Â¶ÍkhÎwÍ#ÁÌ¿º¼¶pÁ#½sÄ»#R¯ÌµÎeÍ#»#nÌ¶Á#½»q]

آمار توصیفی و احتمالات

جزوه درسی

WWW.SHOP.IRANESTEKHDAM.IR

دانلود نمونه سوالات استخدامی

دانلود نمونه سوالات استخدامی

ایران استخدام

آمار توصیفی

شامل مجموعه‌ای از روش‌ها برای برنامه‌ریزی آزمایش‌ها، بدست‌آوردن اطلاعات و سپس سازماندهی، خلاصه‌سازی، ارائه و تجزیه و تحلیل اطلاعات و درنهایت نتیجه‌گیری برمبنای 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 داده‌ها می‌پردازد. هدف آمار توصیفی، توصیف واقعیات موجود است که تنها خطای موجود در آن خطای اندازه‌گیری است، مشخص‌کردن قد دانش آموزان یک کلاس، بررسی روند رشد سرانه ملی در سه سال گذشته نمونه‌ای از مفاهیم مورد بررسی آمار توصیفی هستند. نمادهای مهم در آمار عبارتند از:

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

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 ۰/۵ نمره در اعداد صحیح و در اعداد اعشاری، نیم واحد یعنی ۰/۰۵– ۰/۰۰۵ و مانند آن کسر می‌شود. یعنی: کرانه عدد ۲۵ ← ۲۴/۵ تا ۲۵/۵ و کرانه عدد ۲۵/۵ ← ۲۵/۴۵ تا ۲۵/۵۵ می‌باشد. مفهوم حدود واقعی مخصوصاً زمانی مفید است که اعداد گروه بندی یا طبقه بندی شوند. مثال: پس از اجرای یک آزمون ریاضی مشاهده می شود 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 می‌باشد؛ یعنی ۱۲/۵– ۱۱/۵.

توزیع فراوانی

عبارتست از سازمان دادن اندازه ها یا مشاهدات به صورت طبقات همراه با فراوانی هر طبقه. توزیع فراوانی، داده‌ها را بصورت خلاصه و مرتب، به نحوی که تفسیر آنها آسان شود، نمایش می دهند.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

مراحل ساخت جدول توزیع فراوانی

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

۲- مشخص کردن تعداد دفعاتی که هر عدد تکرار شده است (تعداد فراوانی). زمانی که همه اعداد تک‌تک در جدول آورده شوند، جدول توزیع فراوانی منفرد یا طبقه بندی نشده گفته می شود. اما زمانی که نمره‌ها یا اعداد دارای دامنه گسترده ای هستند و تنظیم اعداد بصورت توزیع فراوانی طبقه بندی نشده وقتگیر و طاقت‌فرسا 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 طبقه‌بندی‌کردن 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

X	F
۱۵	۱
۱۲	۳
۱۱	۵
۱۰	٤

نکته: با توجه به جدول فوق، عدد ۵ در ستون f بیانگر این است که عدد ۱۱ پنج بار تکرار شده است. اگر داده‌های ستون فراوانی (F) را با هم جمع کنیم تعداد کل داده ها بدست می آید. $N = \Sigma F$

یعنی در مثال فوق $N = ۱۳$

توزیع فراوانی طبقه بندی شده

زمانی که تعداد اعداد یک توزیع و همچنین فاصله بین آنها خیلی زیاد باشد، از توزیع فراوانی طبقه‌بندی‌شده استفاده می شود.

نکته: زمانی که تفاضل بین بزرگترین و کوچکترین نمره یا عدد مساوی یا بزگتر از ۲۰ باشد از توزیع فراوانی طبقه‌بندی‌شده استفاده می شود.

طبقات بایستی ناسازگار باشند. یعنی یک عدد معین فقط در یک طبقه قرار داده شود. نمره های فردی در این نوع توزیع‌ها هویت خود را از دست می دهند.

نحوه ساختن توزیع فراوانی طبقه بندی شده

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

الف) روش یکم دارای مراحل زیر است:

۱- تعیین دامنه تغییرات $R = X_H - X_L + ۱$

۲- تقسیم R به ترتیب بر ۱-۲-۳-٤ تا حاصل از ۲۰ کمتر شود. عدد بدست آمده تعداد طبقات و مقسوم علیه فاصله طبقات می‌باشد

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

۳- نوشتن طبقات: اولین عدد باید از کوچکترین عدد توزیع کوچکتر باشد و مضربی از فاصله طبقات نیز باشد.

٤- نوشتن فراوانی طبقات

ب) روش دوم دارای مراحل زیر است:

۱- تعیین دامنه تغییرات

۲- تعیین تعداد طبقات با استفاده از قانون استرژ $K = ۱ + ۳/۳ Log_N$

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

٤- نوشتن طبقات

۵- نوشتن فراوانی طبقات

نکته: تعداد طبقات اختیاری است و معمولاً بین ۲۰-۱۰ است. و اگر تعداد طبقات بزرگتر از۲۰ باشد، تهیه و تنظیم جدول نیاز به وقت و کار بیشتر دارد. اگر تعداد طبقات کوچکتر از۱۰ باشد اندازه طبقات بزرگ می‌شود و اطلاعات بیشتری از دست می رود

نماینده طبقات (نقاط وسط طبقات): نماینده طبقات یا نقاط میانی را با x' نمایش می دهند و ازطریق فرمول زیر به‌دست می‌آید:

$$X' = \frac{\text{حد بالای طبقه+حد پایین طبقه}}{۲}$$

توزیع فراوانی تراکمی: اگر پژوهشگری علاقمند به دانستن تعداد افراد یا نمره هایی باشد که در پایین نمره یا عدد خاصی وجود دارند، نیاز به توزیع فراوانی تراکمی دارد. فراوانی تراکمی با (cf) نشان داده می شود که از جمع‌کردن فراوانی‌های ساده هر طبقه با طبقه بزرگتر به دست می آید.

نکته۱: فراوانی تراکمی کوچکترین طبقه همیشه برابر با فراوانی ساده یا مطلق آن طبقه است.

نکته۲: فراوانی تراکمی بزرگترین طبقه همیشه برابر با مجموع داده ها (ΣF) یا N می باشد.

محاسبه‌ی فراوانی نسبی

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

درصد فراوانی مطلق و تراکمی

$$P = \frac{F}{N} \times ۱۰۰ \qquad P = \frac{\text{فراوانی مطلق هر طبقه}}{\text{تعداد کل فراوانی ها}} \times ۱۰۰$$

$$CF\% = \frac{CF}{N} \times ۱۰۰ \qquad Cf\% = \frac{\text{فراوانی تراکمی هر طبقه}}{\text{تعداد کل فراوانی ها}} \times ۱۰۰$$

نمودارها

نمودار دایره‌ای: 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:

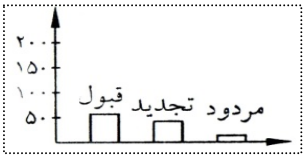
$$\text{درجه}^\circ = \frac{n}{N} \times ۳٦۰$$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

مثال: اگر داخل یک گروه ۱۲۰ نفری۶۰ نفر قبول، ٤۰ نفر تجدید و۲۰ نفر مردود شده باشند، درجه مربوط به تجدیدی‌ها چند است؟

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

نمودار ستونی (میله‌ای): 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 متغیرهای طبقه ای مانند قبولی- مردودی، رنگ چشم و اعتقادات مذهبی نمودار رسم شود، هیچ ضرورتی ندارد که فاصله ها در محور X ثابت می‌ماند. برای این متغیرها از نمودار دایره ای و مانند آن استفاده می کنند (نمودار پای). مثال:

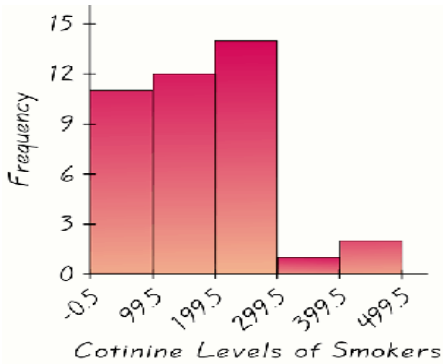


نمودار هیستوگرام: با داده های فاصله ای و نسبتی به کار می رود. نموداری شبیه نمودار ستونی است، ولی در آن ستونها به هم چسبیده است و در محور افقی (X) کرانه (حدود واقعی) طبقات و در محور عمودی (y) فراوانی مطلق قرار می گیرد.

نکته: این نمودار، اغلب در مواردی استفاده می شود که بخواهیم فقط یک نمودار واحد را نشان دهیم، درحالیکه در موقعیت‌های آزمایشی و شبه آزمایشی که مایلیم نمره های آزمودنی های دو گروه مجزا را با هم مقایسه کنیم ناگزیر هستیم از پلیگون استفاده کنیم.

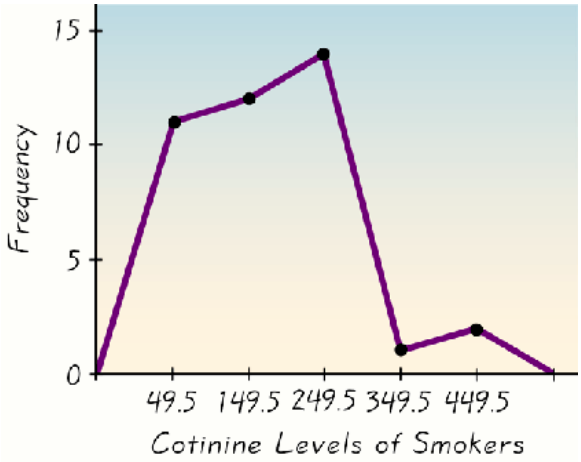
مثال:

X	F_i
٤۰۰-٤۹۹	۲
۳۰۰-۳۹۹	۱
۲۰۰-۲۹۹	۱٤
۱۰۰-۱۹۹	۱۲
۰-۹۹	۱۱



نمودار چندضلعی (پلیگون): با داده های فاصله ای و نسبتی به کار می رود. نموداری شبیه نمودار هسیتوگرام است که در محور X چند نمره میانی و در محور Y فراوانی مطلق قرار می گیرد.

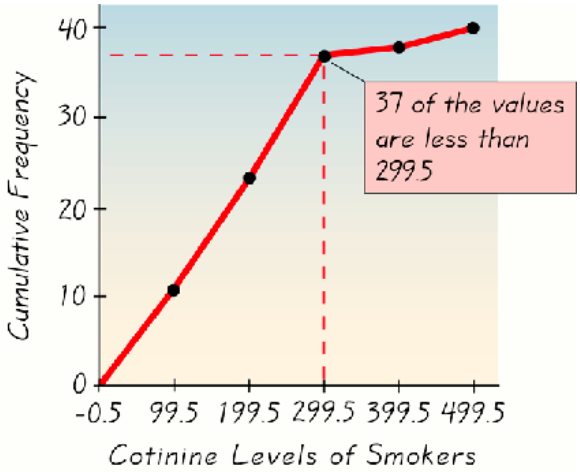
دانلود
نمونه سوالات
استخدامی
ایران استخدامی



نکته: برای مقایسه چند گروه در یک محور مختصات اگر تعداد آزمودنی‌ها در گروه ها برابر باشد، می‌توان از نمودار ستونی و 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 به‌کارمی‌رود. در محور X کرانه طبقات و در محور Y فراوانی تجمعی قرار می گیرد. اگر به جای فراوانی تراکمی درصد فراوانی تراکمی را به صورت هندسی نمایش دهیم حاصل اجایو خوانده می شود.

این نمودار وقتی مفید است که پژوهشگر علاقمند باشد وضعیت یک نمره یا یک فرد را نسبت به بقیه نمره ها یا افراد مشخص کند. برای ترسیم این نمودار در روی محور عرضی (Y) فراوانی تراکمی و در روی محور طولی (X) حدود واقعی طبقات قرار می گیرد. مثل نمودار چند ضلعی است.



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. مثلاً: ۹-۵-۶ -۶-۶-۶-۶-۶-۶-۷-۶-۵-۷-۶-۵-۱

در اینجا نما عدد ٦ است. برخی موارد توزیع، دونمایی یا چندنمایی می شود مثل: ۷-٦-۵-۵-٤-۲-۳-۲-۱

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

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

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

در اینجا عدد ٤ نما است، زیرا عدد چهار، ۹ بار تکرار شده است.

X	F_i
۵	۵
۴	۹
۳	۳
۲	۲
۱	۶

نکته۱: برای محاسبه نما باید ابتدا اعداد را از کوچک به بزرگ مرتب کنیم.

نکته۲: در موقعیت هایی که دو عدد مجاور x فراوانی یکسان داشته باشند که بزرگتر از فراوانی سایر ارزش هـای x باشند، نما را می توان به طور قراردادی به عنوان میانگین دو ارزش مجاور x در نظر گرفت.

۱-۲-۳-۳-٤-٤-۵ میانه $= \frac{۳+٤}{۲} = ۳/۵$

نکته۳: در جایی که ارزش های غیر مجاور فراوانی های بزرگتر از فراوانی های طبقه های مجاور داشته باشند، هرکدام از این ارزش ها را می توان به عنوان نما در نظر گرفت. در چنین شرایطی توزیع دونمایی نامیده می شود.

۱۱-۱۱-۱۲-۱۲-۱۲-۱۳ -۱۳-۱۳-۱۳-۱٤-۱٤-۱٤-۱۵-۱۵-۱۵-۱۵-۱٦-۱٦-۱۷-۱۷-۱۸

در اینجا عدد ۱۳ پنج بار تکرار شده به طوری که این فراوانی بزرگتر از فراوانی ارزش های مجاور است. همچنین عدد ۱۵ چهار بار تکرار شده و این فراوانی نیز بزرگتر از فراوانی ارزش های مجاور است. این مجموعه از مشاهده‌ها دارای دونما است (فرگوسن، تاکانه، ۱۳۸۰).

محاسبه نما در داده ای طبقه بندی شده

از معادله‌ی زیر محاسبه می شود، که در آن:

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

i = طول یا فاصله طبقات

$d_۱$ = تفاوت فراوانی طبقه نمادار از فراوانی طبقه قبل

$d_۲$ = تفاوت فراوانی طبقه نمادار از طبقه بعد.

مثال:

X	F
٥٤-٥٦	۱
٥٧-٥٩	۳
٦۰-٦۲	٦
٦۳-٦٥	۸
٦٦-٦۸	۲

طبقه نما ←

$$MO = L + i(\frac{d_1}{d_۱ + d_۲})$$

$$d_۱ = ۸ - ٦ = ۲$$

$$d_۲ = ۸ - ۲ = ٦$$

$$MO = ٦۲/۵ + ۳(\frac{۲}{۲+٦}) = ٦۳/۲۵$$

نکته: چنانچه توزیع نرمال باشد، نما از فرمول زیر محاسبه می شود: $MO = ۳Md - ۲\overline{X}$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

میانه

میانه جایگاهی در توزیع نمره هاست و توزیع نمره ها را به دو قسمت مساوی تقسیم می کند؛ یعنی جایی است که دقیقاً ۵۰ درصد نمره ها بالای آن و ۵۰ درصد نمره ها زیر آن قرار می گیرند. میانه از نما باثبات‌تر و از میانگین بی‌ثبات‌تر 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 ها را از کوچک به بزرگ مرتب می کنیم و سپس اگر تعداد اعداد فرد باشد، میانه، عدد وسط است.

مثال: ۱۲-۹-٤-۳-۱-٦-۵

مثال: ۱۲-۹-٦-(۵)-٤-۳-۱ در اینجا عدد ۵ میانه است

مثال: در توزیع اعداد ۳۱-٤-۵-۲۱-۳۰-۳۰-۱۷-۸-۱۲ میانه برابر است با ۱۷.

طریقه محاسبه میانه اعداد خام در تعداد زوج

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

۲- دو عدد وسط را با هم جمع و تقسیم بر دو می کنیم. ۱۲-۹-۲-۸-۵-۳

۵+۸ برابر ۱۳ و ۱۳ تقسیم بر ۲ میانه محاسبه می شود که برابر ٦/۵ می باشد. ۱۲-۹-۸-۵-۳-۲

نکته: هنگامی که نمره یا عددی که توزیع را به دو قسمت تقسیم می کند تکراری است، میانه ازطریق محاسبه بدست می‌آید.

الف) ابتدا حد پایین عدد تکراری که میانه یکی از آنها است را می نویسیم.

ب) کسری را در نظر می گیریم که مخرج آن تعداد اعداد تکرای و صورت آن نشان‌دهنده تعداد اعداد تکرای است که در سمت چپ خط رسم کننده میانه قرار می گیرند.

ج) حاصل مراحل (الف) و (ب) را با هم جمع و میانه را بدست می آوریم. (میانه روی عدد ٤ می‌افتد، بنابراین نیمی از آن بعلاوه‌ی یک مورد ٤ قبل از آن جمعاً ۱/۵ تا از ٤ ها به حد پایین اضافه می شود) مانند:

۳-٤-٤-۵-٦ میانه = ۳/۵+$\frac{۱/۵}{۲}$=٤/۲۵

۱-۲-۳-۳⊥۳-٤-۵-٦ میانه = ۳/۵+$\frac{۲}{۳}$=۳/۱۷

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: ۳-۵-۲-٤-۸-٤-٤-٤-٤-۲

$$md = L + \frac{\frac{N}{۲} - F_C}{F_i} i$$

L= حد پایین طبقه میانه دار

طبقه‌ی میانه دار= مجموع فراوانی تقسیم بر دو= $\frac{N}{۲}$

X	F_i	F_c
۸	۱	۱۰
۵	۱	۹
٤	۵	۸
۳	۱	۳
۲	۲	۲
	∑F=۱۰	

۸

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

F_c = فراوانی تجمعی ماقبل سطری که در آن میانه واقع شده است.

F_i = فراوانی مطلق طبقه میانه دار

i = فاصله طبقاتی

F_c=۳ , F_i=۵ , i=۱

مثال:

$$md=L+\frac{\frac{N}{۲}-F_c}{F_i}\times i \Rightarrow md=۳/۵+\frac{۵-۳}{۵}\times ۱ \qquad md=۳/۹$$

نکته: در صورتی که داده در مقیاس فاصله ای یا نسبی باشند، 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 مناسب‌تری است. به عنوان مثال در توزیع ۳۰۰-۱۵-۹-۷-۶-۵ میانه مناسب‌تر است.

ویژگی‌ها

۱- نسبت به اعداد بزرگ یا کوچک حساس نیست. بنابراین بهترین شاخص است که تمرکز اعداد را در وسط توزیع نشان می دهد. به عنوان مثال در توزیع های زیر که دارای میانگین های متفاوتی هستند، میانه برابر و مساوی ۲۰ می‌باشد. 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 داده‌هایی با مقیاس فاصله‌ای و نسبی هم استفاده شود.

۳- مجموع قدرمطلق انحرافهای نمره ها از میانه کوچکتر یا مساوی مجموع قدرمطلق انحرافهای نمره‌ها از هر عدد دیگری است (بدون در نظر گرفتن علامت).

$\sum\lvert X-Md\rvert \le \sum\lvert X-C\rvert$					
نمره ها	قدر مطلق انحرافات				
	میانه (٦)	٤	٥	٧	٩
٤	٢	٠	١	٣	٥
٥	١	١	٠	٢	٤
٦	٠	٢	١	١	٣
٧	١	٣	٢	٠	٢
٩	٣	٥	٤	٢	٠
	٧	١١	٨	٨	١٤

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

میانگین

میانگین، باثبات‌ترین شاخص گرایش مرکزی است که با داده‌های فاصله‌ای و نسبی بکار می‌رود و مرکز ثقل داده‌هاست.

طریقه محاسبه میانگین اعداد خام

مثال: ۱۱-۹-۱۰

$$\overline{X}=\frac{\sum X}{N} \Longrightarrow \text{میانگین} = \frac{\text{مجموع نمره ها}}{\text{تعداد نمره ها}}$$

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

طریقه محاسبه اعداد طبقه بندی شده با i=۱

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

مثال:

X	Fi
۵	۳
٤	۸
۳	۵
۲	۱۲
	$\sum F_i = n = ۲۸$

F.X
۱۵
۳۲
۱۵
۲٤
$\sum F.X = ۸٦$

$$\overline{X}=\frac{۸٦}{۲۸}=۳/۰۷$$

طریقه محاسبه میانگین اعداد طبقه بندی شده با i≠ ۱

$$\overline{X}=\frac{\sum F.X_c}{n}$$

مثال:

X	Fi
۱۷-۱۹	۵
۱٤-۱٦	٦
۱۱-۱۳	۱۱
۸-۱۰	۳
۵-۷	۹
	$\sum F_i = ۳٤$

$\Longrightarrow$

Xc	F.X
۱۸	۹۰
۱۵	۹۰
۱۲	۱۳۲
۹	۲۷
٦	٥٤
	$\sum F.X_c = ۳۹۳$

$$\overline{X}=\frac{۳۹۳}{۳٤}=۱۱/٥٦ \qquad \overline{X}=۱۱/٥٦$$

حد میانی طبقه

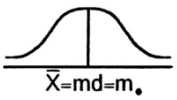
X_c= برای پیدا کردن حد میانی نمرات، دو نمره طبقه را با هم جمع و تقسیم بر ۲ می نمائیم. (۷+۵=۱۲÷۲ =٦)

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

نکته

۱- اگر میانگین و میانه و نما با هم برابر باشد، توزیع متقارن است و کجی آن صفر است.

$\overline{X} = md = mo$

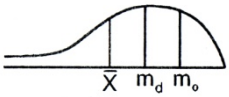


۲- اگر میانگین بزرگتر از میانه و میانه بزرگتر از نما باشد کجی مثبت است، یعنی اکثر نمرات پایین بوده و امتحان سخت است.

$\overline{X} > md > m_o$

۳- اگر میانگین کوچکتر از میانه و میانه کوچکتر از نما باشد، کجی منفی است و اکثر نمرات بالاست و امتحان آسان بوده است.

$\overline{X} < md < mo$



٤- کشیدگی[1] ارتفاع (Y) منحنی را نشان می دهد. به برآمدگی یا خوابیدگی منحنی در مقایسه با منحنی طبیعی کشیدگی گویند.

٥- 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 می شود.

جدید $\overline{X}_۱ = \overline{X} - C$ $\quad \overline{X}_۱ = \overline{X} + C$ $\quad \overline{X}_۱ = \overline{X} : C$ $\quad \overline{X}_۱ = \overline{X}.C$

٦- مجموع انحراف نمره ها از میانگین همیشه صفر است. $\sum (X-\overline{X}) = ۰$ درواقع شاخصی از مکان مرکزی به معنای حداقل مجذورات است (فرگوسن، تاکانه، ۱۳۸۰: ۸۱).

۷- مجموع مجذور انحرافات از میانگین همیشه می‌نیمم است (کمتر از هر عدد دیگری است).

$\sum(x-\mu)^۲ < \sum (x-a)^۲$

رابطه شاخص های گرایش به مرکز

$Mo = ۳md - ۲\overline{X}$ $\qquad md = \frac{mo + ۲\overline{X}}{۳}$ $\qquad \overline{X} = \frac{۳md - mo}{۲}$

میانگین میانگین ها

در صورتی که حجم نمونه ها برابر باشد یعنی $n_۱=n_۲,\ldots,n_n$ $\qquad \overline{X}_T = \frac{\sum \overline{X}_i}{N}$

مثال:

$\overline{X}_۱ = ۵$ $\quad \overline{X}_۲ = ۱۰$ $\quad \overline{X}_۳ = ۱۵$ $\qquad n_۱=۱۰$ $\quad n_۲=۱۰$ $\quad n_۳=۱۰$

[1]. *Kurtosis*

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

خواهیم داشت:

$$\overline{X}_T = \frac{\overline{X}_۱ + \overline{X}_۲ + \overline{X}_۳}{N} = \frac{۵+۱۰+۱۵}{۳} = ۱۰ \qquad \overline{X}_T = ۱۰$$

در صورتی که حجم نمونه ها برابر نباشد به طریق زیر محاسبه می شود:

$$\overline{X}_T = \frac{\sum \overline{X}_i . n_i}{n_i} \quad \Rightarrow \quad \frac{\overline{X}_۱ . n_۱ + \overline{X}_۲ . n_۲ + \overline{X}_۳ . n_۳}{n_۱ + n_۲ + n_۳}$$

مثال:

$\overline{X}_۱ = ۵$ $\quad \overline{X}_۲ = ۱۰$ $\quad \overline{X}_۳ = ۱۵$ $\qquad n_۱=۱۵ \quad n_۲=۱۰ \quad n_۳=۲۰$

$$\overline{X}_T = \frac{(۵\times۱۵)+(۱۰\times۱۰)+(۱۵\times۲۰)}{۱۵+۱۰+۲۰} = ۱۰/۵۵ \qquad \overline{X}_T = ۱۰/۵۵$$

میانگین هارمونیک (همساز)

این نوع میانگین برای مواردی بکار می رود که مقیاس ترکیبی باشد مانند متر در ثانیه و کیلومتر بر ساعت. این میانگین در عینک‌سازی و مطالعه شبکه های برقی به کار می رود.

$$HM = \frac{۱}{\frac{۱}{N}\left[\frac{۱}{x_۱} + \frac{۱}{x_۲} + \ldots + \frac{۱}{x_n}\right]}$$

مثال: میانگین سرعت های ۵ و٦ و۷ و۲ کیلومتر در ساعت ٤ ماشین چند است؟

$$HM = \frac{۱}{\frac{۱}{٤}\times\left[\frac{۱}{۵} + \frac{۱}{٦} + \frac{۱}{۷} + \frac{۱}{۲}\right]} = ۳/۹٦$$

میانگین هندسی

نوعی میانگین است که با G نشان داده می شود و معمولاً هرگاه x_i ها از درصدها یا نسبت ها تشکیل شده باشند، استفاده می‌شود. در کارهای اقتصادی یا جمعیت شناسی به کار می رود.

$$G = \sqrt[n]{X_۱ . X_۲ . X_۳ (X_n)}$$

مثال: میزان سود شرکت مهرداد در ۵ سال گذشته برحسب درصد به ترتیب ۲، ۳، ٤، ٤، ۳، شده است. کدام‌یک از گزینه های زیر وضع سود آوری شرکت را بهتر نشان می دهد.

$$G = \sqrt[۵]{۳\times٤\times٤\times۲\times۳} = ۳/۰۱$$

تفسیر: اعدادی مانند ۳ و٦ را می توان به‌عنوان فاصله‌های بین نقاط نشان داد. حاصل‌ضرب این دو عدد را نیز می‌توان به‌عنوان یک مساحت نشان داد. بنابراین مستطیلی که عرض آن ۸ و طول آن ۱۵ سانتی متر است، مساحتی برابر۱۲۰ سانتی متر مربع دارد. ریشه دوم ۱۲۰ یا ۱۰/۹۵ بعدی از یک مربع است که مساحت آن با مساحت مستطیل ۸ ضرب در ۱۵ برابر است. به همین ترتیب حاصل‌ضرب سه عدد، مبین یک حجم است و ریشه سوم حاصل‌ضرب آنها بعدی از یک مکعب با همان حجم است (فرگوسن، تاکانه، ۱۳۸۰: ۸۱).

دانلود
نمونه سوالات
استخدامی
ایران استخدام

رابطه بین سه میانگین فوق به‌صورت زیر است:

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

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 ٦٥ درجه فارنهایت است. اگر صرفاً با نگاه کردن به میانگین نتیجه بگیریم که این مکان برای زندگی مناسب است، خطا کرده ایم. زیرا این عدد مقدار متوسط ماه‌های زمستان و تابستان است. ما نیاز داریم تغییرپذیری درجات حرارت را بدانیم. یعنی درجه حرارت روزانه چقدر از متوسط ٦٥ درجه فارنهایت متفاوت است. از این رو باید شاخص‌های پراکندگی که نشان می دهد داده ها از میانگین چقدر دور و یا به آن نزدیک هستند را محاسبه کنیم. شاخص‌های پراکندگی، میزان پراکنده بودن نمرات حول و حوش مرکز داده‌ها را نشان می‌دهد که به ترتیب عبارتند از: دامنه تغییرات- انحراف متوسط، انحراف چارکی، واریانس و انحراف معیار.

دامنه تغییرات (R)

دامنه تغییرات یک شاخص پراکندگی و درواقع بی‌ثبات‌ترین شاخص و حساس ترین شاخص پراکندگی است که با داده های فاصله ای بایستی بکار رود و تفاضل بین کوچکترین و بزرگترین عدد در توزیع است (بدون در نظر گرفتن حدود واقعی اعداد)

R = max-min R = max-min+١

مثال: در توزیع نمرات ٣-٢٠-١٤-١٥-٩-٦-٥ دامنه تغییرات را محاسبه کنید:

R= ٢٠-٣+١=١٨ R= ١٨

نکته

١- دامنه تغییرات، برای نمونه های بزرگ شاخص بی‌ثباتی است.

٢- واریانس نمونه برداری دامنه تغییرات برای نمونه های کوچک، بزرگتر از واریانس نمونه برداری انحراف معیار نیست. ولی سریعاً با زیاد شدن N افزایش می یابد.

٣- دامنه تغییرات به جز 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 استفاده می شود.

مثال: برای نمرات ٦٠٠-٣٢-١٥-٩-٦-٥-٣ انحراف چارکی مناسب است.

$Q_١$ md $Q_٣$

دانلود نمونه سوالات استخدامی
ایران استخدامی

$$SIRQ = \frac{Q_۳ - Q_۱}{۲} = \frac{\text{چارک اول} - \text{چارک سوم}}{۲}$$

$$Q = \frac{۳۲ - ۵}{۲} = ۱۳/۵$$

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 تأثیر 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. $Q_۲$

سؤال: میانه اعداد سمت چپ و راست را محاسبه کنید. $Q_۲$ و $Q_۱$ مثال:

۶ – ۸ – ۹ – ۱۱ – ۱۲ – ۱۳ – ۱۴ – ۱۷

$Q_۱$=۸/۵ $Q_۲$=۱۱/۵ $Q_۳$= ۱۳/۵

نکته ۱

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.

نکته۲: پس از محاسبه نمره چارک اول و سوم انحراف چارکی بدست می آید در صورتی که:

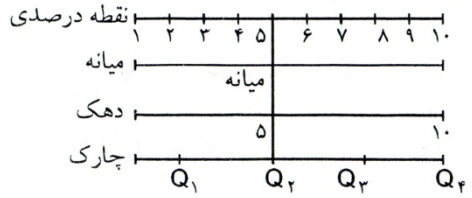
$Q_۳ - Q_۲ < Q_۲ - Q_۱$ کجی منفی $Q_۳ - Q_۲ > Q_۲ - Q_۱$ کجی مثبت $Q_۳ - Q_۲ = Q_۲ - Q_۱$ متقارن

نقطه درصدی P_n

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

$P_{۰/۷۵} = Q_۳$ $P_{۰/۱۰} = D_۱$ $P_{۰/۲۰} = D_۲$ $P_{۰/۵۰} = md = Q_۲$ $P_{۰/۲۵} = Q_۱$

دهک هم مثل نقطه درصدی، موقعیت نمره را در مقیاس۱۰ تایی نشان می دهد و چارک هم موقعیت نمره را در مقیاس ۲۵ تایی در چهار موقعیت نشان می دهد، یعنی:



دانلود
نمونه سوالات
استخدامی
ایران استخدامی

فرمول محاسبه میانه و دیگر نقاط

$$md = L + \left(\frac{\frac{N}{۲} - F_c}{F_i}\right) i \qquad Q_۱ = L + \left(\frac{\frac{۱N}{۴} - F_c}{F_i}\right) i \qquad Q_۳ = L + \left(\frac{\frac{۲N}{۴} - F_c}{F_i}\right) i$$

$$D_۱ = L + \left(\frac{\frac{۱N}{۱۰} - F_c}{F_i}\right) i \qquad D_۶ = L + \left(\frac{\frac{۶N}{۱۰} - F_c}{F_i}\right) i \qquad P_{۴۰} = L + \left(\frac{\frac{۴۰N}{۱۰۰} - F_c}{F_i}\right) i$$

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

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

$$Q_۳ = L + \frac{\frac{۳N}{۴} - F_c}{F_i}.i \qquad N = \frac{۳ \times ۱۰۰}{۴} = ۷۵$$

L=۱۹/۵ $\qquad F_c$=۵۴

F_i=۲۹ $\qquad$ i= ۵

$$Q_۳ = ۱۹/۵ + \left(\frac{۷۵ - ۵۴}{۲۹}\right) \times ۵ = ۲۳/۱۲$$

$$Q_۱ = L + \left(\frac{\frac{N}{۴} - F_c}{F_i}\right) i \qquad F_c = ۱۵ \qquad L = ۹/۵$$

$$Q_۱ = ۹/۵ + \left(\frac{۲۵ - ۱۵}{۱۸}\right) \times ۵ = ۱۲/۲۸$$

(انحراف چارکی) $\quad Q = \frac{Q_۳ - Q_۱}{۲} = \frac{۲۳/۱۲ - ۱۲/۲۸}{۲} = ۵/۴۲$

$$D_۲ = L + \left(\frac{\frac{۲N}{۱۰} - F_c}{F_i}\right) i \qquad L = ۹/۵ \qquad F_c = ۱۵$$

$$D_۲ = ۹/۵ + \left(\frac{۲۰ - ۱۵}{۱۸}\right) \times ۵ \quad \Rightarrow \quad D_۲ = ۱۰/۸۹$$

$$P_{N(۰/۶۸)} = L + \left(\frac{\frac{۶۸n}{۱۰۰} - F_c}{F_i}\right) i \qquad F_c = ۵۴ \qquad L = ۱۹/۵$$

$$I = ۵ \qquad F_i = ۲۹ \qquad \frac{۰/۶۸ \times ۱۰۰}{۱۰۰} = ۶۸$$

$$P_{۰/۶۸} = ۱۹/۵ + \left(\frac{۶۸ - ۵۴}{۲۹}\right) \times ۵ \quad \Rightarrow \quad P_{۰/۶۸} = ۲۱/۹۱$$

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.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

$$P_R = (\frac{F_c + \frac{F_i}{۲}}{n}) \times ۱۰۰$$

نکته: رتبه درصدی موقعیت فرد را در گروه و نقطه درصدی موقعیت نمره فرد را در داخل نمره ها (X)، نشان می‌دهد. بطور 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.

مثال: اگر بخواهیم رتبه درصدی عدد ۲۲ را در جدول فوق مشخص کنیم، خواهیم داشت.

$$P_R = (\frac{F_c + \frac{F_i}{۲}}{n}) \times ۱۰۰ \qquad P_R = (\frac{۵٤ + \frac{۲۹}{۲}}{۱۰۰}) \times ۱۰۰ = ٦۸/۵ \qquad P_R = ٦۸/۵$$

نکته۱: در توزیع های نامتقارن اغلب از میانه به عنوان شاخص مرکزی و انحراف چارکی به عنوان شاخص پراکندگی استفاده می شود.

نکته۲: اگر انحراف چارکی اندازه ها برابر صفر باشد، **۵۰** درصد اندازه‌هایی که در وسط قرار گرفته اند، با هم برابرند. درنتیجه چارک‌های یکم و دوم و سوم با هم برابرند و برعکس.

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

حدود طبقات	۲۰-۵۰	۲۰-٤۰	٤۰-٦۰	٦۰ و بیشتر
فراوانی	۲۵	۳۵	۳۰	۱۰

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

پاسخ: گزینه (الف) صحیح است.

انحراف متوسط (*MD*)

انحراف متوسط، یک شاخص پراکندگی است که به آن میانگین قدر مطلق انحراف نمره از میانگین گفته می‌شود.

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

مثال: برای اعداد ۵-٤-۳-۲-۱ خواهیم داشت:

X	$X-\overline{X}$	$\lvert X - \overline{X} \rvert$
۵	۲	۲
٤	۱	۱
۳	۰	۰
۲	-۱	۱
۱	-۲	۲
		$\sum \lvert X - \overline{X} \rvert = ٦$

$$md = \frac{٦}{۵} = ۱/۲$$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

نکات مهم

۱- با این شاخص عملیات جبری را نمی‌توان انجام داد.

۲- در انحراف متوسط علائم اعداد و در انحراف چارکی کلیه اعداد مورد مطالعه قرار نمی‌گیرند.

۳- تأثیر انحرافات بزرگ را در شرایطی که تعداد زیادی انحرافات کوچک در برابر تعداد کمی انحراف بزرگ باشد، نشان نمی دهد (مهمترین کاستی).

٤- انحراف متوسط اعداد ثابت صفر است.

واریانس (s^2)

یک شاخص پراکندگی است که میانگین مجموع مجذورات انحراف نمره از میانگین می باشد؛ یعنی:

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

محاسبه واریانس از راه انحراف نمره از میانگین

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

مثال: برای اعداد ۵-٤-۳-۲-۱

X	$X-\overline{X}$	X^2
۵	۲	٤
٤	۱	۱
۳	۰	۰
۲	-۱	۱
۱	-۲	٤
	$\sum\lvert X-\overline{X}\rvert = ۰$	$\sum X^2 = ۱۰$

$$S^2 = \frac{۱۰}{۵} = ۲$$

محاسبه واریانس از راه نمرات خام

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

نکته: هنگامی که واریانس را در نمونه محاسبه می کنیم در مخرج کسر n-۱ که به آن درجه آزادی (df) گفته می‌شود استفاده می کنیم.

درجه آزادی: تعداد مشاهداتی است که می تواند آزادانه تغییر کند. بنابراین درجه آزادی واریانس df =n-۱ می‌باشد.

نکته مهم: واریانس ۵ عدد متوالی در نمونه برابر ۲/۵ و واریانس ۵ عدد متوالی در جامعه برابر ۲ می باشد.

X	X^2
۵	۲۵
٤	۲٦
۳	۹
۲	٤
۱	۱
$\sum X = ۱۵$	$\sum X^2 = ۵۵$

$$S^2 = \frac{۵۵ - \frac{(۱۵)^2}{۵}}{۵} = ۲$$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

انحراف معیار (انحراف استاندارد) S

جذر واریانس را انحراف معیار می گویند، یعنی در مثال فوق: $S = \sqrt{۲} = ۱/٤۱$

نکته: توان دوم انحراف معیار، واریانس نام دارد. انحراف معیار، باثبات‌ترین شاخص پراکندگی است. هرچقدر 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 می‌شود که میانگین به‌عنوان شاخص مرکزی مورد استفاده قرار می‌گیرد. کلیه شاخص‌های پراکندگی (دامنه تغییرات، انحراف متوسط، انحراف معیار) با مقیاس حداقل فاصله ای بکار می روند.

ثبات پراکندگی به ترتیب

بیشتر کمتر

انحراف معیار ← واریانس ← انحراف متوسط ← چارک متوسط ← دامنه تغییرات

$R \leftarrow Q \leftarrow MD \leftarrow S^۲ \leftarrow S$

نکته مهم $S > MD > Q$

تأثیر چهار عمل اصلی در واریانس

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

تفریق $S^۲_۱ = S^۲$ جمع $S^۲_۱ = S^۲$

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

تقسیم $S^۲_۱ = S^۲ : C^۲$ ضرب $S^۲_۱ = S^۲ . C^۲$

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

$S^۲_۱ = ۲ \times ۳^۲ = ۱۸$

اگر واریانس ۵ و عدد ثابت ٦ با کلیه نمرات جمع شود، واریانس جدید برابر است با: $S^۲_۱ = ۵$ (تغییری نمی کند).

تأثیر 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 می شود.

تفریق $S_۱ = S$ جمع $S_۱ = S$

تقسیم $S_۱ = S : C$ ضرب $S_۱ = S . C$

مثال: اگر انحراف معیار یک توزیع ٦ و یک نمره ثابت ۲ در کلیه نمرات توزیع ضرب شود، انحراف معیار جدید برابر است با: $S_۱ = ٦ \times ۲ = ۱۲$

اگر انحراف معیار یک توزیع ٦ و نمره ثابت ۳ با کلیه نمرات توزیع جمع شود انحراف معیار جدید برابر است با: $S = ٦$ (تغییری نمی کند).

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

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

٤- واریانس در مجذور آن عدد ضرب یا تقسیم می شود.

۵- واریانس اعداد ثابت صفر است.

٦- 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 که میانگین جوامع برابر باشد که در آن صورت واریانس آنها هم برابر است.

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

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

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

$$S_{\bar{x}_1-\bar{x}_2}^2 = S_{\bar{x}_1}^2 + S_{\bar{x}_2}^2 = \frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}$$

۹- واریانس مجموع برابر است با:

$$S_{(X+Y)}^2 = S_X^2 + S_Y^2 \pm 2Cov(X,Y) = (S_X^2 + S_Y^2) 2rS_X S_Y$$

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:

$$S_c = \sqrt{25 - \frac{10^2}{12}} = 4/97$$

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.

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

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

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؟

الف) $V = \frac{۲۰۰۰}{۱۰۰۰۰} \times ۱۰۰ = ۲۰$ ب) $V = \frac{۱۰۰۰}{۱۱۰۰۰} \times ۱۰۰ = ۹$

پاسخ: نوع (ب) بهتر است، زیرا هم میانگین عمر آن بیشتر است و هم ضریب تغییر آن کوچکتر.

موارد استفاده

زمانی که دو یا چند جامعه در مقایسه با هم دارای مشاهدات ناهمگون از نظر واحد اندازه گیری باشند، مانند یک جامعه برحسب متر و یک جامعه برحسب اینچ، و یا چند جامعه دارای میانگین های متفاوتی باشند، استفاده می‌شود. گاهی نیز مقیاس صفت مورد اندازه‌گیری در دو جامعه یکسان است ولی بزرگی مشاهدات آنها به طور قابل ملاحظه ای تفاوت دارد. مانند مقایسه پراکندگی سود و زیان در صنایع دستی با صنایع سنگین.

نکته۱: ضریب تغییرات اعداد ثابت برابر صفر است.

پرسش: برای تعیین آنکه در۳۰ 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

پاسخ: گزینه (ب) صحیح است.

گشتاوردهای پیرامون میانگین

گشتاور اول: $m_۱ = \frac{\sum(X-\overline{X})}{N} = ۰$

گشتاور دوم: $m_۲ = \frac{\sum(X-\overline{X})^۲}{n} = S^۲$

گشتار سوم: $m_۳ = \frac{\sum(X-\overline{X})^۳}{n}$

گشتاور چهارم: $m_٤ = \frac{\sum(X-\overline{X})^٤}{n}$

گشتاور اول همیشه صفر است. گشتاور دوم همان واریانس است. گشتاور سوم برای محاسبه چولگی (کجی) بکار می رود، یعنی:

$$SK = \frac{m_۳}{m_۳\sqrt{m_۳}}$$

گشتاور چهارم برای محاسبه کشیدگی بکار می رود، یعنی:

$$Kg = \frac{m_٤}{(m_۳)^۲} - ۳$$

نکته۲: کشیدگی معیاری است بدون واحد که ارتفاع را نشان می دهد و رابطه معکوس با پراکندگی دارد.

اگر k=۰ باشد، کشیدگی هم اندازه و هم ارتفاع توزیع نرمال است، اگر k >۰ باشد، از نرمال بزرگتر و پراکندگی کمتر و اگر k < ۰ باشد، از نرمال کوتاهتر و پراکندگی بیشتر است.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

فرمول کجی پیرسون

$$g_1 = \frac{\overline{X} - m_0}{S}$$

نکته

۱- اگر $|sk| \leq ۰/۱$ باشد، تقریباً کجی وجود ندارد و جامعه نرمال است.

۲- اگر $۰/۱ \leq |sk| \leq ۰/۵$ باشد، چولگی موجود اندک ولی غیر قابل اغماض است. درحقیقت جامعه ازنظر تقارن اندکی با توزیع نرمال متفاوت است.

۳- اگر $|sk| > ۰/۵$ باشد، چولگی زیاد و غیر قابل اغماض است. به‌عبارت دیگر جامعه ازنظر قرینگی دارای تفاوت فاحشی با توزیع نرمال است.

نمره های استاندارد

نمره های استاندارد موقعیت فرد را در گروه معین می کنند. با داده‌های مقیاس حداقل فاصله ای کاربرد دارند. در تبدیل نمرات خام به نمره استاندارد از فرمول مقابل استفاده می کنیم:

$$Z = \frac{X - \overline{X}}{S}$$

مثال: در امتحانی که میانگین آن ۳۸ و انحراف معیار آن ۳ باشد، کسی که نمره ٤٤ گرفته است دارای $Z = ۲$ می‌باشد.

$$Z = \frac{٤٤ - ۳۸}{۳} = \frac{٦}{۳} = ۲$$

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

$$X = \overline{X} - Z_{\frac{a}{۲}} S_{\overline{X}}$$

مثال: کسی که نمره او $Z = ۲$ باشد، نمره t او برابر خواهد بود با:

$$t = ۵۰ + ۲ \times ۱۰ = ۷۰$$

نمره های استاندارد	میانگین	انحراف استاندارد
Z	۰	۱
T	۵۰	۱۰
CEEB	۵۰۰	۱۰۰
نه‌گانه	۵	۲
هوش وکسلر	۱۰۰	۱۵
هوش بینه	۱۰۰	۱٦
AGCT ارتش آمریکا	۱۰۰	۲۰

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

هوش وکسلر $= ۱۰۰ + ۱۵Z$ $\quad$ $t = ۵۰ + ۱۰Z$ $\quad$ $CEEB = ۵۰۰ + ۱۰۰Z$

نه‌گانه $= ۵ + ۲Z$ $\quad$ هوش بینه $= ۱۰۰ + ۱٦Z$ $\quad$ $AGCT = ۱۰۰ + ۲۰Z$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

مثال:

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

$$Z=\frac{X-\overline{X}}{S}=\frac{۸۰۰-۵۰۰}{۱۰۰}=۳$$

۲- فردی که در یک آزمون نمره نه‌گانه او ۳ شده است، نمره Z او چقدر است؟

$$\text{stanine}=۵+۲z\rightarrow Z=\frac{۳-۵}{۲}=-۱$$

۳- نمره IQ فردی ۱٤۰ شده است. نمره Z او چقدر است؟

$$Z=\frac{۱٤۰-۱۰۰}{۱۵}=۲/۷$$

٤- نمره t فردی ۳۰ شده است. نمره Z او چقدر است؟

$$Z=\frac{۳۰-۵۰}{۱۰}=-۲$$

t = ۵۰+ ۱۰ z

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

$$X=\overline{X}+Z.S\Rightarrow ۵۰۰+(۲\times ۱۰۰)=۷۰۰$$

٦- نمره Z فردی ۱/۵ می باشد. نمره نه‌گانه او چقدر است؟

$$X=۵+(۱/۵\times ۲)=۸$$

۷- نمره Z فردی۲- است. نمره AGCT او چقدر است؟

X=۱۰۰+(-۲)×۲۰=٦۰

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

t = ۵۰+(-۱/۵)×۱۰ =۳۵

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

$$Z=\frac{X-\overline{X}}{S}\Rightarrow\frac{۳۵۰-۵۰۰}{۱۰۰}=۱/۵$$

$$X=\overline{X}+ZS\Rightarrow ۱۰۰+(-۱/۵)\times ۱۵=۷۷/۵$$

۱۰- نمره نه‌گانه فردی ۸ شده است، نمره t وی چقدر است؟

$$Z=\frac{۸-۵}{۲}=۱/۵$$

t = ۵۰+ ۲/۵×۱۰ =۷۵

۱۱- نمره IQ فردی ٤۵ شده است. نمره t او چقدر است؟

$$Z=\frac{٤۵-۱۰۰}{۱۵}=-۳/٦۷$$

t = ۵۰+(-۳/٦۷)×۱۰ =۱۳

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

منحنی طبیعی و سطوح زیر منحنی

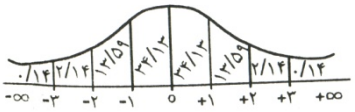
منحنی طبیعی (زنگوله ای یا گوس) یک توزیع طبیعی است.

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 در میانگین است.

٤- در منحنی طبیعی میانگین و میانه و نما با هم برابر هستند.



به‌طوریکه می بینیم مشخص می شود که ۳٤/۱۳ درصد افراد بین ۱+ تا $\overline{X}$ قرارمی‌گیرند. همینطور ۱۳/۵۹ درصد افراد بین ۲+ تا ۱+ قرار می‌گیرند. و ۲/۱٤ درصد افراد بین ۳+ تا ۲+ و ۰/۱٤ درصد افراد بین ∞+ تا ۳+ قرار می‌گیرند. این درصدها برای طرف چپ منحنی به دلیل تقارن هم صادق است. اگر بخواهیم رتبه 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 ۱= Z چقدر است؟ ۵۰+۳٤/۱۳ ⇒ ۸٤/۱۳ درصد

رتبه درصدی ۱- = Z چقدر است؟ ۵۰-۳٤/۱۳=۱۵/۸۷

درصورتیکه سطوح بین دو نمره Z خواسته شود دیگر ۵۰ کاربردی ندارد و سطوح زیر منحنی محاسبه می‌شود.

مثال: سطوح زیر منحنی بین ۱± چقدر است؟ ۳٤/۱۳+۳٤/۱۳=٦۸/۲٦ درصد

سطوح دیگر عبارتند از:

$\overline{X}$ تا ۲+ ⇒ ٤۷/۷۲ درصد

$\overline{X}$ تا ۲- ⇒ ٤۷/۷۲ درصد

$\overline{X}$ تا ۳+ ⇒ ٤۹/۸٦ درصد

$\overline{X}$ تا ۳- ⇒ ٤۹/۸٦ درصد

سطوح بین مثبت و منفی

۲- تا ۱+ ⇒ ۸۱/۸۵ درصد

۳+ تا ۳- ⇒ ۹۹/۷۷ درصد

۲+ تا ۲- ⇒ ۹۵/٤٤ درصد

درصورتیکه بصورت اعشاری باشد با روش زیر عمل می کنیم:

رتبه درصدی ۱/۵= Z چقدر است؟

$$۵۰ + ۳٤/۱۳ + ۱۳/۵۹ \times \frac{۵}{۱۰} = ۹۰/۹۳$$

رتبه درصدی ۲/۵- = Z چقدر است؟

$$۵۰ - \left[۳٤/۱۳ + ۱۳/۵۹ + ۲/۱٤ \times \frac{۲۵}{۱۰۰} \right] = ۱/۷۵$$

رتبه درصدی ۱/۰۲= Z چقدر است؟

$$۵۰ + ۳٤/۱۳ + ۱۳/۵۹ \times \frac{۲}{۱۰۰} = ۸٤/٤۰$$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

سطوح زیر منحنی بین Z = ۱/۷۵ تا Z = −۱/۵ چقدر است؟

$$P_R = -۱/۵ \Rightarrow \left[۳٤/۱۳ + ۱۳/۵۹ \times \frac{۵}{۱۰} \right] = ٤۰/۹۳$$

$$P_R = ۱/۷۵ \Rightarrow \left[۳٤/۱۳ + ۱۳/۵۹ \times \frac{۷۵}{۱۰} \right] = ٤٤/۳۲ \qquad ٤۰/۹۳+٤٤/۳۲ = ۸۵/۲۵$$

نکات مهم

۱- استفاده از روش بالا حدود رتبه درصدی را نشان می دهد. برای محاسبه دقیق باید از جداول مربوط که سطوح بین Z های مختلف را ارائه کرده استفاده کرد. (پیوست۱)

۲- انتقال از نمره های خام به نمره تراز Z شکل نمره را تغییر نمی دهد.

۳- اگر توزیع نمره‌های خام دارای کجی باشد، توزیع Z نیز دارای کجی است.

٤- برخلاف رتبه های درصدی اختلاف در نمره های Z اختلاف در نمره های خام را نشان می دهد.

۵- نسبت اختلاف نمره‌ها در توزیع اصلی یا نمره‌های خام، مساوی نسبت اختلاف بین نمره‌های Z آنها است. بنابراین فاصله بین اندازه نمره های اصلی در تبدیل به نمره های Z تغییر نمی کند.

٦- در نمره های نه‌گانه، ۹ فاصله یا واحد وجود دارد که هر فاصله یا واحد آن، مساوی نصف واحد انحراف معیار است. واحد میانی (پنجمین نمره نه‌گانه) نقطه میانی توزیع را نشان می دهد که دربرگیرنده ۲۰ درصد موارد است. چهارمین، سومین، دومین، و نخستین نمره نه‌گانه، از مرکز توزیع به سمت کرانه انتهایی پایین توزیع، به ترتیب ۱۷، ۱۲، ۷، ٤ درصد از موارد را دربرمی‌گیرد. (و برعکس) نمره ۹ گانه مانند رتبه درصدی نقطه ای بر روی مقیاس نیست، بلکه دربرگیرنده محدوده‌ی وسیعی از توزیع است. به همین دلیل نسبت به اختلاف‌های قابل‌توجه در طول مقیاس حساس نیست، اما نسبت به اختلاف‌های موجود در کرانه‌های انتهایی توزیع، به صورت گمراه‌کننده‌ای حساس است. (آیزاک، ۱۳۸٤: ۱۱۲).

۷- نمره های کلاسی مانند نمره های هوشی دارای توزیع بهنجار هستند. و بیشتر برای دوره ابتدایی که مطالب از پیوستگی بیشتری برخوردارند مناسب است. (آیزاک، ۱۳۸٤: ۱۱۵).

همبستگی

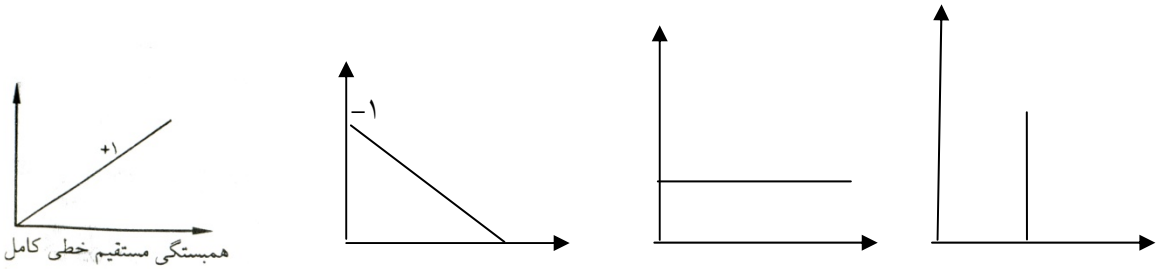
همبستگی بررسی رابطه بین دو متغیر با چند متغیر است، یعنی به‌دنبال این امر هستیم که آیا افزایش یا کاهش یک متغیر تأثیری روی متغیر دیگر دارد. فنون توصیفی در اینگونه مطالعات (که در آنها فقط ارتباط بین متغیرها، بدون آنکه هیچ یک از آنها دستکاری یا کنترل شود، مورد بررسی قرار می‌گیرد) عبارتست از میزان همخوانی (مانند ضریب فای یا ضرایب مشابه دیگر) برای مواردی که دو متغیر اندازه‌گیری‌شده طبقه‌ای باشد. ضریب رتبه‌ای اسپیرمن برای مواقعی که دو متغیر مورد اندازه گیری در مقیاس ترتیبی بیان می‌شوند، و سرانجام ضریب همبستگی پیرسون برای مواقعی که دو متغیر مورد بحث دارای مقیاس فاصله‌ای باشند. هرکدام از اینها نشان دهنده قدرت ارتباط بین دو متغیر است (هومن، ۱۳۸۲: ۱۲۷).

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

اگر هدف بررسی یک متغیر مثل X روی متغیر دیگر مثل y باشد، بررسی همبستگی تک متغیری است. اصطلاحاً به پژوهش‌های تک متغیری موقعیت های مصنوعی گفته می شود. اگر هدف بررسی چند متغیرمثل $X_1X_2...X_n$ بر روی یک متغیر دیگر یعنیy باشد، بررسی همبستگی چند متغیری است. اگر هدف بررسی چند متغیر $X_1X_2...X_n$ بر روی چند $y_1...y_n$ باشد، همبستگی کانونی است. در این مطالعات متغیر مستقل را متغیر پیش‌بینی‌کننده و متغیر وابسته را متغیر پیش‌بینی‌شونده گویند.

جهت همبستگی و شدت همبستگی

جهت همبستگی: اگر افزایش یک متغیر با افزایش متغیر دیگر همراه 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.

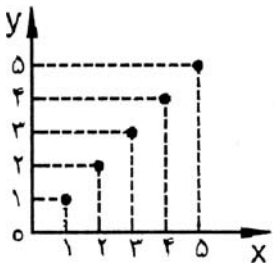


شدت همبستگی: دامنه همبستگی بین ۱+ و ۱- یعنی همبستگی ۱- معکوس و کامل، همبستگی ۱+ مستقیم و کامل است. مقدار همبستگی صفر تا ۱ می باشد (شیولسون، ۱۳۸۰: ۱۷۶). هر داده ای بین این دو حد ناقص است. یعنی همبستگی ۰/۹۹ ناقص است. هرچه همبستگی به قدر مطلق ۱ نزدیکتر باشد شدت آن بیشتر است. مثال: همبستگی ۰/۹۰- از همبستگی ۰/۷۰+ شدت بیشتری دارد.

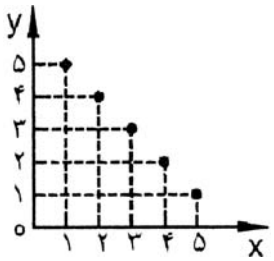
نکته: اگر ضریب همبستگی از ۰/۷۰ بالاتر باشد می گوئیم رابطه خطی است. اگر پایین تر باشد رابطه غیرخطی است و در صورت غیر خطی بودن رابطه بین دو متغیر از ضریب همبستگی ایتا η استفاده کنیم.

$$\eta = 1 - \frac{SS_{reg}}{SS_{res}}$$

نام	X	Y
حسن	۵	۵
علی	٤	٤
احمد	۳	۳
مهدی	۲	۲
محمد	۱	۱



X	Y
۵	۱
٤	۲
۳	۳
۲	٤
۱	۵



دانلود
نمونه سوالات
استخدامی
ایران استخدامی

نکته۱: برای محاسبه همبستگی ابتدا باید نمودار پراکندگی را رسم کرد. اگر نقاط به خط فرضی نزدیک باشند، همبستگی زیاد است.

نکته۲: چهار عمل اصلی (×÷+−) یک نمره در کلیه نمرات یک توزیع هیچ تأثیری روی ضریب همبستگی ندارد. مثلاً اگر همبستگی بین y,x (قد و وزن) ۰/۵ باشد همبستگی بین ۲x و y همان ۰/۵ خواهد بود.

مثال:

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 یکی اسمی و دیگری رتبه ای باشند: ضرایب عبارتند از کرامر- توافق c – لاندا- تاو گودمن و کراسکال.

- **همبستگی کرامر**: (جدول از ۲×۲ 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. (بی‌طرف، تندرو، محافظه کار). معادله آن عبارتست از: $\upsilon = \sqrt{\frac{\chi^۲}{n(L-۱)}}$ که n تعداد حجم نمونه و L تعداد ردیف یا ستون، هرکدام که تعداد کمتری دارند.

گرایش / سیاسی جنسیت	بیطرف	تندرو	محافظه کار	جمع
خانم ها	۳۲	۲۲	۵	۵۹
آقایان	۵	۳۰	۱۰	۴۵
جمع	۳۷	۵۲	۱۵	۱۰۴

$$\upsilon = \sqrt{\frac{۲۱.۰۹۸۲}{۱۰۴(۲-۱)}} = ۴۵$$

نکته: هنگامی که تعداد کمتر سطر یا ستون برابر ۲ باشد، فرمول بالا به علت این که مقدار L برابر با یک می‌شود، به‌صورت زیر که به همبستگی فی معروف است، تبدیل می شود.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

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

– **همبستگی c توافقی**: (جدول از ۲×۲ بیشتر باشد و حداقل یکی از متغیرها چند ارزشی است). معادله آن عبارتست از:

$$c=\sqrt{\frac{\chi^2}{\chi^2+n}}$$

– **ضریب همبستگی لاندا:** این ضریب تفسیر روشن تری از همبستگی می دهد. شما تا چه اندازه می‌توانید از روی جنس افراد، گرایشات سیاسی آنها را پیش بینی کنید و یا برعکس. این ضریب امکان چنین پیش‌بینی را می‌دهد. معادله آن عبارتست از: $\gamma=\frac{e_1-e_2}{e_1}$. در این رابطه e_1 اشتباه گروه‌بندی در موقعیت اول و e_2 اشتباه گروه‌بندی در موقعیت دوم می باشد.

ب) ضریب همبستگی در حالتی که هر دو متغیر دارای مقیاس رتبه ای باشند: عبارتند 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 تغییری ایجاد نمی شود.

– **ضریب گاما (G):** رابطه‌ی بین دو متغیر دو مقوله ای ترتیبی را به دست می دهد. مانند رابطه‌ی سطح تحصیلات مادران با نگرش آنان نسبت به تحصیلات دختران.

– **ضریب لامبدا (λ):** رابطه‌ی بین دو متغیر نامتقارن 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 می شود. مثلاً اگر k داور N شیئ را به m مقوله اسمی تبدیل کنند، آماره کپا میزان توافق داورها در مورد اینگونه طبقه بندی را نشان می دهد. مقدار آن بین صفر و یک است.

– **ضریب هماهنگی (w) یا توافق کندال (u):** در پژوهش هایی که در آنها بیش از دو مجموعه رتبه وجود دارد و مایلیم بدانیم که بین رتبه هایی که توسط m داور به n فرد داده شده تا چه حد توافق وجود دارد، از این دو شاخص استفاده می شود. وقتی داده ها به صورت زوج های همتا و نه به صورت رتبه جمع آوری شده باشد، روش ضریب توافق کندال (u) مناسب‌تر است (سرمد و دیگران، ۱۳۸۰: ۲۲٤). مقدار w بین صفر و یک است. مقدار u اگر k زوج باشد ($1-\frac{1}{k}$) و اگر k فرد باشد ($\frac{1}{k}$)، حداکثر برابر یک است.

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

m = تعداد داوران

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

n = تعداد افراد رتبه بندی شده

SS_T= واریانس مجموع رتبه ها

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

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

نکته: هرگاه پژوهشگر بخواهد اثر متغیر سوم را ثابت نگهدارد و میزان رابطه‌ی دو متغیر دیگر را مشخص کند، از ضریب همبستگی رتبه ای تفکیکی کندال باید استفاده کند (سرمد و دیگران، ۱۳۸۰: ۲۲٤)

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.

٤- مقیاس متغیرها 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

$$r_{xy} = \frac{\sum\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{S_x.S_y} = \frac{\text{cov}}{S_x.S_y} = \frac{Zx.Zy}{n}$$

ضریب همبستگی رتبه‌ای اسپیرمن

هنگامی که فقط دو دسته رتبه وجود داشته باشد (تعداد رتبه‌ها کم)، از ضریب اسپیرمن و یا تاو کندال می توان استفاده کرد. ضریب اسپیرمن از زوج های همتا استفاده می کند. زمانی کاربرد دارد که داده‌ها رتبه‌ای باشد. یکی از متغیرها یا هر دو متغیر و یا اینکه داده‌ها فاصله ای، ولی مفروضه های آمار پارامتریک رعایت نشده باشد. مثل: ترتیب تولد و رتبه در کلاس. وقتی رابطه غیرخطی باشد، داده های فاصله‌ای باید به رتبه تبدیل شوند. (شیولسون،۱۳۸۰: ۱۹۰). این ضریب برای رابطه های کاملاً غیر خطی مناسب نیست (شیولسون،۱۳۸۰: ۱۹٦).

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

X خلاقیت	Y پیشرفت تحصیلی	R_x	R_y	D	D^2
۱۲۰	۱۸	۲	۳	-۱	۱
۱۱۲	۱۹	٤	۲	۲	٤
۱۱۸	۱۲	۳	۵	-۲	٤
۱۲۵	۱۷	۱	٤	-۳	۹
۱۰۱	۱۹/۵	۵	۱	٤	۱٦
					$\sum D^2 = 34$

$$r_s = 1 - \frac{6\times 34}{5(5^2-1)} \Rightarrow 1 - \frac{204}{120} = 1 - 1/7 \Rightarrow 0/7$$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

نکته: زمانی که اعداد تکرار می‌شوند، برای تعیین رتبه اعداد مشابه باید بشرح زیر عمل کرد:

X	Y	R_x	R_y	D	$D^۲$
۱۸	۲۹	۵	٦	-۱	۱
۱۹	٤۲	۳	۲	۱	۱
۱۷	۳۵	۷/۵	٤	۳/۵	۱۲/۲۵
۱۸	۳۵	۵	٤	۱	۱
۲۵	۲۸	۱	۷	-٦	۳٦
۱۲	۲۱	۹	۸	۱	۱
۲۳	۳۵	۲	٤	-۲	٤
۱۱	٦٦	۱۰	۱	۹	۸۱
۱۸	۱۱	۵	۱۰	-۵	۲۵
۱۷	۱٤	۷/۵	۹	-۱/۵	۲/۲۵
					$\sum D^۲ = ۱٦٤/۵$

رتبه عدد تکراری $\frac{٤+۵-٦}{۳} = ۵$ $\qquad \frac{۷+۸}{۲} = ۷/۵$ $\qquad \frac{۳+٤+۵}{۳} = ٤$

$$r_s = ۱ - \frac{٦ \times \sum D^۲}{n(n^۲ - ۱)} = ۱ - \frac{٦ \times ۱٦٤/۵}{۱۰(۱۰^۲ - ۱)} = ۱ - \frac{۹۸۷}{۹۹۰} = ۰/۰۰۳$$

نکته: اگر $n \leq ۳۰$ از جدول استفاده کنید. اگر $n \geq ۳۰$ از فرمول زیر مقدار بحرانی را بیابید. اگر بین مقادیر مثبت و منفی بود همبستگی وجود ندارد.

$$r_s = \frac{\pm z}{\sqrt{n - ۱}}$$

در مثال بالا اگر فرض کنیم افراد نمونه ٦۵ نفر هستند. بنابراین داریم:

$r_s = \frac{\pm ۱/۹۶}{\sqrt{٦۵ - ۱}} = \pm ۰/۲۴۵$ اگر آماره آزمون $r_s = ۰/٤۰$ باشد، چون خارج از دامنه‌ی $\pm ۰/۲٤۵$ است فرض صفر رد می‌شود. یعنی بین رتبه ها همبستگی وجود دارد.

ضریب همبستگی دورشته‌ای نقطه‌ای[۱]

زمانی مورد استفاده قرار می گیرد که یک متغیر پیوسته و دیگری دوارزشی واقعی باشد یا یک متغیر پیوسته و دیگری دوارزشی و فرض نرمال بودن توزیع رعایت نشده باشد. مثل: همبستگی بین نمره کل آزمون و یک سؤال یا جنسیت و نمره در یک آزمون که می تواند از یک کمتر باشد تابع p و q است. در حالت بیشینه مقدار آن از ۱ کمتر است. (r = ۰/۷۹۷۸)

[۱]. *Combination*

دانلود
نمونه سوالات
استخدامی
ایران استخدام

۱- ضریب همبستگی دو رشته‌ای: یک متغیر پیوسته و دیگری دو ارزشی ساختگی است. می تواند از یک بیشتر باشد. توزیع نرمال و رابطه خطی است. منظور از ساختگی یعنی ما مثلاً نمرات ریاضی ۱۵-۱۵-۹-۸ را به صورت مردود – قبول طبقه بندی کنیم (۰ و ۱).

X	X'
۱۵	۱
۱۵	۱
۹	۰
۸	۰

۲- ضریب همبستگی فی: زمانی کاربرد دارد که هر دو متغیر دوارزشی واقعی باشند یا هر دو متغیر دوارزشی و فرض نرمال بودن رعایت نشده باشد و رابطه متقارن باشد. یعنی جای متغیرها عوض شود نتیجه تغییر نمی کند. مانند رابطه جنسیت و وضع سواد (بی‌سواد- باسواد) و یا همبستگی بین دو سؤال (در کتاب خانم سرمد برای این مورد تتراکوریک پیشنهاد شده است، ص ۲۲۲) که البته بیشتر فی بکار می رود.

۳- ضریب همبستگی تتراکوریک: زمانی کاربرد که یک یا هر 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 سؤال ۱.

٤- ضریب همبستگی کانونی: هنگامی که بخواهیم رابطه بین دو یا چند متغیر را با دو یا چند متغیر دیگر بررسی کنیم، از این ضریب همبستگی استفاده می‌کنیم.

۵- همبستگی تفکیکی (پاره ای): ضریب همبستگی دو متغیری است. برای داده های ناپیوسته که اثر متغیر سوم و بعدی را می خواهیم روی هر دو متغیر حذف کنیم، بکار می‌رود.

$r_{XY.Z_{fg}}$	$r_{XY.Z_F}$	$r_{XY.Z}$
مرتبه سوم	مرتبه دوم	مرتبه اول

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

٦- ضریب نیمه تفکیکی: اثر متغیر سوم و بعدی را روی یکی از متغیرها حذف می کنیم:

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

نکته: تفسیر ضریب همبستگی نباید برحسب درصد و نسبت باشد. مثلاً r_{xy} =۰/۷۰ هفتاد درصد رابطه بین متغیرها را تبیین نمی کند و r_{xy} =۰/۹۰ دقیقاً دو برابر r_{xy} =۰/٤٥ نیست.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

خلاصه همبستگی ها

روش	نماد	متغیر ۱	متغیر ۲	توضیح
گشتاوری پیرسون	R	پیوسته	پیوسته	باثبات‌ترین روش است.
تفاوت رتبه ها (rho)	ρ	رتبه ای	رتبه ای	بیشتر به جای همبستگی گشتاوری به کار برده می‌شود، وقتی که تعداد موارد کمتر از ۳۰ است.
تای کندال	τ	رتبه ای	رتبه ای	زمانی که تعداد موارد ۱۰ کمتر است، به جای rho به کار برده می شود.
دو رشته ای	r_{bis}	دو ارزشی ساختگی	پیوسته	مقدار این همبستگی گاهی اوقات از ۱ بیشتر می‌شود و خطای استاندارد آن بزرگتر از r است. از این روش در تجزیه و تحلیل سؤال استفاده می شود.
دو رشته ای گسترده	r_{wbis}	دو ارزشی ساختگی گسترده	پیوسته	زمانی به کار برده می شود که علاقمند به مطالعه افرادی هستید که در کرانه های انتهایی متغیر دو ارزشی قرار دارند.
دو رشته ای نقطه ای	r_{pbis}	دو ارزشی واقعی	پیوسته	ضریب همبستگی حاصل در این روش کوچکتر از r و بسیار کوچکتر از r_{pbis} است.
چهارخانه ای (تتراکوریک)	r_t	دو ارزشی ساختگی	دو ارزشی ساختگی	زمانی به کار برده می شود که هر دو متغیر را بتوان در نقطه‌ی بحرانی معینی به دو قسمت تقسیم کرد.
ضریب فی	φ	دو ارزشی واقعی	دو ارزشی واقعی	برای محاسبه همبستگی درونی بین سؤال ها در آزمون‌های چندگزینه ای یا دوگزینه ای به کار برده می شود.
ضریب توافقی	c	دو یا چند طبقه ای	دو یا چند طبقه ای	در برخی شرایط معین، با r_t قابل مقایسه است. ارتباط نزدیکی با خی دو دارد.
شاخص رابطه نامتقارن سامرز	d	هر دو متغیر دو مقوله ای یا رتبه ای		اینکه کدام متغیر مستقل یا وابسته نامگذاری می‌شود در نوع محاسبه‌ی شاخص تأثیر دارد.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

نکته

۱- **متغیر پیوسته**: به متغیری گفته می شود که پیوستار زیربنایی آن تمایل به توزیع طبیعی یا بهنجار دارد. نمونه‌هایی از این متغیر عبارتند از: وزن، توانایی یا پیشرفت تحصیلی که به وسیله آزمون های استاندارد اندازه‌گیری می‌شوند.

۲- **متغیر دوارزشی ساختگی**: وقتی حاصل می شود که یک متغیر پیوسته بر اساس یک معیار قراردادی که غالباً نزدیک مرکز داده‌هاست، به دو گروه تقسیم شود. به عنوان مثال، تقسیم بندی دانش آموزان به موفق- ناموفق، بالای متوسط، پایین متوسط، قبول- رد، و صمیمی- بی‌تفاوت در یک مقیاس نگرش‌سنج.

۳- **متغیر دوارزشی واقعی**: نقطه برش نسبتاً روشنی دارد (البته، نه ضرورتاً مطلق) به طوری که سرانجام داده های این متغیر به دو گروه تقسیم می‌شوند. نمونه‌هایی از این متغیر عبارتند از: مذکر- مؤنث، مرده- زنده، معلم- غیر معلم، مردود- نامردود، و سیگاری- غیرسیگاری. متغیرهای دیگری وجود دارند که برای محاسبه همبستگی، می‌توان آنها را دو ارزشی واقعی تلقی کرد. نظیر کوررنگی- غیرکوررنگی، الکلی- غیر الکلی و پاسخ های صحیح- غلط یک سؤال معین در یک آزمون. توزیع‌های زیربنایی دو ارزش های واقعی اگر دارای تفاوت های مطلق نباشند، دونمایی و یا به طور نسبی ناپیوسته هستند.

کوواریانس (واریانس مشترک)

کوواریانس تغییرپذیری مشترک بین دو متغیر را گویند (واریانس یک متغیر). مقدار و جهت رابطه بین دو متغیر را اندازه می گیرد. ولی دو محدودیت دارد. مقدار کوواریانس به مقدار تغییرپذیری نمره های X و Y بستگی دارد. اگر پراکندگی زیاد کوواریانس نیز زیاد باشد، درنتیجه نمی توان کوواریانس اندازه های مختلف را با هم مقایسه کرد. تفاوت در مقدار کوواریانس ممکن است در اثر رابطه بین متغیرها، تفاوت در انحراف معیارها و یا تفاوت در رابطه و انحراف معیارها باشد. شاخصی که از تفاوت در اندازه مقیاس ها تأثیر نپذیرد همبستگی است. این عمل تصحیح از رابطه الف بدست می آید (شیولسون، ۱۳۸۰: ۱۷۵).

$$\text{cov} = \frac{\sum(xy)}{n-۱} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-۱}$$

فرمول محاسبه

پرسش: اگر کوواریانس x و y صفر باشد، کدام عبارت درباره x و y 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) صحیح است. اگر کوواریانس صفر شد، نمی توان حکم به استقلال دو متغیر داد، چون ممکن است بین آنها رابطه‌ای غیرخطی وجود داشته باشد که بوسیله کوواریانس مشخص نمی شود. چون کوواریانس فقط نوع رابطه خطی دو متغیر را تعیین می کند (آذر، ۱۳۸۰: ۱۳۲).

مثال۱: اگر cov برابر ۵۵ و $S_y^۲$=٦٤ باشد، همبستگی بین y,x چقدر است؟

$$r_{xy} = \frac{cov}{S_x.S_y} = \frac{۵۵}{۸\times۹} = \frac{۵۵}{۷۲} = +۰/۷٦$$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

مثال۲: اگر $\sum(X-\overline{X})=۵$ و $\sum(Y-\overline{Y})=۸$ باشد و انحراف معیار x برابر۱۰ و انحراف معیار y برابر۱٦ باشد، ضریب همبستگی چقدر است؟

$$r_{xy}=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{S_x.S_y}=\frac{۵\times۸}{۱۰\times۱٦}=\frac{٤۰}{۱٦۰}=+۰/۲۵$$

مثال۳: اگر $Z_x=۲$ و $Z_y=۳/۵$ باشد و حجم نمونه برابر ٤۰ نفر، ضریب همبستگی چقدر است؟

$$r_{xy}=\frac{Zx.Zy}{n}=\frac{۲\times۳/۵}{٤۰}=\frac{۷}{٤۰}=+۰/۱۷۵$$

مثال٤: چنانچه میانگین وزن دانش آموزان (n =۱۰) برابر ۵۰ و میانگین قد همین گروه ۱۱۰cm باشد و فردی قد ۱۱۵cm و وزن ۵۸kg داشته باشد، میزان همبستگی بین قد و وزن او چقدر است؟ $S_۱=۵$ وزن و $S_۲=۳$ قد

قد $Zy=\frac{Y-\overline{Y}}{S}=\frac{۱۱۵-۱۱۰}{۳}=۱/٦۷$ وزن $Zx=\frac{X-\overline{X}}{S}=\frac{۵۸-۵۰}{۵}=۱/٦$

$$r_{xy}=\frac{Zx.Zx}{n}=\frac{۱/٦\times۱/٦۷}{۱۰}=+۰/۲۷$$

ضریب تعیین (درصد واریانس مشترک) $V=r^۲_{xy}\times۱۰۰$

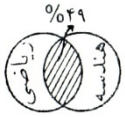
ضریب تعیین نشان دهنده میزان تأثیری است که متغیر X (مستقل) در متغیر Y (وابسته) ایجاد می کند یا به‌عبارتی با محاسبه این ضریب می توان تعیین کرد چند درصد از کل واریانس X ناشی از واریانس Y است.

$$r^۲_{xy}=\frac{\text{cov}^۲}{S^۲_x.S^۲_y}=\frac{\text{واریانس مشترک}}{\text{کل واریانس}}$$

نکته: ضریب تعیین هیچ وقت منفی نخواهد شد، زیرا برای محاسبه آن ضریب همبستگی مجذور می شود.

سؤال: اگر همبستگی بین نمره ریاضی و نمره هندسه ۰/۷ باشد، درصد واریانس را بدست آورده و با شکل نشان دهید.

$V=(۰/۷)^۲\times۱۰۰=٤۹$



۵۱٪ عوامل دیگر

عواملی که بر ضریب همبستگی تأثیر می گذارند:

۱- اساس رابطه از جامعه ای به جامعه دیگر فرق می کند. مثلاً در افراد بشر در سنین ۱٦-۱۰سالگی بین سن تقویمی و توانایی فیزیولوژکی همبستگی بالایی وجود دارد. ولی بین این دو متغیر در سنین ۲٦-۲۰ سالگی همبستگی وجود ندارد.

۲- پراکندگی متغیرها در جوامع مختلف متفاوت است. بدین معنی که هرچه تجانس بیشتر باشد (واریانس کمتر باشد)، همبستگی کمتر است. به عنوان مثال، در یک پژوهش اگر همه باهوش باشند، دامنه محدود و مقدار r کاهش می‌یابد.

۳- همبستگی بین دو متغیر تحت تأثیر همبستگی آنها با متغیر سوم قرار دارد. به عنوان مثال، همبستگی بین فیزیک و ریاضی ممکن است به دلیل همبستگی این متغیرها با هوش باشد.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

٤- استفاده از گروه های انتهایی ضریب همبستگی را زیاد می کند. نمره هـای انتهـایی در حجـم هـای کوچـک بـر همبستگی تأثیر زیاد می گذارد.

٥- اگر n افزایش یابد، احتمال معناداری r بیشتر است.

٦- وقتی متغیرها نامرتبط باشند، r کاهش می یابد.

٧- رابطه غیر خطی امکان دارد مقدار ضریب همبستگی پیرسون را به صفر نزدیک کند.

٨- محدودیت در دامنه تغییرات ضریب همبستگی را کاهش می دهد. برای وجود محدودیت در دامنه تغییر باید واریانس‌ها یا انحراف معیارهای متغیرهایی که ضریب همبستگی آنها محاسبه می گردد مورد بررسی قرار گیرند. کوچک بودن واریانس ها می تواند نشانه محدودیت در دامنه تغییر باشد (شیولسون، ١٣٨٠: ١٩٧).

٩- اگر واریانس کوچک باشد، تفسیر ضریب همبستگی باید با احتیاط لازم صورت گیرد.

نکته١: آزمون معناداری برای همبستگی در دو متغیر مستقل آزمون t فریدلی است. و اگر گروه‌ها مستقل است آزمون z فیشر و اگر ضریب همبستگی را در دو گروه وابسته مقایسه کنیم t استیودنت است. (فرگوسن، تاکانه، ١٣٨٠).

نکته٢: درجه آزادی همبستگی برابر است با: $d.f = n - ٢$

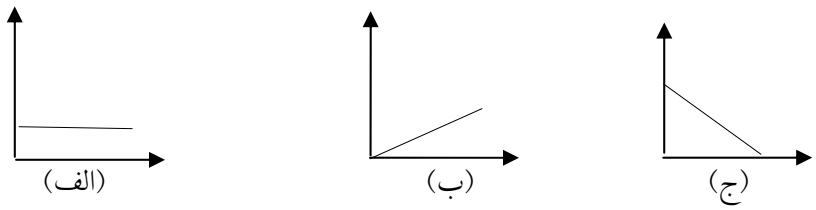
تحلیل همبستگی از نوع پیش بینی: رگرسیون و پیش بینی

زمانی که بین دو متغیر همبستگی وجود داشته باشد، می‌توان ازطریق رگرسیون مقدار یک متغیر (Y) را از روی یک متغیر دیگر (X) پیش بینی یا برآورد کرد، و هرچه همبستگی بین متغیرها بالاتر باشد، به همان اندازه پیش‌بینی دقیقتر است. (معدل نمره دروس دانشگاهی را از روی نمره آزمون استعداد پیش بینی می کنیم) این پیش بینی از 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 ها حول خط رگرسیون را کمینه می‌سازد. می‌توان معادله این خط را بدست آورد و سپس نمودار آن را رسم کرد. خط رگرسیون درحقیقت یک میانگین متحرک یا خط کمترین مجذورات است. رگرسیون درحقیقت معادله خط است. $Y=a+bx$

نکته: رابطه بین متغیر پیش‌بینی‌شونده (ملاک) (Y) و متغیر پیش‌بینی‌کننده (X) تابع علامت و شدت ضریب همبستگی است. اگر ١ = r_{xy} باشد، رگرسیون صفر و اگر ٠ = r_{xy} باشد، همبستگی کامل است.

پیش بینی Y از روی X

شیوه شهودی: که ازطریق ترسیم نمودار ممکن است. تعبیر و تفسیر آن تابع فرد مشاهده گر است.



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 پیش بینی را دارد:

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

۱- میانگین متغیر Y یعنی $\overline{Y}$ و شیب خط فرضی در نمودار پراکندگی. ۲- این نمودار نشان می دهد اگر شیب خط (ضریب زاویه) صفر باشد، 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 می‌یابد.

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

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

برای پیش بینی بطور کل 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.

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

پیش بینی با استفاده از معادله رگرسیون خطی

Y=a+bx

a= عرض از مبدأ (مقدار Y وقتی X صفر است).

B= شیب خط (مقدار متغیر در Y به ازای یک واحد تغییر در x)

$$b = \frac{Y_2 - Y_1}{X_2 - X_1}$$

از این معادله هنگامی استفاده می شود که همه نقاط یک توزیع مشترک روی یک خط مستقیم قرار می گیرند.

پیش بینی نمره های استاندارد (Z)

مقدماتی‌ترین روشی که در استفاده از ضریب همبستگی پیرسون برای پیش بینی به کار برده می شود نمره‌های استاندارد است.

$$Z_Y = (Z_X)(r_{XY})$$

مهم: اگر همبستگی کامل (۱) باشد، نمره استاندارد پیش بینی نیز کامل، در غیر این صورت غیرکامل می‌باشد. به عنوان مثال اگر نمره استاندارد Z برابر ۲ و همبستگی کامل (۱) باشد، نمره پیش بینی استاندارد نیز ۲ می باشد.

مهم: هنگامی که همبستگی بین دو متغیر کم باشد، نمره استانداردی که پیش‌بینی می‌کنیم، نزدیک به میانگین خواهد بود.

نکته۱: زمانی که همبستگی بین دو متغیر پایین باشد نمرات پیش‌بینی‌شده نزدیک به میانگین نمره پیش‌بینی‌شونده هستند تا نمره واقعی. به این پدیده رگرسیون می گویند.

نکته۲: میزان همبستگی بین دو متغیر حدود یا مقدار اتفاق رگرسیون را تعیین می کند.

نکته۳: پدیده رگرسیون اولین بار بوسیله گالتن مورد استفاده قرار گرفت. بر اساس مطالعات گالتن، فرزندان والدین بلندقد، بلندقد هستند اما نه به اندازه والدین خود. به همین ترتیب فرزندان والدین کوتاه‌قد، کوتاه‌قد هستند اما نه به کوتاهی والدین خود.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

نکته٤: خط رگرسیون، خطی است که خطاهای پیش بینی را به حداقل می رساند (خط حداقل مجذورها). یعنی اینکه مجموع مجذور فاصله Yها از خط رگرسیون کوچکتر از فاصله هر خط دیگری تا محور Yها می‌باشد. خط رگرسیون را خط برازنده نیز می نامند.

نکته۵: اختلاف بین نمره واقعی (Y) و نمره پیش بینی شده (y`) را خطای پیش بینی e می‌گویند.

$$e = y - y'$$

نکات بسیار مهم

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

۲- اگر همبستگی صفر باشد، رگرسیون کامل و اگر یک باشد رگرسیون صفر است.

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

٤- 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، همبستگی بین دو متغیر در حد متوسط و مقدار آن نزدیک به ۰/۵ است (آیزاک، ۱۳۸۱: ۱۷۲).

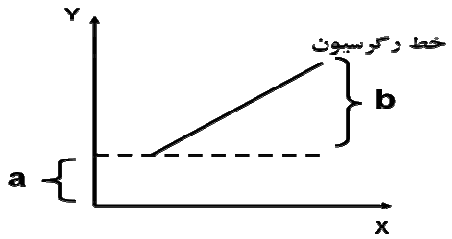
معادله خط رگرسیون

برای پیش بینی یک متغیر از روی متغیر دیگر از معادله خط رگرسیون استفاده می شود: Y'=a+bx که در آن:

B = شیب خط Y' = متغیری که می خواهیم پیش بینی کنیم

a= محل تلاقی خط رگرسیون با محور عرض‌ها یا عرض از مبدأ X = متغیری که مقدار آن را در اختیار داریم

نکته: بعضی مواقع مقدار a و b در اختیار است. در این صورت با جایگزینی مقادیر، محاسبه خیلی ساده می باشد.



مثال:

X = ۰	Y = ٤ +۲(۰) =٤
X = ۱	Y = ٤ +۲ (۱) =٦
X = ٤	Y = ٤ +۲ (٤) =۱۲
X = ۸	Y = ٤ +۲ (۸) =۲۰

روش محاسبه ضرایب a و b: $bxy = rxy\frac{S_X}{S_Y}$ $byx = rxy\frac{S_Y}{S_X}$ $b_{YX} = \frac{Cov_{XY}}{S_X^2}$

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

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

محاسبه a و b از روی داده های خام

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

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^{۲} - (\Sigma X)^{۲}}$$

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

X	Y	$X^{۲}$	$Y^{۲}$	XY
٦	٨	٣٦	٦٤	٤٨
٩	٩	٨١	٨١	٨١
٥	٧	٢٥	٤٩	٣٥
٢٠	٢٤	١٤٢	١٩٤	١٦٤

$$byx = \frac{٣(١٦٤) - (٢٠)(٢٤)}{٣(١٤٢) - (٢٠)^{۲}} = ٠/٤٦$$

$$a_y = \frac{٢٤ - (٠/٤٦)(٢٠)}{٣} = ٤/٩٣$$

$$y = ٤/٩٣ + ٠/٤٦(x)$$

خطای استاندارد برآورد

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

نکته: هرچه همبستگی بین متغیرها بیشتر باشد، خطای پیش بینی کمتر خواهد بود. این خطا به دو روش محاسبه می‌شود.

$$S_{YX} = \sqrt{\frac{\sum(Y-\bar{Y})^{۲}}{N-۲}} \quad \text{یا} \quad S_{yx} = \sqrt{\frac{\Sigma e^{۲}}{N}}$$

خطای استاندارد پیش بینی $S_{yx} = S_y\sqrt{۱ - r^{۲}xy}$

خطای پیش بینی = e ضریب همبستگی = rxy

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

رابطه پایایی و خطا

$S_E = S\sqrt{۱ - r_{xx}}$ 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.

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

چند متغیری: چند متغیر مستقل، چند متغیر وابسته را پیش‌بینی می کنند.

نکته۱: وقتی همبستگی چند متغیر x را با چند متغیر y محاسبه می کنیم، همبستگی کانونی یا قانونی گفته می شود.

نکته۲: مقدار همبستگی چندگانه معمولاً از همبستگی بین هریک از متغیرهای پیش بین و ملاک بیشتر است. (کازبی، ۱۳۸٦: ۲۷۸).

راهبردهای تحلیلی در انتخاب متغیرهای مستقل برای تشکیل معادله‌ی رگرسیون

۱- **راهبرد همزمان**: همزمان همه‌ی متغیرهای مستقل با هم وارد تحلیل می شوند.

۲- **رگرسیون گام به گام**: در این روش نحوه‌ی ورود متغیرها به تحلیل در اختیار محقق نیست. از روش گام به گام برای پیش بینی متغیر وابسته استفاده می شود. از این روش برای تبیین استفاده نمی‌شود.

۳- **رگرسیون سلسله مراتبی**: ترتیب ورود متغیرها با توجه به چارچوب نظری و توسط محقق مشخص شده است. چنانچه پژوهشگر بخواهد در پیش بینی پیشرفت تحصیلی از متغیرهای خلاقیت و هوش استفاده کند، می‌تواند با استفاده از این روش و با توجه به این که هوش بر خلاقیت تقدم دارد، عمل کند.

نکته۱: برای پی‌بردن به معنادار بودن آماره‌های محاسبه شده از آزمون‌های آماری زیر استفاده می‌شود: آزمون معنادار بودن R^2 و ΔR^2 که از توزیع F پیروی می کند و آزمون معناداری ضرایب رگرسیون که از توزیع T پیروی می‌کند.

نکته۲:دررابطه با ضرایب استاندارد نشده b و استاندارد شده β باید توجه داشت، هرگاه بخواهید اثر متغیرهای متفاوت را در یک جامعه‌ی واحد مقایسه کنید از β و هرگاه خواسته باشید اثر متغیرهای خاصی را در جوامع متفاوت مقایسه کنید از b استفاده نمایید (سرمد و همکاران، ۱۳۸۰: ۲۳۲).

روش های دیگر تحلیل رگرسیون

۱- آزمون χ^2 نسبت درست نمایی (آزمون G^2)

این آزمون مانند χ^2 معمولی برای داده های ۲ × ۲ به کار می رود. اما به دلایل زیر بر آزمون χ^2 معمولی برتری دارد:

- کمتر تحت تأثیر حجم نمونه قرار می گیرد.
- قابل افراز به مؤلفه هایی است که مانند مجموع مجذورات در تحلیل واریانس جمع پذیرند.
- در مدل لگاریم خطی که در آن با جداول چند بعدی سروکار داریم مورد استفاده قرار می گیرد.

۲- تحلیل ممیز[۱]

در مواردی استفاده می شود که متغیر وابسته اسمی و متغیرهای مستقل کمی باشند. در تحلیل ممیز باید گروه‌ها ناهمپوش و عضویت آزمودنی ها در آنها مشخص گردد.

برای مثال فرض می کنیم دانشگاهی می‌خواهد موفقیت یا شکست متقاضیان ورود به دوره‌ی دکترای رشته‌ی خاصی را پیش‌بینی کند. این دانشگاه با داشتن ویژگی‌های داوطلبان سال قبل مانند پیشرفت تحصیلی، توصیه‌نامه‌ها و نمرات امتحانات ورودی که به عنوان متغیر مستقل در نظر گرفته می شود، با تحلیل ممیز می‌تواند به پیش بینی دست بزند.

۱. *Discriminant*

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

۳- رگرسیون لجستیک

در مدل تحلیل ممیز همه‌ی متغیرهای پیش بین با مقیاس کمی (فاصله ای یا نسبتی) اندازه گیری شده و متغیر ملاک (وابسته) مقوله ای است. در صورتی که متغیرهای پیش بینی هم در مقیاس کمی و هم در مقیاس مقوله‌ای اندازه‌گیری شده باشند، از تحلیل ممیز نمی توان استفاده کرد. در این موارد از رگرسیون لوجستیک استفاده می شود که متغیر، متغیر وابسته مقوله ای و دو سطحی است. این دو مقوله معمولاً به عضویت یا عدم عضویت در یک گروه یا بلی و خیر اشاره دارد. در معادله‌ی رگرسیون معمولی از تعدادی متغیرهای پیش بین با ضرایبی (وزن ها) برای پیش‌بینی متغیر وابسته استفاده می شود. در رگرسیون لوجستیک آنچه که پیش بینی می شود یک احتمال است که ارزش آن بین ۰ و ۱ تغییر می کند. در این رگرسیون از مفهوم بخت استفاده می شود که عبارتست از نسبت احتمال وقوع یک پدیده بر احتمال عدم وقوع آن.

نکته: واژه‌ی کلیدی در رگرسیون لوجستیک سازه ای به نام لوجیت[۱] است که لگاریتم طبیعی بخت می باشد.

مثال: فرض کنید پژوهشگری می‌خواهد کمک‌کردن یا نکردن افراد را به دیگری برحسب احساس همدردی، احساس توانایی و قومیت فرد، پیش بینی کند. در این مثال احساس همدردی و توانایی در مقیاس فاصله ای و قومیت اسمی می باشد.

٤- مدل لگاریم خطی

در مواردی که تمام متغیرهای مدل در مقیاس اسمی اندازه گیری شده و داده ها به صورت فراوانی باشد، روش معمولی برای پی بردن به رابطه‌ی بین دو متغیر مقوله ای استفاده از آزمون χ^2 است. هرگاه تعداد متغیرهای مقوله‌ای بیش از دو باشد، تفسیر جداول خی دو χ^2 مشکل یا غیر ممکن می شود. در چنین مواردی می توان از مدل لگاریتم خطی استفاده کرد. که اساس آن شبیه روش رگرسیون چندگانه است. با این تفاوت که در این روش متغیر مستقل، متغیرهای مقوله‌ای و اثرمتقابل آنها، و متغیر وابسته لگاریتم طبیعی فراوانی های خانه‌ای است. مثال: اثر جنس، قومیت، سطح درآمد و اثرهای متقابل آنها بر فراوانی خانه های یک جدول توافقی.

نکته: همانند تحلیل واریانس اثرهای متقابل نامگذاری می شوند. به عنوان مثال در سه عاملی اثر متقابل (مرتبه سوم) نامیده می‌شود. مدل هایی که تمامی اثرات ممکن را دربرداشته باشد، مدل اشباع شده نامیده می شود. که مثال بالا این چنین است. چون ساده نیست مطلوب ترین روش محسوب نمی شود. تحلیل مدل لگاریتم خطی به دو صورت سلسله مراتبی و غیر سلسله مراتبی است.

نکته: تفاوت عمده مدل لگاریتم خطی با مدل رگرسیون لوجستیک که مدل لوجیت نیز نامیده می شود، این است که در مدل لوجیت یک متغیر مقوله ای دو سطحی به عنوان متغیر وابسته وجود دارد. درحالیکه در لگاریتم خطی متغیر وابسته، فراوانی خانه ها است (سرمد و دیگران، ۱۳۸۰: ۲٦۵).

---

۱. *Logit*

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

تحلیل پروبیت[1] و لوجیت

برای پاسخگویی به پرسش هایی از قبیل: چند ساعت مطالعه در هفته برای بدست آوردن نمره ای معادل ۸۰۰ در یک آزمون مانند تافل لازم است؟ که در چنین پرسش هایی متغیر تأثیرگذار که در یک مقیاس پیوسته اندازه‌گیری می‌شود (ساعات مطالعه) و یک متغیر برونداد (وابسته) که مقوله ای است و موفقیت یا شکست را اندازه‌گیری می‌کند، از این تحلیل‌ها استفاده می‌شود. این شیوه‌ها برای تبدیل داده‌های خام به داده‌های استاندارد بکار می‌روند. مقدار هر دوی آنها بین ۰/۲ و ۰/۸ تغییر کرده و مانند z عمل می کند. برای تبدیل به پروپیت، نمره‌ی z را با ۵ جمع می کنیم. برای تبدیل به لوجیت از فرمول زیر استفاده می شود.

$$\text{نمره لوجیت} = \frac{\ln \frac{p}{(p-۱)}}{۲} + ۵$$

Ln= لگاریتم طبیعی و p = نسبت پیامدهای موفقیت آمیز است.

مثال: اگر ۱۵۰ دانشجو ۳۰ ساعت در هفته مطالعه‌ی زبان انگلیسی کرده باشند و ۱۱۱ نفر از آنها توانسته باشند نمره‌ای بالاتر از ۱۲۰۰ به دست آورند، نسبت موفقیت برابر است با: $\frac{۱۱۱}{۱۵۰} = ۰/۷٤$

نمره‌ی z متناظر با آن (۰/۷٤ درصد) در توزیع نرمال برابر۰/٦٤ است. از این رو پروپیت این پیامد برابر است با: ۵/٦٤ = ۵+۰/٦٤. لوجیت این پیامد نیز عبارتست از:

$$\text{لوجیت} = Ln\frac{۰/۷٤(۱-۰/۷٤)}{۲} + ۵ = ۵/۵۲$$

تحلیل عاملی

برای بررسی روایی سازه از روش آماری تحلیل عاملی استفاده می شود که با تشکیل ماتریس همبستگی سازه‌ی موردنظر بررسی می‌شود. برای انجام تحلیل عاملی ابتدا باید اطلاعات موجود در ماتریس معنادار باشد. برای این منظور از آزمون χ^2 بارتلت استفاده می‌شود. معنادار بودن χ^2 و آزمون بارتلت[2] حداقل شرط لازم برای انجام محاسبات تحلیل عاملی است. در آزمون بارتلت فرض صفر این است که متغیرها فقط با خودشان همبستگی دارند. رد فرض صفر به این معنا است که ماتریس حاوی اطلاعات معناداری است. این آزمون را آزمون کرویت[3] گویند.

احتمالات و توزیع دو جمله ای

احتمال یک نسبت است که صورت آن تعداد دفعاتی است که پیشامد رخ داده و مخرجش کل کوشش ها است.

$$P = \frac{f}{N}$$

مثال: همیشه تعداد تولد پسرها بیشتر از دخترها است. در یک گروه معمولی ۲۰۵ نوزاد به‌دنیا آمده است. از بین این نوزادان ۱۰۵ نفر پسر هستند. اگر یک نوزاد به طور اتفاقی از بین این گروه انتخاب شود، احتمال اینکه این نوزاد پسر نباشد چقدر است؟

---

[1]. *Probit*
[2]. *Bartlett*
[3]. *Sphericity*

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

چون از ۲۰۵ نوزاد ۱۰۵ نفر پسر هستند پس ۱۰۰ نوزاد دخترند. درنتیجه: $P_c = \frac{۱۰۰}{۲۰۵} = ۰/٤۸۸$

نکته: احتمال یک پیشامد غیر ممکن صفر است. احتمال رخ دادن یک پیشامد مسلم یک است.

فضای نمونه: از تمام پیشامدهای ساده تشکیل می شود. یعنی فضای نمونه از تمامی پیشامدهایی تشکیل می‌شود که دیگر نمی‌توان آنها را به مؤلفه های ساده تر تبدیل کرد.

متمم رویداد A: شامل تمام پیشامدهایی می شود که در آنها پیشامد A اتفاق نمی افتد.

نکته: وقتی مقدار احتمال را بیان می کنیم باید یا باید عدد دقیق اعشاری را بیان کنیم یا اینکه عدد را تا ۳ رقم معنادار گرد کنیم. اگر عدد کسر ساده مانند $\frac{2}{3}$ است لازم نیست گرد شود.

پیشامد مرکب[1]: به پیشامدی گفته می شود که دارای ۲ یا بیشتر پیشامد ساده باشد.

فرمول فضای نمونه کل حالتهای ممکن: $S = P^n$

محاسبه‌ی احتمال

در احتمال محاسبات بر اساس دو قانون جمع و ضرب انجام می شود. قانون جمع زمانی به کار می رود که یکی از دو حادثه اتفاق بیفتد، درصورتیکه قانون ضرب در شرایطی به کار برده می شود که هر دو حادثه با هم اتفاق بیفتند.

دو رویداد ناسازگار: به حوادثی گفته می شود که وقوع یکی از آنها مانع وقوع دیگری باشد. برای مثال در پرتاب سکه، نشستن یک روی سکه مانع نشستن طرف دیگر می شود. در این صورت احتمال اینکه یکی از دو رویداد اتفاق بیفتد مساوی است با مجموع احتمال هریک از آنها.

$$P_{(A \text{ or } B)} = P_A + P_B \quad \Rightarrow \quad P=(A\cap B)=۰$$

دو رویداد سازگار: هنگامی که دو رویداد سازگارند امکان وقوع آنها به صورت همزمان وجود دارد.

$$P_{(A \text{ or } B)} = P_A + P_B - P_{(A\cap B)}$$

مثال: اگر از یک دسته ورق دو کارت بکشیم، احتمال آمدن ۱۰ تا دل چقدر است؟

$$P(A \text{ یا } B) = \frac{٤}{۵۲} + \frac{۱۳}{۵۲} - \frac{۱}{۵۲} = \frac{۱٦}{۵۲}$$

قاعده ی ضرب احتمال ها

رویدادهای مستقل: در ضرب، هر دو رویداد با هم اتفاق می افتد. به هم این و هم آن معرف است. در این صورت دو حالت وجود دارد. یا رویدادها مستقلند یا وابسته:

– **ضرب رویدادهای مستقل:** به رویدادهایی مستقل گویند که وقوع یا عدم وقوع هریک در احتمال وقوع یا عدم وقوع حادثه‌ی دیگر تأثیر نداشته باشد. برای مثال تولد فرزندان و یا احتمال آمدن عدد ۵ در انداختن دو تاس سبز و آبی.

$$P(A\times B) = P(A)\times P(B) = \frac{۱}{٦}\times\frac{۱}{٦} = \frac{۱}{۳٦}$$

[1]. *Compound even*

دانلود
نمونه سوالات
استخدامی
ایران استخدام

– **ضرب رویدادهای وابسته:** اگر دو رویداد وابسته باشند، یعنی احتمال وقوع یکی از آنها به احتمال وقوع دیگری وابسته باشد، احتمال آن که هر دو اتفاق بیفتند عبارتست از:

$$P_{(A;B)} = P_A \times P_{B/A} = P_A \times \frac{P(A \cap B)}{A} = P(\frac{B}{A}) = \frac{P(A \cap B)}{P(A)}$$

مثال: در یک کیسه ۱۸ مهره وجود دارد که ٦ مهره مشکی و۱۲ مهره آن سفید است. احتمال اینکه اگر دو مهره خارج کنیم هر دو سفید باشد، چقدر است؟ (بدون جایگزینی)

$$P_A = \frac{۱۲}{۱۸} \Rightarrow P_B = \frac{۱۱}{۱۷} \rightarrow P = \frac{۱۲}{۱۸} \times \frac{۱۱}{۱۷} = \frac{۱۳۲}{۳۰٦}$$

فاکتوریل: n!

نشان‌دهنده حاصلضرب اعداد صحیح مثبت است، به طوری که هر عدد در عدد صحیح قبل از خود ضرب می شود. این عمل تا رسیدن به عدد یک ادامه پیدا می کند.

مثال: به چند روش می توان ٤ نوع کتاب را کنار هم چید؟ ٤!= ٤×۳×۲×۱=۲٤

نکته: ۰! = ۱

فرمول ترکیب[۱]: زمانی به کار می رود که ترتیب چیدن افراد یا اشیا برای ما مهم نیست. ترکیب abc با cba فرقی نمی کند.

$$C_r^n = \frac{n!}{r!(n-r)!}$$

نکته: باید ایتم ها متفاوت باشند. اگر مشابه باشند از این قاعده استفاده نمی شود.

ما باید r مورد را از n آیتم انتخاب کنیم (بدون جایگزینی)

مثال: ٦ نفر می خواهند دو به دو کنار هم بنشینند. اینکه چه افرادی کنار هم بنشینند، مهم نیست. تعداد حالت‌های ممکن برای نشستن را محاسبه کنید.

$$C_۲^٦ = \frac{٦!}{۲!(٤!)} = \frac{٦ \times ٥ \times ٤ \times ۳ \times ۲ \times ۱}{۲ \times ۱(٤ \times ۳ \times ۲ \times ۱)} = ٥$$

فرمول ترتیب: زمانی بکار می رود که ترتیب قرار گرفتن موارد برایمان مهم است.

$$A_r^n = \frac{n!}{(n-r)!}$$

مثال: ٦ نفر می خواهند دو به دو کنار هم بنشینند. این که چه افرادی کنار هم بنشینند، مهم است. تعداد حالت‌های ممکن برای نشستن را محاسبه کنید.

$$A_۲^٦ = \frac{٦!}{(٦-۲)!} = \frac{٦ \times ٥ \times ٤ \times ۳ \times ۲ \times ۱}{٤ \times ۳ \times ۲ \times ۱} = ۳۰$$

بسط دو جمله ای نیوتن (پاسکال)

$$[p+q]^n = p^n + c.p^{n-۱}.q + c.p^{n-۲}.q^{n+۱} + ...q^n$$

[۱]. *Combination*

دانلود
نمونه سوالات
استخدامی
ایران استخدامی

شرایط استفاده

۱- آزمایش‌ها فقط نشان دو نتیجه پیروزی و شکست باشد. $p+q=۱$

۲- آزمایش ها مستقل از هم باشند.

۳- تعداد دفعات کوشش مشخص باشد.

هنگام مواجه با مجموعه ای از پرسش های مشابه زیر از بسط دو جمله ای که راحت‌تر است استفاده کنید:

مثال: در امتحان ۳ سؤالی، ٤ جوابی، احتمال اینکه دانش آموزی بصورت تصادفی به یک سؤال پاسخ دهد، چقدر است؟

$$\frac{۱}{٤}, q=\frac{۳}{٤} \rightarrow (p+q)^۳ = p^۳ + ۳p^۲q^۱ + ۳p^۱q^۲ + q^٤ = ۱$$

- احتمال اینکه به همه‌ی سؤال ها از روی شانس پاسخ درست دهد برابر است با: $p^۳ = (\frac{۱}{٤})^۳ = \frac{۱}{٦٤}$

- احتمال اینکه به دو سؤال پاسخ درست دهد برابر است با: $۲p^۲q^۱ = ۳(\frac{۱}{٤})^۲(\frac{۳}{٤}) = \frac{۹}{٦٤}$

- احتمال اینکه به یک سؤال پاسخ درست دهد برابر است با: $۳p^۱q^۲ = ۳(\frac{۱}{٤})(\frac{۳}{٤})^۲ = \frac{۲۷}{۳٤}$

- احتمال اینکه اصلاً پاسخ درست ندهد برابر است با: $q^٤ = (\frac{۳}{٤})^۳ = \frac{۲۷}{٦٤}$

- احتمال اینکه حداقل به یک سؤال پاسخ درست دهد چقدر است؟

$$p^۳ + ۳p^۲q^۱ + ۳p^۱q^۲ = (\frac{۱}{٦٤}) + (\frac{۹}{٦٤}) + (\frac{۲۷}{٦٤}) = \frac{۳۷}{٦٤}$$

فرمول میانگین در توزیع دو جمله ای $\overline{X} = P.n$

n: تعداد دفعات تکرار آزمایش

P: احتمال پیروزی

مثال: اگر سکه ای ده بار پرتاب شود، میانگین آن چقدر است؟ $\overline{X} = \frac{۱}{۲} \times ۱۰ = ۵$

فرمول خطای استاندارد برآورد در توزیع دو جمله ای: $S_f = \sqrt{N.p.q}$

مثال: اگر در امتحان ۱۰ سؤالی شرکت کرده باشیم و امتحان ٤ گزینه ای باشد، خطای استاندارد برآورد آنرا محاسبه نمایید؟

$$S_p = \sqrt{۱۰ \times \frac{۱}{٤} \times \frac{۳}{٤}} = \sqrt{\frac{۳۰}{۱٦}} = ۱/۳۷$$

انواع تحلیل داده

تحلیل داده های کیفی: مستلزم سه فعالیت است:

۱- تلخیص داده ها[۱]: که منظور انتخاب، تمرکز، تنظیم و تبدیل داده‌ها است. رمزگذاری داده ها و اینکه پژوهشگر مشخص کند چه قسمتی از داده ها مفید و چه قسمتی نمی تواند مورد استفاده قرار گیرد در این بخش قرار دارد.

۱. *Data reduction*

دانلود نمونه سوالات استخدامی
ایران استخدامی

۲- عرضه داده ها[۱]: منظور ظاهرساختن مجموعه‌ای از داده های سازمان‌یافته است. مانند مطالب روزنامه، رایانه. شکل رایج عرضه داده های کیفی در گذشته استفاده از متن های داستان‌گونه بوده است.

۳- نتیجه گیری/ تأیید[۲]: پی بردن به معنای هر رویداد، الگوهای وقوع آنها، تبیین آنها. با نتیجه‌گیری نمی‌توان فعالیت پژوهشی را پایان یافته تلقی کرد. زیرا این فعالیت باید مورد تأیید واقع شود. یعنی باید از جنبه موجه‌بودن، استحکام و قابلیت تأیید مورد بازبینی قرار گیرد. این فرایند همان تعیین اعتبار نتایج است.

تحلیل داده های کمی: چنانچه داده ها کمی و پیوسته باشند، دارای توزیع بهنجار بوده و آزمودنی‌ها با نمونه‌گیری تصادفی انتخاب شده باشند، برای تحلیل داده ها می توان از آزمون های پارامتری استفاده کرد. این آزمون‌ها دارای مفروضه‌های نرمال‌بودن توزیع نمرات، یکسانی واریانس در گروه‌ها و داشتن مقیاس فاصله‌ای و نسبتی و پیوسته‌بودن داده‌ها است. مانند آزمون t. اگر این مفروضه ها وجود نداشته باشند، باید از آزمون های غیر پارامتری که با متغیرهای اسمی و رتبه ای سروکار دارد استفاده کرد. در مواردی که مقیاس فاصله ای است، ولی حجم نمونه کوچک است نیز از آزمون های ناپارامتری استفاده می شود.

نکته: در تحلیل‌های پارامتری می‌توان پارامترهای جامعه را برآورد کرد، درحالیکه در تحلیل‌های ناپارامتری فقط آزمون فرض صورت می گیرد (سرمد و دیگران، ۱۳۸۰: ۲۱۳).

---

۱. *Data display*

۲. *Conclusion drawing/ verification*

دانلود
نمونه سوالات
استخدامی
ایران استخدامی