ها یا عرض از مبدا
X=متغیری که مقدار ان را در اختیار داریم

✓ نکته :بعضی مواقع مقدار a وb در اختیار است. ر این صورت با جایگزینی مقادیر محاسبه خیلی ساده است.

شیب خط رگرسیون= bxy و Byx	انحراف استاندارد y= S_Y
ضریب همبستگی= rxy	انحراف استاندارد x= S_X

محاسبه a و b از روی داده های خام

$$a_{yx} = \frac{\Sigma Y - b_{YX}\Sigma X}{N}$$

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^{2} - (\Sigma X)^{2}}$$

مثال: برای داده های زیر معادله خط رگرسیون را به دست آورید؟

X	Y	X۲	Y۲	XY
٦	٨	٣٦	٦٤	٤٨
٩	٩	٨١	٨١	٨١
٥	٧	٢٥	٤٩	٣٥
٢٠	٢٤	١٤٢	١٩٤	١٦٤

❁ خطای استاندارد

اکثر مواقع بین نمرات پیش بینی شده و نمرات مشاهده تفاوت وجود دارد بیه این حالت اختلاف خطای استاندارد براورد می گویند.

- ✓ نکته ۱ ؛ هر چند همبستگی بین متغیرها بیشتر باشد خطای پیش بینی کمتر خواهد بود ، این خطا به دو روش محاسبه می شود .

S_{yx} خطای استاندارد پیش بینی $S_{yx} = S_y\sqrt{1 - r^2xy}$

$$= \sqrt{\frac{\sum e^2}{N}} \quad \text{یا} \quad S_{yx} = \sqrt{\frac{\sum(Y - \bar{Y})^2}{N - 2}}$$

خطای پیش بینی $= e$ ضریب همبستگی $= r\ xy$

نکته: N −۲ را درجه آزادی خطای استاندارد پیش بینی گویند.

رابطه پایایی و خطا

$S_E = S\sqrt{1 - r_{xx}}$ همانطور که ملاحظه می شود $r_{\ xx}$ ضریب اعتبار(پایایی) است.

نکته۱: اگر همبستگی کامل باشد خطا صفر است. یعنی همبستگی با خطا رابطه عکس دارند.

نکته۲: انحراف معیار نمونه و خطای پیش بینی رابطه مستقیم دارند (به فرمول ها توجه کنید).

نکته۳: بین خطای اندازه‌گیری و ضریب پایایی رابطه عکس وجود دارد.

انواع رگرسیون

تک متغیری: از روی یک متغیر، متغیر دیگر را پیش‌بینی می‌کنیم. از روی هوش، خلاقیت فرد را پیش‌بینی می‌کنیم.

چندگانه: از روی چند متغیر، یک متغیر پیش‌بینی می شود. از روی اعتماد به نفس و اضطراب و خلاقیت، پیشرفت تحصیلی پیش‌بینی می شود.