جزوه آمار آزمون های استخدامی

{این جزوه توسط دانشجویان کارشناسی ارشد دانشگاه تهران تهیه شده است}

مقدمه

علم آمار تقریباً در سال ۱۹۰۰ با مقاله کارل پیرسون در مورد کای ـ اسکوئر متحول شد. پیش از آن در دوران کتله، تحلیل داده‌ها و آمارهای مربوط به سرشماری، به سؤالات ساده ولی مهمی از جمله تعیین نسبت زنان به مردان، نرخ مرگ و میر و ... پاسخ می‌داد.با شروع قرن بیستم، به‌منظور پاسخ‌گویی به همان سؤال ساده، تمرکز و توجه بیشتری روی آمارها در مقیاسی کوچک‌تر صورت گرفت. نظریه‌پردازان قرن بیستم سعی در پاسخ‌گویی به همان سؤالات ساده با استفاده از مجموعه داده‌های کوچک داشتند. تاکنون این‌گونه آمارها و نظریاتی که در مورد آن وجود دارند به تصمیم‌سازان و سیاست‌گذاران کمک کرده تا بتوانند با استفاده از آنها تصمیم‌های صحیح‌تری اتخاذ کنند.بدیهی است که آمار و اطلاعات دقیق، مهم‌ترین ابزار برای تصمیم‌سازان و سیاست‌گذاران برای برنامه‌ریزی دقیق و صحیح با کمترین هزینه است؛ ولی آیا این روال برای آینده هم مناسب است؟ با توجه به اینکه تغییرات با سرعت بیشتری به‌وقوع می‌پیوندد و تغییرات فناوری و در پی آن تغییر در دیگر جنبه‌های زندگی، افزایش روزافزون وابستگی متقابل کشورها و ملل، تمرکززدایی جوامع و نهادهای موجود که به‌دلیل گسترش فناوری اطلاعات شتاب بیشتری یافته است، تمایل روزافزون به جهانی‌شدن به همراه حفظ ویژگی‌های ملی، قومی و فرهنگی و بسیاری از عوامل دیگر، لزوم داشتن درکی بهتر از «تغییرات» و «آینده» را برای دولت‌ها، کسب‌وکارها، سازمان‌ها و مردم ایجاب می‌کند. آینده‌نگری یکی از ابزارهای بسیار مؤثر و مفید در راستای یاری به تصمیم‌گیران و سیاست‌گذاران است که در مقیاس قابل ملاحظه‌ای قابلیت استفاده از علم آمار را دارد

مقدمه‌ای بر آمار و احتمالات

سرفصل های این بخش

۱) تعریف آمار

۲) سیر تحول علم آمار

۳) جامعه و نمونه

۴) صفت ها و متغیر ها

۵) مقیاس های اندازه گیری

۶) عمل جمع

۷) پرسش و تمرین

TESTDOWNLOAD.IR

تعریف آمار :

آمار مجموعه ای از روشهارا برای جمع آوری و خلاصه کردن داده ها ، طبقه بندی آنها و روشهای تحلیلی برای پیش بینی ، برآورد و تصمیم گیری درشرایط مختلف ارائه می دهد .

سیر تحول آمار

سیر تحول آمار از نظر موضوعی به سه مرحله تقسیم می شود:

۱ - آمار توصیفی

قسمتی از علم آمار است که در باره خلاصه کردن و توصیف خصوصیات مهم مجموعه داده ها بحث می کند بدون آنکه استنباط آماری انجام شود

{در آمار توصیفی داده ها را در قالب نمودارها جداول فراوانی و شاخص های عددی خلاصه می کنیم. }

۲- آمار استنباطی

در این نوع آمار، محقق با استفاده از مقادیر نمونه آماره ها را محاسبه ، و سپس به کمک تخمین و آزمون فرض آماری، نتایج به دست آمده از آماره ها به پارامترها ی جامعه تعمیم میدهد.

{آمار استنباطی شامل روشهایی است که با استفاده از آنها اطلاعات موجود در نمونه به کل جامعه تعمیم داده می شوند .}

۳- آمار ناپارامتریک

این نوع آمار در مقابل آمار پارامتریک بیان می شود .

TESTDOWNLOAD.IR

فرض اساسی در آمار پارامتریک برخوردار بودن مشاهدات از یک توزیع خاص مانند توزیع نرمال است. در صورتی که در این نوع آمار این فرض ضرورتی ندارد .

{بیشتر در علوم رفتاری که متغیرهای آن با مقیاس های کیفی سنجیده می شوند از فنون آمار ناپارامتری استفاده می شود.}

جامعه آماری

هر مجموعه از افراد یا اشیاء و یا ... را که حداقل دارای یک خصوصیت مشترک باشند، جامعه می گویند مانند:

- اتومبیل های سواری شخصی با سن بالای ۱۵ سال
- مدیران با مدرک تحصیلی دکتری
- فاکتورهای فروش کالا در یک ماه معین

جمع آوری اطلاعات از همه اعضای جامعه را سرشماری می گویند.

جامع اماری را با N نشان می دهیم که ممکن است متنهاهی یا نا متناهی باشد

نمونه

نمونه تعداد محدودی از آحاد جامعهٔ آماری است که بیان کننده ویژگی های اصلی جامعه باشند.

تعداد نمونه ها را با n نمایش می دهیم و تکنیک های انتخاب نمونه را **نمونه گیری** می گوییم.

پارامتر

TESTDOWNLOAD

برای بدست آوردن برخی از شاخص ها در جامعه ، اگر این شاخصها را با اندازه گیری از تمامی عناصر جامعه بدست آوریم آنها را پارامتر می نامند .

آماره

اگر شاخص های مورد نظر در جامعه با استفاده از بخشی از جامعه (نمونه گیری) بدست آمده باشند آماره نامیده می شوند .

مثال

اگر بخواهیم میانگین درآمد کارکنان دولت را بدست آوریم :

در صورتی که این کار با استفاده از درآمد ***کلیه*** کارکنان دولت محاسبه گردد

پارامتر، و اگر با استفاده از درآمد ***بخشی*** از کارکنان بدست آید آماره نامیده می شود.

اندازه گیری

اندازه گیری عبارت است از نسبت دادن اعداد به خصوصیات اشیاء و وقایع یا افراد بر طبق قواعدی منطقی و قابل قبول .

به کمک تکنیک های مختلف، هر چیزی را می توان اندازه گیری کرد.

مقیاسهای اندازه گیری

۱- مقیاس اسمی

۲ - مقیاس ترتیبی

۳- مقیاس فاصله ای

TESTDOWNLOAD.IR

۴ – مقیاس نسبتی (نسبی)

خصوصیات انوع مقیاس	ترتیب	ارزش	مبدا صفر قرار دادي	مبدا صفر مطلق
اسمی	ندارد	ندارد	ندارد	ندارد
رتبه ای	دارد	ندارد	ندارد	ندارد
فاصله ای	دارد	دارد	دارد	ندارد
نسبتی	دارد	دارد	دارد	دارد

انواع صفت ها

۱– **صفت مشخصه:** صفت هایی که در همه اعضا جامعه به طور یکسان وجود دارند.

۲– **صفت متغیر:** صفت هایی هستند که از هر عضو به عضو دیگر تغییر می کنند. از این جهت آنها را به اختصار **متغیر** می گوییم. این متغیرها هستند که عملا در حین تحقیق مورد سوال و اندازه گیری قرار می گیرند.

داده ها مقادیر اندازه گیری شده یک متغیر هستند.

جامعه آماری و محدوده مطالعه با این صفتها تعیین می شود.

دسته بندی متغیرها کمک می کند مسیر مناسبی را برای رسیدن به نتایج درست، طی کنیم.

ASTDOWNLOAD

متغیرهای کیفی: بعضی از متغیرهایی که غیر قابل شمارش و اندازه گیری اند مثل رنگ چشم، گروه خون، درجه کیفیت کالا و مانند اینها را **متغیرهای کیفی** می گوییم. این گونه متغیرها ماهیت اندازه پذیری ندارند و تنها با حالات یا وضعیت هایی که دارند، ثبت می شوند. متغیرهای کیفی میتوانند **اسمی** یا **ترتیبی** باشند.

اگر به هریک از گروههای خونی چهارگانه اعداد ۱ تا ۴ را اختصاص دهید، با یک متغیر کیفی اسمی روبرو هستید و اگر میزان مهارت کارگران را با اعداد ۱ تا ۳ رتبه بندی کنید، یک متغیر کیفی ترتیبی خواهید داشت.

متغیرهای کمی

از آنجا که ریاضیات و آمار علم کمیت ها است، به همین دلیل دامنه وسیعی از تحلیل های آماری به متغیرهای کمی اختصاص دارند. متغیرهای کمی متغیرهایی هستند که قابل شمارش و اندازه گیری اند، و همیشه نتیجه اندازه گیری آنها، یک عدد است.

تعداد فرزندان یک خانواده، طول یک قطعه، دمای هوا و مانند اینها متغیرهای کمی هستند.

متغیرهایی را که فقط قابل شمارش هستند، **متغیرهای گسسته** می گویند.

مثل تعداد فرزند، تعداد دندان های پوسیده، تعداد تماس های نا موفق تلفن همراه و...

TESTDOWNLOAD.IR

متغیرهایی را که اندازه پذیرند و مقادیر آنها در فاصله ای از اعداد حقیقی قرار دارد، **متغیرهای پیوسته** می گویند.

مثل دمای هوا، مدت زمان تولید یک قطعه و وزن یک کالا و ...

عمل جمع در مطالعه آمار همه جا با عمل جمع داده‌ها سروکار داریم. برای اجتناب از زیاد نوشتن علامت (+)، نماد Σ (سیگما) را به عنوان اختصار برای عمل جمع به کار می‌بریم.

مجموعه داده‌ها متشکل از تعدادی اندازه است که به طور نمادی به صورت به ترتیب نشان داده می‌شوند. مجموع این n عدد را به این صورت می نویسیم:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

$$\begin{aligned}\sum_{i=1}^{5} x_i &= x_1 + x_2 + x_3 + x_4 + x_5 \\ &= 3 + 5/3 + 4 + 6 + 3/7 \\ &= 22\end{aligned}$$

۱. در آمار توصیفی و استنباطی چه اهدافی را دنبال می کنیم، هر کدام را با یک مثال بیان کنید.

۲. تفاوت پارامتر و آماره چیست؟

۳. تفاوت داده با متغیر را بایک مثال شرح دهید.

۴. متغیرهای کمی چه تفاوتی با متغیرهای کیفی دارند؟

۵. مقیاس‌های اندازه گیری را با مثال توضیح دهید.

۶. مقیاس هر یک از متغیرهای زیر را تعیین کنید.

TESTDOWNLOAD.IR

- میزان علاقمندی به درس؛
- درآمد خانواده؛
- مارک اتومبیل؛
- موسیقی مورد علاقه؛
- تعداد کتابهایی که در یک ماه گذشته مطالعه کرده‌اید؛
- مدت تماشای تلویزیون در روز؛
- تعداد اتاق های منزل؛
- شغل؛

توصیف اطلاعات

سرفصل های این بخش

۱) انواع طبقه بندی

۲) طبقه بندی داده های کیفی و گسسته

۳) فراوانی ها

۴) طبقه بندی داده های پیوسته

۵) نمودارهای آماری

اندازه‌های ثبت شده در مجموعه داده‌ها، اجزای اساسی اطلاعاتی هستند که باید به دقت آنها را مورد بررسی قرار داده و به خوبی توصیف کنیم. به همین دلیل در آمار توصیفی اهداف زیر را دنبال می کنیم.

- خلاصه‌کردن، دسته‌بندی کردن و تهیه جدول‌های توزیع فراوانی
- بدست آوردن شاخصهایی که گرایش به مرکز داده‌ها را معین می کنند (شاخصهای مرکزی)
- بدست آوردن شاخصهایی که میزان انحراف از مرکز داده‌ها را معلوم می کنند (شاخصهای پراکنگی)
- رسم نمودارهای مناسب برای توصیف بصری داده ها
- بررسی کلی نمودارها و پارامترها از لحاظ خصوصیات گرایش به مرکز یا انحراف از مرکز

اندازه‌های ثبت شده در مجموعه داده‌ها، اجزای اساسی اطلاعاتی هستند که در دسترس محقق قرار دارد. بنابراین در مواجهه با تعداد زیادی داده، ذهن انسان نمی‌تواند محتوای کلی اطلاعات ثبت شده در مجموعه داده‌ها را بلافاصله درک کند. یکی از راههای توصیف اطلاعات دسته بندی کردن داده ها است.

TESTDOWNLOAD.IR

در نگاه اول ممکن است داده‌ها اطلاعات مفیدی به ما ندهند و برای اینکه بدانیم این اعداد شامل چه اطلاعاتی هستند، باید آنها را مرتب و دسته‌بندی کنیم تا قابل تفسیر و بهره برداری شوند. طبقه بندی کردن داده های گسسته و کیفی به دلیل محدود بودن پاسخها بسیار ساده است. در صورتی که طبقه بندی متغیرهای پیوسته همواره با پیچیدگی های مختلفی روبرو است که در ادامه به آنها نیز خواهیم پرداخت.

برای دسته بندی داده های کیفی کافی است فراوانی هر یک از سطوح متغیر کیفی را در یک جدول قرار دهیم

برای دسته بندی داده های کمی گسسته باید هریک از مقادیر متغیر را به همراه فراوانی آنها در یک جدول قرار دهیم.

تعداد فرزندان ۵۰ نفر از کارمندان یک موسسه را پرسیده ایم و نتایج به دست آمده شامل ۵۰ عدد است که پس از دسته بندی به صورت زیر خواهند شد.

تعداد دندانها	فراوانی fi
0	2
1	13
2	18
3	10
4	7

در مدیریت ، انسان ها را به لحاظ ارتباط به ۴ دسته تصویری، احساسی، صوتی و ارقامی تقسیم می کنند. اطلاعات جدول زیر نتایج به دست آمده از ۱۰۰ نفر است.

نوع ارتباط	فراواني fi
تصويري	35
احساسي	20
صوتي	32
ارقامي	13

به چنین جدولهایی، جدول، توزیع فراوانی میگوییم و میتوان در آنها فراوانی های دیگری را نیز تعیین کرد.

TESTDOWNLOAD

فراوانی مطلق

به تعداد داده ها در هر طبقه فراوانی مطلق آن طبقه می گویند و آن را با fi نشان می دهند.

فراوانی نسبی

اگر فراوانی های مطلق را بر کل فراوانی ها تقسیم کنیم، فراوانی نسبی (ri) به دست می آید.

$$r_i = \frac{f_i}{n}$$

فراوانی تجمعی

به مجموع فراوانی های مطلق طبقه های قبل و همان طبقه، فراوانی تجمعی آن طبقه می گویند و آن را با Fi نمایش میدهند.

$$F_i = \sum_{j=1}^{i} f_j$$

فراوانی تجمعی نسبی

می توان از تقسیم فراوانی های تجمعی بر تعداد داده ها، این فراوانی را به دست آورد.

$$R_i = \frac{F_i}{n}$$

نماینده طبقه

ضرورت دارد برای هر طبقه یک نماینده تعیین کنیم. نقطه وسط هر طبقه را نماینده آن طبقه می گوییم و آن را با x'i نشان می دهیم.

$$x'_i = \frac{L_i + H_i}{۲}$$

مثال

جدول توزیع فراوانی همراه با فراوانیهای مطلق، تجمعی، نسبی، تجمعی نسبی برای تعداد فرزندان ۵۰ کارمند در مثال قبل

+13+18
2

۱۳/۵۰ =۰.۲۶

۲/۵۰ =۰.۰٤

٤۳/۵۰ =۰.۸٦

x_i	f_i	F_i	r_i	R_i
0	2	2	0/0	0/04
1	1	1	4	0/3
2	3	5	0/2	0/66
3	1	3	6	0/86
4	8	3	0/3	1
	1	4	6	
	0	3	0/2	
	7	5	0/1	
		0	4	

$$\sum_{i=1}^{k} r_i = 1 \qquad \sum_{i=1}^{k} f_i = n, \qquad R_i = \sum_{j=1}^{i} r_j$$

به خاطر داشته باشید که

4- حدود هر طبقه (Li-Hi) را تعیین کنید وداده هایی که در هر طبقه قرار می گیرند، شمارش کرده و به عنوان فراوانی (fi) آن طبقه ثبت کنید.

مثال

معدل ۵۰ دانشجو با تقریب تا یک رقم اعشار، به شرح زیر است. این داده ها را به ۸ طبقه تقسیم بندی کنید و فراوانی های مطلق نسبی تجمعی و تجمعی نسبی را به دست آورید.

2/1	1/9	1/6	2/2	2/1	2/2	2/4	1/8	1/5	2/9
1/8	2/3	1/8	1/7	2/3	2/3	2/0	2/5	2/1	2/6
1/8	1/2	1/9	1/7	1/7	2/0	1/9	2/2	2/6	1/4
2/9	2/4	1/8	1/9	2/2	2/2	2/5	2/0	2/0	2/0
1/4	2/5	1/9	1/8	1/6	2/4	2/9	1/9	1/6	1/4

برای طبقه بندی این داده ها، مراحل دسته بندی کردن داده های کمی پیوسته را پی می گیریم

[بزرگترین و کوچکترین عدد در بین داده ها را مشخص کرده ایم]

4- حدود هر طبقه (Li-Hi) را تعیین کنید وداده هایی که در هر طبقه قرار می گیرند، شمارش کرده و به عنوان فراوانی (fi) آن طبقه ثبت کنید.

مثال

معدل ۵۰ دانشجو با تقریب تا یک رقم اعشار، به شرح زیر است. این داده ها را به ۸ طبقه تقسیم بندی کنید و فراوانی های مطلق نسبی تجمعی و تجمعی نسبی را به دست آورید.

2/1	1/9	1/6	2/2	2/1	2/2	2/4	1/8	1/5	2/9
1/8	2/3	1/8	1/7	2/3	2/3	2/0	2/5	2/1	2/6
1/8	1/2	1/9	1/7	1/7	2/0	1/9	2/2	2/6	1/4
2/9	2/4	1/8	1/9	2/2	2/2	2/5	2/0	2/0	2/0
1/4	2/5	1/9	1/8	1/6	2/4	2/9	1/9	1/6	1/4

برای طبقه بندی این داده ها، مراحل دسته بندی کردن داده های کمی پیوسته را پی می گیریم

[بزرگترین و کوچکترین عدد در بین داده ها را مشخص کرده ایم]

معمولاً برای توصیف بهتر داده‌های آماری از نمودارها استفاده می‌کنیم. عموما این نمودارها در ارتباط با داده‌های پیوسته به کار گرفته می شوند.

منظور از نمایش نموداری داده ها، تجسم عینی اطلاعات نهفته در آنها ست.

تعدادی از نمودارهای مهم آماری

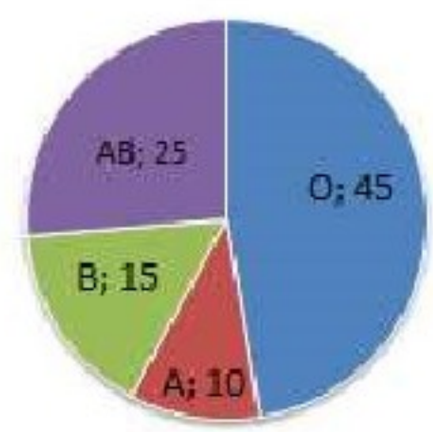
۱- نمودار دایره ای (معمولا برای داده های کیفی استفاده می شود)
۲- نمودار ستونی (معمولا برای داده های کمی گسسته استفاده می شود)
۳- نمودار مستطیلی یا هیستوگرام (کمی پیوسته)
٤- نمودار چند ضلعی فراوانی (کمی پیوسته)
٥- نمودار ساقه و برگ (کمی گسسته و پیوسته)
٦- نمودار جعبه ای (کمی پیوسته)

مثال گروه خون ۱۰۰ بیمار بستری در یک بیمارستان را در اختیار داریم نمودار دایره ای داده ها را رسم کنید.

نحوه رسم نمودار دایره ای را در این مثال ببینید.

۰/۱۵×۳٦۰=٥٤

گروه خون	فراوانی fi	فراوانی نسبی ri	قطاع دایره (درجه)
O	45	0/45	162
A	10	0/1	36
B	15	0/15	54
AB	25	0/25	90



روش رسم نمودار ستونی

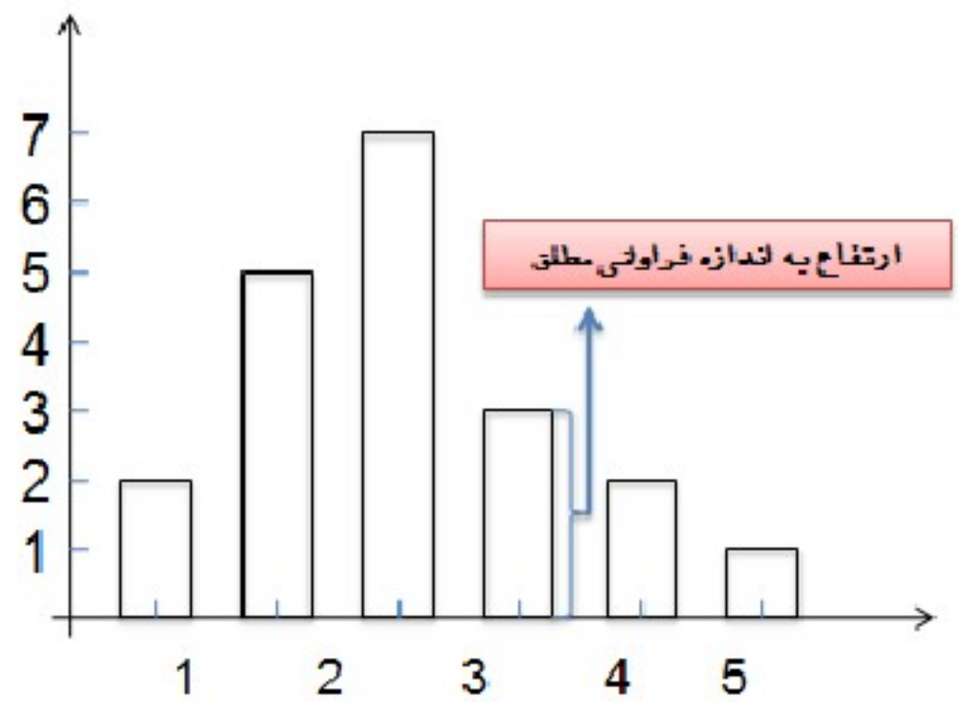
برای رسم نمودار ستونی کافی است مقادیر متغیر را روی محور افقی تعیین کنید سپس مستطیل هایی با فاصله و به ارتفاع فراوانی مطلق برای هر مقدار متغیر رسم کنید.

مثال برای داده های زیر نمودار ستونی رسم کنید.

2 2 0 1 2 3 2 1 1 5 2 0 1 2 2 4 4 3

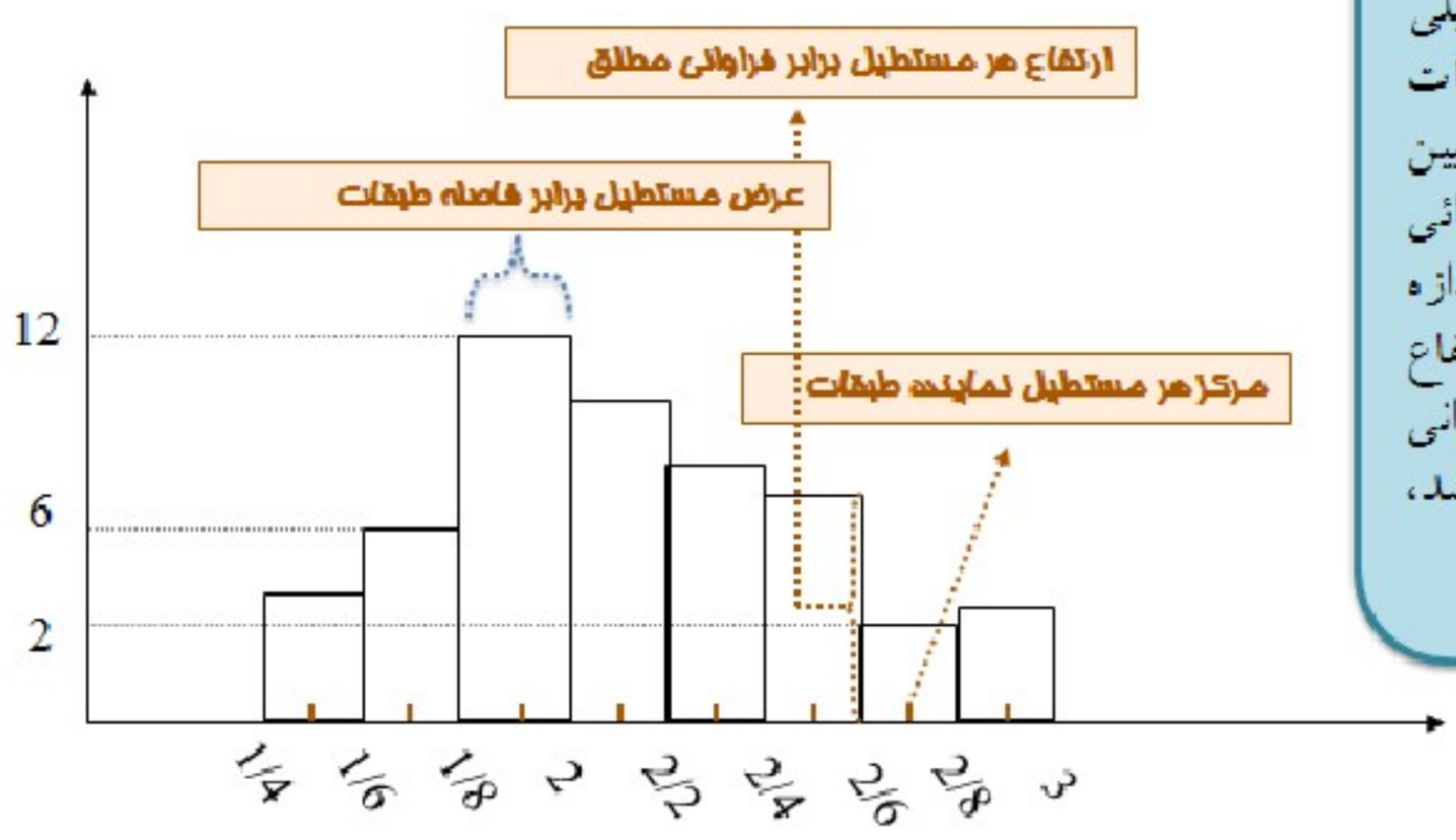
x	fi
0	2
1	5
2	7
3	3
4	2
5	1

ابتدا جدول توزیع فراوانی را تشکیل می دهیم



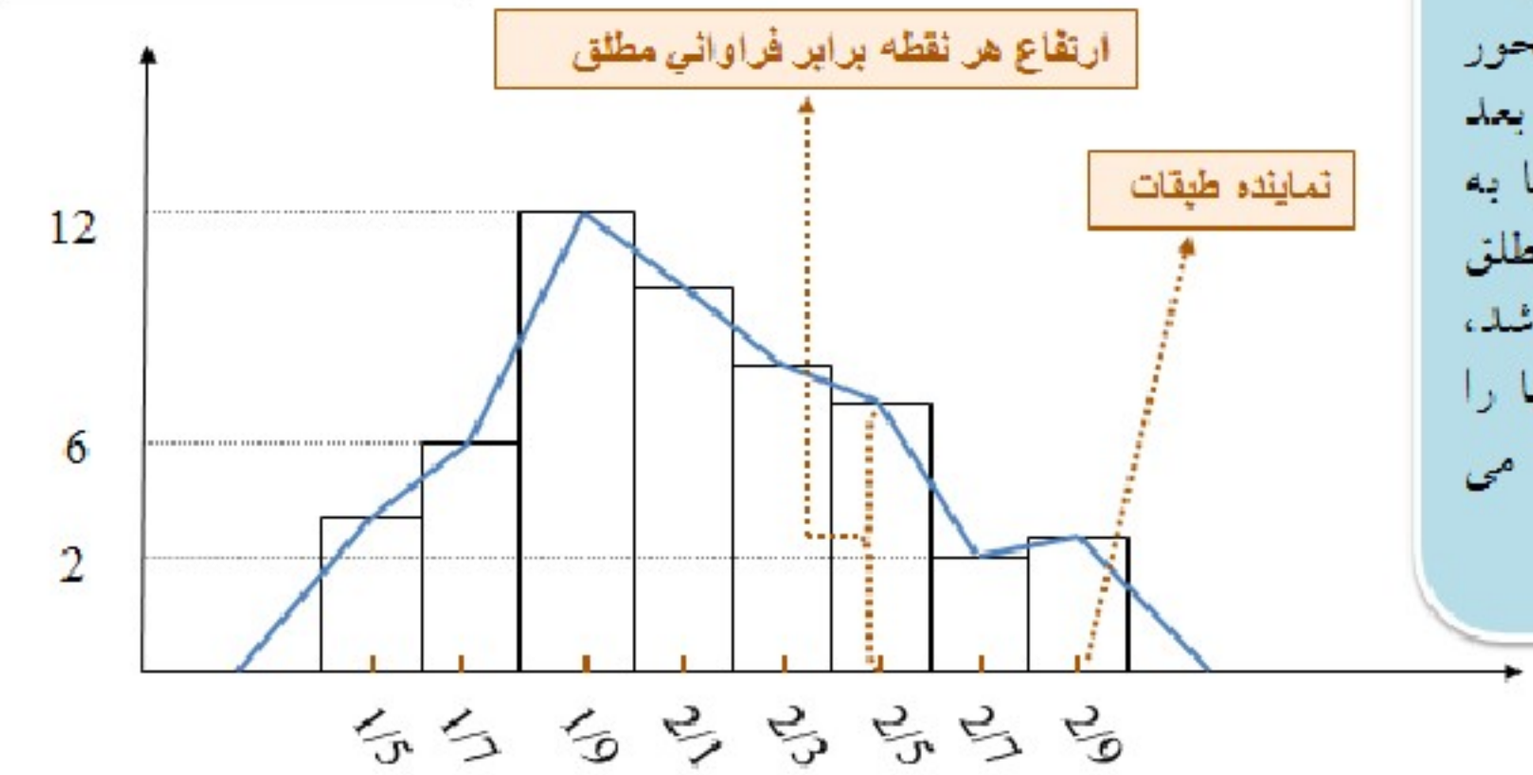
روش رسم نمودار مستطیلی

برای رسم هیستوگرام فراوانی یا نمودار مستطیلی داده ها ابتدا حدود طبقات را روی محور افقی تعیین می‌کنیم. بعد مستطیل هائی که عرض آنها به اندازه فاصله طبقات و ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد، رسم می‌کنیم.



روش رسم نمودار چند ضلعی

مستطیلها را روی این نمودار برای درک بهتر نمودار چند ضلعی قرار داده ایم



برای رسم چند ضلعی فراوانی داده ها نماینده طبقات را روی محور افقی تعیین می‌کنیم. بعد نقاطی که ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد، مشخص کرده و آنها را به یکدیگر متصل می کنیم.

روش رسم نمودار چند ضلعی فراوانی تجمعی

	F_i
1/4 - 1/6	4
1/6 - 1/8	10
1/8 - 2	22
2 - 2/2	31
2/2 - 2/4	39
2/4 - 2/6	45
2/6 - 2/8	47
2/8 - 3	50

برای رسم چند ضلعی فراوانی تجمعی داده ها نماینده طبقات را روی محور افقی تعیین می‌کنیم. بعد نقاطی که ارتفاع آنها به اندازه فراوانی تجمعی طبقه مربوطه باشد، مشخص کرده و آنها را به یکدیگر متصل می کنیم.

شاخصهای مرکزی

گام اول در توصیف داده ها، طبقه بندی آنها است که در قسمت دوم این مجموعه به آن پرداخته شد. اینک برای اینکه داده‌ها را به صورت بهتری توصیف کنیم، باید گام دیگری برداریم و آن اینکه مشخص کننده‌های عددی را برای داده ها بدست آوریم. به این مشخص‌کننده‌های عددی شاخص یا معیار می‌گویند و بر دو نوعند:

شاخصهای مرکزی شاخصهایی هستند که میزان گرایش به مرکز داده ها را اندازه می گیرند.

شاخصهای پراکندگی شاخصهایی هستند که میزان پراکندگی داده ها از مرکز را اندازه می گیرند

توجه: بسیاری از شاخص های آماری برای داده های کیفی تعریف نمشوند

شاخصهای مرکزی

- میانگین
- میانه
- نما
- چندک ها

شاخصهای پراکندگی

- واریانس
- انحراف معیار
- دامنه تغییرات
- ضریب تغییرات
- ضریب چولگی

شاخصهای مرکزی

میانگین حسابی

میانگین حسابی

یکی از مهمترین شاخص های گرایش به مرکز، میانگین حسابی است. میانگین حسابی را از این رابطه به دست می آوریم. چنانچه میانگین حسابی را از یک نمونه به دست آوریم آن را با $\overline{X}$ و اگر میانگین مقادیر مربوط به همه جمعیت را محاسبه کنیم، آن را با μ (مو) نمایش می‌دهیم.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

مثال: فرض کنید داده‌های زیر، نمرات ۱۰ درس مربوط به یک دانش‌آموز باشد. می خواهیم میانگین نمرات او را حساب کنیم، داریم:

۱۰ ۱۶ ۱۲ ۱۴ ۱۳ ۱۵ ۱۴ ۹ ۱۶ ۱۱

$$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10}(11+10+16+12+14+13+15+14+9+16) = \frac{130}{10} = 13$$

توجه داشته باشید:

اگر داده‌ها طبقه بندی شده باشند یعنی مقادیر $x_1, x_2, \ldots x_k$ را با فراوانی های $f_1, f_2, \ldots, f_k$ در اختیار داشته‌باشیم، در این صورت فرمول میانگین به شکل زیر خواهد بود:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{1}{n}(f_1 x_1 + f_2 x_2 + \ldots f_k x_k)$$

شاخصهای مرکزی

میانگین حسابی

مثال: میانگین حسابی را در داده‌های جدول زیر محاسبه می کنیم

x_i	f_i	$f_i x_i$
5	3	3×5=15
7	3	3×7=21
8	2	2×8=16
9	5	5×9=45
10	5	5×10=50
	18	$\sum f_i x_i = ۱٤۷$

$$\bar{x} = \frac{۱}{۱۸}\sum_{i=۱}^{۵} f_i x_i = \frac{۱}{۱۸}(۱٤۷) = ۸/۱۷$$

توجه داشته باشید

در داده های پیوسته که طبقه بندی شده باشند، در فرمول بالا باید به جای x_i از x'_i (نماینده طبقات) استفاده کنیم.

مثال: برای داده های جدول توزیع فراوانی زیر میانگین حسابی را به دست می آوریم.

حدود طبقات	f_i	x'_i	$f_i x'_i$
1-3	1	2	2
4-6	5	5	25
7-9	15	8	120
-12	6	11	66
10	3	14	42
-15			255
13			30

$$\bar{x} = \frac{۱}{۳۰}\sum_{i=۱}^{۵} f_i x'_i = \frac{۱}{۳۰}(۲۵۵) = ۸/۵$$

شاخصهای مرکزی

میانگین حسابی

خواص میانگین حسابي

1- همیشه مجموع اختلاف داده‌ها از میانگین حسابی صفر است.

$$\sum_{i=1}^{n}(x_i-\bar{x})=0$$

2-هرگاه به هر يك از داده‌ها عد ثابتي را مانند a اضافه يا كم كنيم، میانگین داده‌های حاصل، برابر با میانگین داده‌های قبلی به اضافه (یا منهای) عدد ثابت a خواهد بود.

$$Y_i = X_i \pm a \rightarrow \bar{Y} = \bar{X} \pm a$$

3- هرگاه عدد ثابت a را در هريك از داده‌ها ضرب كنيم، میانگین داده‌های حاصل، برابر میانگین داده‌های قبلی ضرب در عدد ثابت a خواهد بود.

$$Y_i = aX_i \Rightarrow \bar{Y} = a\bar{X}$$

4- درمورد تقسیم هم به صورت ضرب عمل می کنیم.

$$Y_i = \frac{X_i}{a} \rightarrow \bar{Y} = \frac{\bar{X}}{a}$$

شاخصهای مرکزی

میانگین کل

میانگین کل

اگر یك مجموعه n_1 تائی از داده‌ها با میانگین $\bar{X}_1$ و یك مجموعه n_2 تائی از داده‌ها با میانگین $\bar{X}_2$ داشته باشیم میانگین کل این دو مجموعه داده برابر است با :

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

این موضوع برای n مجموعه نیز صادق است

مثال: در یک مجموعه، ۱۰ داده با میانگین $\bar{x}_1 = ٤/٢$ و در مجموعه دیگری ۸ داده با میانگین $x_2 = ٤$ در اختیاراست. میانگین کل داده‌ها را به دست می آوریم.

$$\bar{x} = \frac{٤/٢ \times ١٠ + ٤ \times ٨}{١٠ + ٨} = \frac{٦٤}{١٨} = ٣/٥٥$$

کاربرد این میانگین بیشتر برای محاسبه حد متوسط شاخص‌ها، نسبت ها و درصدها است.

میانگین هندسی

فرض کنید $X_1, X_2, ..., X_n$ داده‌های مثبت و مخالف صفر باشند، میانگین هندسی (Geometric Mean) آنها از فرمول مقابل بدست می‌آید و آن را با G نمایش می‌دهیم:

$$G = \sqrt[n]{\prod_{i=1}^{n} X_i} = \sqrt[n]{X_1 \times X_2 \times ... \times X_n}$$

شاخصهای مرکزی

میانگین هندسی

مثال: جدول زیر سرمایه در هر سال در یک فروشگاه را نمایش می‌دهد با توجه به مقادیر جدول، نرخ رشد سرمایه (میانگین نسبت رشد) را محاسبه می کنیم.

سال	سرمایه	نسبت رشد
1373	15	-
1374	20	$\frac{۲۰}{۱۵}=\frac{۴}{۳}$
1375	30	$\frac{۳۰}{۲۰}=\frac{۳}{۲}$
1376	30	$\frac{۳۰}{۳۰}=۱$
1377	35	$\frac{۳۵}{۳۰}=\frac{۷}{۶}$
1378	40	$\frac{۴۰}{۳۵}=\frac{۸}{۷}$

$$G = \sqrt[۵]{\frac{۴}{۳} \times \frac{۳}{۴} \times \frac{۷}{۶} \times \frac{۸}{۷} = \frac{۸}{۳}} \quad \sqrt[۵]{\ }$$

روش کوتاهتر این است که حجم تولید در سال آخر یعنی 40 را برحجم تولید در سال اول یعنی 15 تقسیم کنیم و سپس ریشه (n - 1) ام را به دست آوریم.

برای سهولت در محاسبه میانگین هندسی می‌توان از خواص لگاریتم استفاده کرد. بدین ترتیب که اگر از طرفین فرمول آن، لگاریتم بگیریم می توان از فرمول زیر که مشابه میانگین حسابی است، استفاده کرد.

$$\log G = \frac{۱}{n}\sum_{i=۱}^{n} \log x_i$$

شاخصهای مرکزی

میانگین همساز

میانگین همساز

فرض کنید داده‌های x_1 x_2 ,... x_n مخالف صفر باشند، میانگین هارمونیک آنها از فرمول زیر بدست می‌آید و آن را با H نمایش می‌دهیم:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

این میانگین بیشتر در مطالعه شبکه های برق، اپتیک و برخی نسبت ها استفاده می شود.

مثال: فرض کنید یک اتومبیل مسیر ۶۰ کیلومتری را با سرعت ۴۰ کیلومتر طی کرده و همین مسیر را با سرعت ۶۰ کیلومتر برگشته است. متوسط سرعت در طی این ۱۲۰ کیلومتر چقدر است؟

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{2}{\frac{1}{40} + \frac{1}{60}} = 48$$

توجه داشته باشید

اگر از میانگین حسابی متوسط سرعت را حساب کنیم $\bar{x} = \frac{1}{2}(40 + 60) = 50$ را به دست خواهیم آورد که عدد درستی نیست. زیرا زمان رفتن این اتومبیل یک ساعت و زمان برگشتن او یک ساعت ونیم طول کشیده است. پس او مسیر ۱۲۰ کیلومتری را در مدت دو ساعت و نیم طی کرده که به طور متوسط سرعتی معادل ۴۸ کیلومتر داشته است.

میانه

میانه

میانه

میانه؛ عددی است از بین داده‌ها که اگر داده‌ها را (به طور غیر نزولی) مرتب کرده باشیم، نیمی از داده‌ها کمتر از آن واقع شوند. به عبارت دیگر داده‌ای از میان داده‌ها که دقیقاً در وسط قرار دارد.

توجه داشته باشید

اگر تعداد داده‌ها فرد باشد، داده ای که در وسط قرار می گیرد، میانه است. ولی اگر تعداد داده‌ها زوج باشد، میانگین دو عدد وسط، میانه خواهد بود.

مثال: میانه مجموعه داده‌های زیر را بدست می آوریم. ٦ ٣ ٤ ١ ٢ ١ ٣ ٢ ١ ٧ ٠ ١ ٥ ٣ ٠ ٣

ابتدا داده‌ها را مرتب می کنیم.

٠ ٠ ١ ١ ١ ٢ ٢ ٣ ٣ ٣ ٣ ٤ ٥ ٦ ٧

چون تعداد داده ها ١٥و عددی فرد است، داده هشتم میانه است زیرا:

$$١٥ \times \frac{١}{٢} = ٧/٥ \cong ٨ \rightarrow Md = ٣$$

شاخص های مرکزی

میانه

مثال: برای داده‌های زیر میانه را محاسبه می کنیم.

۱ ۳ ۷ ۵ ۹ ۱ ۸ ٦ ٤ ۷ ۱ ۱ ۹ ۹ ۲ ۲

ابتدا داده‌ها را مرتب می‌کنیم.

۱ ۱ ۱ ۱ ۲ ۲ ۳ ٤ ٥ ٦ ۷ ۷ ۸ ۹ ۹ ۹

چون ١٦= n عددی زوج است، پس معدل داده هشتم و نهم را به عنوان میانه در نظر می گیریم

$$Md = \frac{٤ + ٥}{٢} = ٤/٥$$

۱ ۱ ۱ ۱ ۲ ۲ ۳ ٤ ٥ ٦ ۷ ۷ ۸ ۹ ۹ ۹

توجه کنید که:

در دو مثال اخیر داده ها گسسته و دسته بندی نشده اند

شاخص های مرکزی

میانه

اگر داده ها گسسته و دسته بندی هستند مراحل زیر را دنبال کنید

۱- ستون فراوانی تجمعی را بدست می آورید.

۲- مقدار $\frac{۱}{۲} \times n$ را بدست آورده و از ستون فراوانی تجمعی، طبقه میانه‌دار را معلوم کنید.

۳- مقدار متغیر آن طبقه، میانه داده ها است.

مثال: برای داده های جدول زیر میانه را به دست آورید.

x_i	f_i	F_i	
5	3	3	
7	4	7	
8	5	12	طبقه میانه دار
9	18	30	
10	10	40	
	40		

فراوانی تجمعی

چون $۲۰ = \frac{۱}{۲} \times ٤۰$ است پس داده بیستم میانه است.

داده بیستم در طبقه سوم قرارا دارد.

$md = ۸$

شاخص های مرکزی

میانه

اگر داده ها پیوسته و دسته بندی شده هستند مراحل زیر را دنبال کنید

۱- ستون فراوانی تجمعی را بدست می آورید.

۲- مقدار $\frac{۱}{۲} \times n$ را بدست آورده و از ستون فراوانی تجمعی، طبقه میانه‌دار را معلوم کنید.

۳- میانه را از فرمول زیر به دست آورید.

$$Md = L_i + \frac{\frac{n}{۲} - F_{i-۱}}{f_i} \times C$$

در این فرمول داریم:

$L_i =$ کران پائین طبقه‌ای است که میانه در آن قرار دارد.

$n =$ تعداد داده‌ها

$F_{i-۱} =$ فراوانی تجمعی یک طبقه، قبل از طبقه‌ای که شامل میانه است.

$f_i =$ فراوانی مطلق طبقه‌ای است که میانه در آن قرار دارد.

$c =$ فاصله طبقات

شاخص های مرکزی

میانه

مثال: میانه را برای داده های جدول زیر به دست آورید.

حدود طبقات	f_i	F_i
1 – 4	8	8
4 – 7	5	13
7 – 10	13	26
10 – 13	15	41
13 - 16	9	50

فراواني تجمعي

طبقه میانه دار

ابتدا ستون فراوانی تجمعی را تشکیل می‌دهیم سپس داریم

$$۵۰ \times \frac{۱}{۲} = ۲۵$$

بنا بر این بیست و پنجمین عدد در طبقه سوم قرار دارد حال باید از فرمول استفاده کنیم داریم:

$$Md = ۷ + \frac{۲۵ - ۱۳}{۱۳} \times ۳ = ۷ + ۲/۷۷ \Rightarrow Md = ۹/۷۷$$

فاصله طبقات

شاخص های مرکزی

نما

نما

نما یا مد (Mode) داده‌ای است که بیشتر از سایر داده‌ها تکرار شده باشد و آن را با نماد Mo نمایش می‌دهیم.

مثال: نما در داده‌های زیر عدد ۳ است. زیرا بیشتر از همه، عدد ۳ تکرار شده است،

3 1 5 9 2 3 1 7 3 2 3 4 3 3 2

اگر داده‌ها گسسته و طبقه بندی شده باشند مانند مثال زیر عمل کنید

مثال: نما در داده‌های جدول زیر، عدد ۱۵ است. چون فراوانی عدد ۱۵ بیشتر است.

x_i	f_i
10	5
15	19
20	13
25	8

همچنان به زوج یا فرد بودن تعداد داده ها توجه داشته باشید

شاخص های مرکزی

نما

نما در داده‌های طبقه بندی شده پیوسته

اگر داده ها پیوسته و طبقه‌بندی شده هستند، به صورتی که در زیر توضیح داده می‌شود، نما را محاسبه کنید.
ابتدا طبقه‌ای را که فراوانی آن ماکزیمم است معلوم کنید (طبقه نمادار)
سپس از فرمول مقابل برای محاسبه نما استفاده کنید .

$$Mo = l_i + \frac{D_1}{D_1 + D_2} \times C$$

که در این فرمول:

l_i = کران پائین طبقه‌ای است که فراونی مطلق آن ماکزیمم است.(طبقه نمادار)

D_1 = تفاضل فراوانی مطلق طبقه نمادار با طبقه قبل از آن است، یعنی $D_1 = f_i - f_{i-1}$

D_2 = تفاضل فراوانی مطلق طبقه نمادار با طبقه بعد از آن است. یعنی $D_2 = f_i - f_{i+1}$

C = فاصله طبقات

شاخص های مرکزی

نما

مثال: برای داده‌های جدول فراوانی زیر نما را محاسبه می کنیم.

چون طبقه چهارم بیشترین فراوانی را دارد، پس طبقه نمادار است. حال از فرمول استفاده می کنیم داریم:

حدود طبقات	f_i
1 – 4	8
4 – 7	5
7 – 10	13
10 – 13	15
13 - 16	9

$D_۱ = ۱۵ - ۱۳ = ۲$

$D_۲ = ۱۵ - ۹ = ۶$

فاصله طبقات

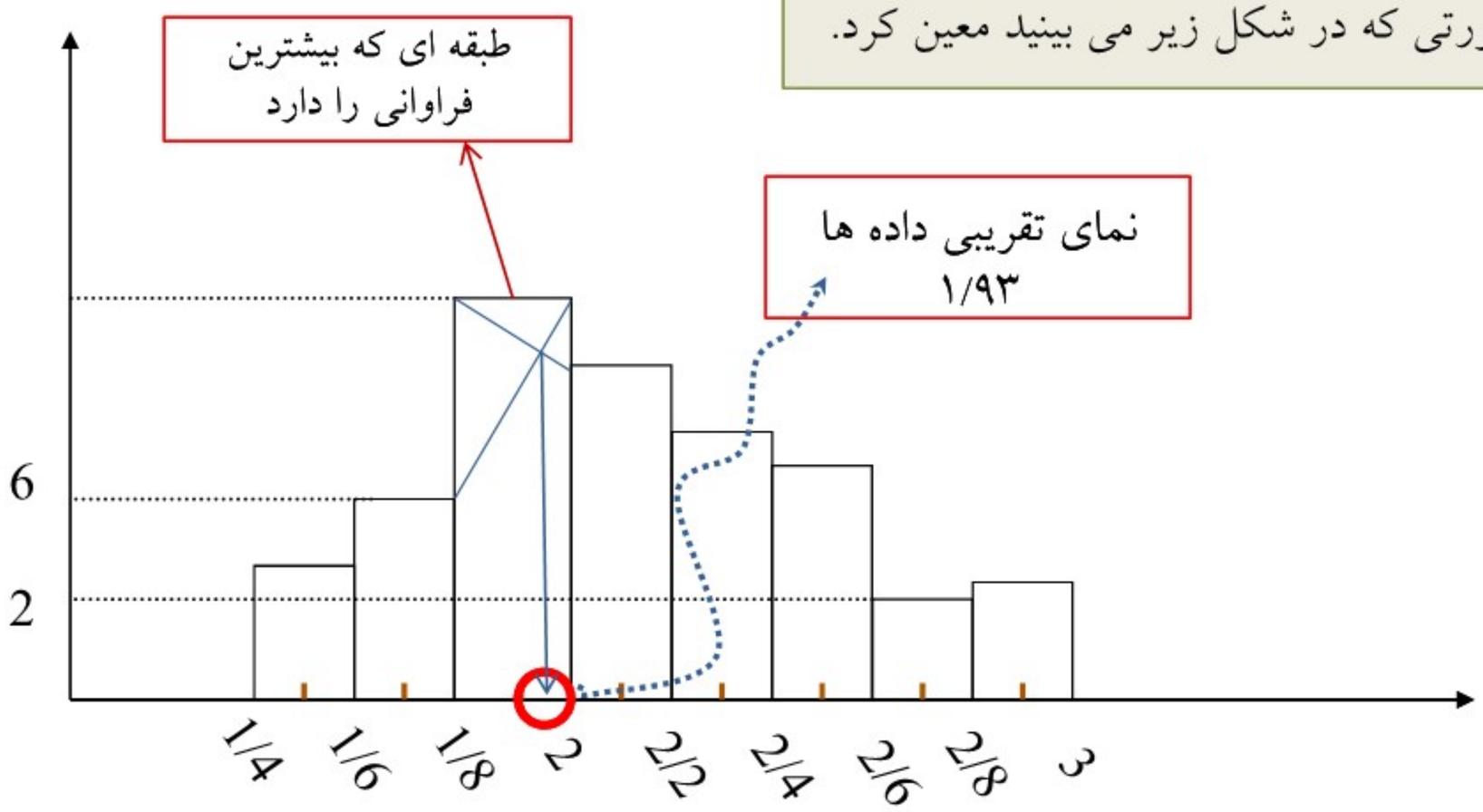
$$Mo = ۱۰ + \frac{۲}{۲+۶} \times ۳ = ۱۰ + ۰/۷۵ \Rightarrow Mo = ۱۰/۷۵$$

شاخص های مرکزی

نما

نما روی نمودار مستطیلی

می توان نمای داده ها را با استفاده از نمودار مستطیلی و به طور تقریبی و به صورتی که در شکل زیر می بینید معین کرد.



شاخص های مرکزی

چندک ها

صدک ها

صدک (p ۱۰۰)ام داده‌ای است از مجموعه داده‌ها که اگر آنها را به ترتیب صعودی مرتب کرده باشیم، (p ۱۰۰) درصد داده‌ها از آن کوچکتر و [(p−۱) ۱۰۰] درصد داده‌ها از آن بزرگتر باشند.

$۰ < P < ۱$

مثلاً $p_{۳۷}$ ، صدک ۳۷ ام است و داده‌ای است که ۳۷ درصد داده‌ها از آن کوچکتر است. بدیهی است که ٦٣ درصد داده‌ها از آن بزرگتر باشد.

دهک ها

دهک ها، حالت خاص صدک ها هستند و به ترتیب به صدک های دهم، بیستم، سی ام و... و نودام، به ترتیب دهک اول، دوم، سوم و ... نهم گفته می‌شود.

مثلاً $p_{٤٠}$ ، صدک چهلم یا دهک چهارم است و داده‌ای است که ٤٠ درصد داده‌ها از آن کوچکتر است. بدیهی است که ٦٠ درصد داده‌ها از آن بزرگتر باشد.

شاخص های مرکزی

چندک ها

چارک ها

برای داده ها سه چارک تعریف تعریف می کنیم. چارک اول Q_1 و چارک دوم Q_2 و چارک سوم Q_3 که به ترتیب باصدک های ۲۵ و ۵۰ و ۷۵ برابر اند.

مثلاً Q_1 ، چارک اول است و داده‌ای است که یک چهارم (۲۵٪) داده‌ها از آن کوچکتر است. بدیهی است که سه چهارم (۷۵٪) داده‌ها از آن بزرگتر باشد.

همینطور Q_2، چارک دوم است و داده‌ای است که نصف (۵۰٪) داده‌ها از آن کوچکتر است. بدیهی است که نصف (۵۰٪) دیگر داده‌ها از آن بزرگتر باشد.

چارک دوم با میانه یکسان است

همینطور Q_3، چارک سوم است و داده‌ای است که سه چهارم (۷۵٪) داده‌ها از آن کوچکتر است. بدیهی است که یک چهارم (۲۵٪) دیگر داده‌ها از آن بزرگتر باشد.

شاخص های مرکزی

چندک ها

توجه داشته باشید که:

- به شاخص های میانه، چارک ها، دهک ها و صدکها، چندک ها اطلاق می شود.
- هر چندک با یک نسبت که آن را با p نشان میدهیم، معین می شود.
- چندک ها در مقیاس اسمی قابل تعریف نیسند.

روش محاسبه چندك ها

ابتدا داده ها را مرتب کنید سپس تعداد داده ها (n) را در نسبت مربوط به چندک (p) ضرب کنید (در صورت اعشاری بودن آن را به بالا گرد کنید) تا محل قرار گرفتن چندک معلوم شود.

مثال: صدک ٦٥ داده‌های زیر را معلوم می کنیم. (داده‌ها مرتب شده اند)

٢/٣ ٥/٣ ٥/٣ ٦/٣ ٧/٣ ٨/٣ ٩/٣ ٤ ٣/٤ ٤/٤

برای محاسبه صدک ٦٥ باید تعداد داده ها را در ٦٥٪ ضرب کنیم. (٦٥٪ نسبت مربوط به صدک ٦٥ است)

١٠×٠ /٦٥ = ٦ / ٥ ٧

بنابراین عدد هفتم، صدک ٦٥ ام است یعنی: $P_{٦٥}$ =٩/٣

شاخص های مرکزی

چندک ها

مثال: دهک چهارم داده‌های زیر را معلوم می کنیم. (داده‌ها مرتب شده اند)

(۰/٤ نسبت مربوط به دهک چهارم است)

۲/۳ ۳ ۳/۱ ۳/۲ ۳/٤ ۳/۸ ٤ ٤/٤ ٥/۳ ٥/۳ ٦/۱ ٦/۹ ۷/٥ ۸/۳ ۹/۸

برای محاسبه دهک چهارم باید تعداد داده ها (n=۱٥) را در ۰/٤ ضرب کنیم، داریم:

$$n \times p = ۱٥ \times ۰/٤ = ٦$$

چون $n \times p$ اعشاری نیست، معدل عدد ششم و هفتم را حساب کنیم، یعنی:

$$p_{٤} = \frac{۳/۸ + ٤}{۲} = ۳/۹$$

مثال: برای داده‌های جدول زیر چارک اول را بدست آورید

x_i	f_i	F_i
5	3	3
7	4	7
8	5	12
9	18	30
10	10	40

تعداد داده ها (n=٤۰) را در ۲٥٪ ضرب کنیم داریم:

$$n \times p = ٤٠ \times ۰/۲٥ = ۱۰$$

عدد دهم در طبقه سوم است (از فراوانی تجمعی استفاده کنید)

پس چارک اول عدد ۸ است.

شاخص های مرکزی

چندک ها

اگر داده ها پیوسته و دسته بندی هستند، برای محاسبه چندک ها مراحل زیر را دنبال کنید

- ستون فراوانی تجمعی را تشکیل دهید
- مقدار $n \times p$ را محاسبه کنید
- از ستون فراوانی تجمعی طبقه چندک دار را معلوم کنید.
- از فرمول مقابل استفاده کنید

$$Q = L_i + \frac{np - F_{i-1}}{f_i} \times C$$

مثال: برای داده‌های جدول زیر دهک سوم را بدست آورید.

برای دهک سوم ستون فراوانی تجمعی را به دست می آوریم

$$n \times p = ۵۰ \times ۰/۳ = ۱۵$$

عدد پانزدهم در طبقه سوم است.

حدود طبقات	f_i	F_i
1 – 4	8	8
4 – 7	5	13
7 – 10	13	26
10 – 13	15	41
13 - 16	9	50

$$p_۳ = ۷ + \frac{۱۵ - ۱۳}{۱۳} \times ۳ = ۷ + ۰/۴۶ \Rightarrow p_۳ = ۷/۴۶$$

فاصله طبقه

شاخص های مرکزی

میانگین اصلاح شده

داده های دور افتاده

وجود معدود داده‌هائی غیر معمول یا دورافتاده که خیلی بزرگ یا کوچک باشند، تأثیر نامطلوبی در نتایج به دست آمده از تحلیل های آماری دارند. به ویژه این تاثیر در کاهش یا افزایش میانگین بسیار مشهود است. از طرفی میانه بجز با یک یا دو مقدار وسطی داده‌ها از بقیه آنها تاثیر نمی پذیرد. بنا بر این در چنین مواردی میانگین های *پیراسته* و *وینزوری* را می‌توان تلفیقی از میانگین حسابی و میانه تصور کرد.

میانگین پیراسته

اگر داده‌های کوچکتر از چارک اول و بزرگتر از چارک سوم را از مجموعه داده ها حذف کنیم و میانگین داده‌های باقیمانده را به دست آوریم به آن میانگین پیراسته می‌گوئیم

میانگین وینزوری

اگر به جای هر یک از داده‌های کوچکتر از چارک اول، چارک اول و به جای داده‌های بزرگتر از چارک سوم، مقدار چارک سوم را بگذاریم و میانگین داده‌ها را بدست آوریم، به آن میانگین وینزوری می گوییم.

شاخص های مرکزی

میانگین اصلاح شده

مثال: در داده‌های زیر میانگین پیراسته و میانگین وینزوری را بدست می آوریم.

۱ ۱۲ ۱۴ ۱۵ ۱۷ ۲۰ ۲۵ ۲۸ ۳۰

ابتدا چارک اول وسوم داده‌ها را حساب می‌کنیم.

$$\begin{cases} \frac{۱}{٤} \times ۱۰ = ۲/۵ \quad ۳ \\ Q_۱ = ۱٤ \end{cases} , \quad \begin{cases} \frac{۳}{٤} \times ۱۰ = ۷/۵ \quad ۸ \\ Q_۳ = ۲۵ \end{cases}$$

برای محاسبه میانگین پیراسته، مقادیر ۱و ۱۲ که کمتر از چارک اول و همچنین ۲۸ و ۳۰ را که بیشتر از چارک سوم هستند،از بین داده‌ها حذف می‌کنیم ومیانگین داده های باقیمانده را حساب می کنیم. داریم:

$$\text{میانگین پیراسته} = \frac{۱}{٦}(۱٤ + ۱۵ + ۱۷ + ۲۰ + ۲۵ + ۲۵) = ۱۸/۳$$

برای محاسبه میانگین وینزوری، به جای مقادیر ۱و ۱۲ که کمتر از چارک اول هستند مقدار ۱٤و همچنین به جای ۲۸ و ۳۰ که بیشتر از چارک سوم هستند، مقدار ۲۵ را جایگذاری می کنیم و میانگین داده هارا حساب می کنیم. داریم:

$$\text{میانگین وینزوری} = \frac{۱}{۱۰}(۱۲ + ۱۲ + ۱٤ + ۱۵ + ۱۷ + ۲۰ + ۲۵ + ۲۵ + ۲۸ + ۳۰) = ۱۹/۶$$

شاخص های مرکزی

میانگین اصلاح شده

میانگین اصلاح شده

روش کاملتر برای حذف اثر سوء داده های دورافتاده، این است که pدرصد ابتدا و pدرصد انتهای داده ها را حذف کنیم و میانگین باقیمانده داده ها را بدست آوریم. این میانگین به میانگین اصلاح شده pدرصد موسوم است.

اگر تعداد نمونه ها کم و مقادیر دور افتاده قابل تشخیص هستند، میانگین اصلاح شده را از فرمول مقابل به دست می آوریم و به آن میانگین اصلاح شده مرتبه kام می گوییم.

$$T_k = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} x_i$$

در این رابطه kتعداد داده هایی است که دور افتاده اند و قرار است از مجموعه داده ها حذف شوند. البته باید داده ها مرتب شده باشند و $k < \frac{n}{2}$ باشد.

معمولا وقتی تعداد نمونه ها زیاد است، از میانگین اصلاح شده ۵ درصد، استفاده می کنند

توجه داشته باشید که:

میانگین پیراسته، یک میانگین اصلاح شده ۲۵ درصد است

شاخص های پراکندگی

وقتی یک جدول توزیع فراوانی را با دقت بررسی کنید مشاهده خواهید کرد عاملی که باعث می‌شود فراوانی ها بین مقادیر مختلف یا در فواصل مختلف توزیع گردند، ناشی از پراکندگی داده‌ها حول میانگین است که شاخصهای مرکزی قادر به بیان آنها نیستند.

شاخص های پراکندگی

شاخص هایی عددی هستند که میزا ن پراکندگی داده ها حول میانگین را اندازه می گیرند

مثال: دو گروه از دانشجویان را از نظر مطالعه تعداد کتاب های غیر درسی مورد بررسی قرار داده ایم که نتایج در دو جدول زیر آمده است.

x_i	f_i
٢	٢
٣	٣
٤	١٥
٥	١٠
٦	١٠
٧	٤٠
٨	١٥
٩	٣
١٠	٢

x_i	f_i
٤	٥
٥	١٠
٦	٧٠
٧	١٠
٨	٥

برای این دو گروه، میانگین، میانه و نما یکسان و برابر ٦ است

مهمترین شاخصهای پراکندگی

- دامنه تغییرات
- واریانس
- انحراف معیار
- ضریب تغییرات
- ضریب چولگی

شاخص های پراکندگی

دامنه تغییرات

دامنه تغییرات

دامنه تغییرات داده ها حدود نوسان تغییرات صفت متغیر را معلوم می کند و آن را از رابطه زیر به دست می آوریم

$$R = X_{max} - X_{min}$$

دامنه میان چارکی

این شاخص مانند دامنه تغییرات است و حدود تغییرات بین چارک های اول و سوم را معین می کند.

$$Q = Q_{۳} - Q_{۱}$$

مثال: در داده‌های زیر دامنه تغییرات و دامنه میان چارکی را بدست می آوریم.

۱ ۱۲ ۱۴ ۱۵ ۱۷ ۲۰ ۲۵ ۲۸ ۳۰

دامنه تغییرات داده ها

$$R = X_{max} - X_{min} = ۳۰ - ۱ = ۲۹$$

چارک اول وسوم داده‌ها را حساب می‌کنیم.

$$\frac{۱}{۴} \times ۱۰ = ۲/۵ \quad ۳$$

$Q_{۱} = ۱۴$ چارک اول

$$\frac{۳}{۴} \times ۱۰ = ۷/۵ \quad ۸$$

$Q_{۳} = ۲۵$ چارک سوم

دامنه میان چارکی $R = Q_{۱} - Q_{۳} = ۲۵ - ۱۴ = ۱۱$

شاخص های پراکندگی

واریانس

واریانس

واریانس یکی از مهمترین شاخص های پراکندگی است که قادر است پراکندگی داده‌ها نسبت به میانگین را به خوبی بیان کند. واریانس از رابطه زیر محاسبه می شود.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (1)$$

میانگین توان دوم داده ها را به صورت $\overline{x^2} = \frac{1}{n}\sum_{1}^{n} x_i^2$ تعریف می کنیم. بنابراین واریانس را می توان به طور معادل از فرمول (۲) نیز حساب کرد.

$$s^2 = \overline{x^2} - \bar{x}^2 \quad (2)$$

مثال: برای داده‌های زیر واریانس را به دست می آوریم. ۳ ۲ ۴ ۳ ۶ ۳ ۵

ابتدا میانگین داده‌ها را حساب می کنیم و سپس از فرمول (۱) واریانس را محاسبه می کنیم، داریم:

$$\bar{x} = \frac{1}{8}(3+2+4+3+6+3+5+6) = \frac{32}{8} = 4$$

$$S^2 = \frac{1}{8}[(3-4)^2 + (2-4)^2 + (4-4)^2 + (3-4)^2 + (6-4)^2 + (3-4)^2 + (5-4)^2 + (6-4)^2] = \frac{1}{8}(16) = 2$$

توجه کنید که در این مثال داده ها دسته بندی نیستند

شاخص های پراکندگی

واریانس

مثال: برای داده‌های جدول توزیع فراوانی زیر واریانس را محاسبه کنید.

X_i	f_i	$f_i x_i$	$f_i x_i^2$
0	3	0	0×3=0
1	7	7	1×7=7
2	21	42	4×21=84
3	6	18	9×6=54
4	7	28	16×7=112
5	6	30	25×6=150
جمع	50	125	407

ابتدا با تشکیل دو ستون آخر جدول مقادیر زیر را به دست می آوریم

$$\bar{x} = \frac{۱۲۵}{۵۰} = ۲/۵ \qquad \overline{x^۲} = \frac{۴۰۷}{۵۰} = ۸/۱۴$$

سپس واریانس از فرمول (۲) به صورت زیر بدست می آید.

$$S^۲ = ۸/۱۴ - (۲/۵)^۲ = ۱/۸۹$$

توجه کنید که در این مثال داده ها گسسته و دسته بندی شده اند

مثال: برای داده‌های جدول توزیع فراوانی زیر، واریانس را محاسبه کنید

حدود طبقات	f_i	x'_i	$f_i . x'_i$	$f_i . x'^2_i$
7 -11	2	9	18	162
11 -15	3	13	39	507
15 -19	5	17	85	1445
19 -23	10	21	210	4410
23 -27	2	25	50	1250
	22		402	7774

برای محاسبه واریانس از فرمول (۲) استفاده می کنیم
ابتدا با اضافه کردن سه ستون آخر جدول، مقدار آن را به دست می آوریم، داریم:

$$\bar{x} = \frac{۱}{۲۲}(٤٠٢) = ١٨/٢٧$$

$$\overline{x^{۲}} = \frac{۱}{۲۲}(٧٧٧٤) = ٣٥٣/٣٦$$

$$\Rightarrow \quad S^{۲} = ٣٥٣/٣٦ - (١٨/٢٧)^{۲} = ١٩/٥٧$$

{ین جزوه توسط وب سایت تست دانلود عرضه شده است، testdownload.ir}

خواص واریانس

۱- اگر همه داده‌ها باهم برابر باشند واریانس داده‌ها صفر است و برعکس

$$x_1 = x_2 = \dots = x_n \Rightarrow S_x^2 = 0$$

۲- واریانس داده‌ها مقداری مثبت است.

$$S^2 \geq 0$$

۳- اگر به داده‌ها عدد ثابتی مانند a را اضافه یا کم کنیم، واریانس تغییری نمی‌کند.

$$Y_i = a \pm X_i \rightarrow S_Y^2 = S_x^2$$

٤- اگر داده‌ها را در عدد ثابتی مانند a ضرب کنیم، واریانس آنها در مجذور a ضرب می‌شود.

$$Y_i = ax_i \Rightarrow S_Y^2 = a^2 S_x^2$$

به همین صورت در مورد تقسیم بر عدد ثابتی مانند a، داریم:

$$Y_i = \frac{X_i}{a} \rightarrow S_Y^2 = \frac{S_x^2}{a^2}$$

شاخص های پراکندگی

واریانس

مثال: جدول توزیع فراوانی زیر را در نظر بگیرید. اگر $Y_i = X_i + ۵$ و $Z_i = ۳X$ باشد مطلوب است:

$S_y^۲, S_z^۲, S_x^۲$?

X_i	0	1	2	3
f_i	1	2	8	4

حل: ابتدا واریانس x را به دست می آوریم سپس از خواص ۳ و ٤ واریانس استفاده می کنیم

X_i	f_i	$x_i f_i$	X_i^2	$x_i^۲ f_i$
0	1	0	0	0
1	2	2	1	2
2	8	16	4	32
3	4	12	9	36
جمع	15	30	14	70

$$\bar{x} = \frac{۳۰}{۱۵} = ۲$$

$$\bar{x}^۲ = \frac{۷۰}{۱۵}$$

$$\Rightarrow \quad S_x^۲ = \frac{۷۰}{۱۵} - ٤ = ۰/٦٦$$

بنابر این واریانس y و z به این صورت به دست می آید.

$$S_y^۲ = S_x^۲ = ۰/٦٦$$

$$S_z^۲ = ۹\, S_x^۲ = ۹ \times ۰/٦٦ = ۵/۹٤$$

واریانس کل

اگر یک مجموعه n_1 تائی از داده‌ها با واریانس S_1^2 و یک مجموعه n_2 تائی از داده‌ها با واریانس S_2^2 داشته باشیم، می توانیم واریانس کل داده‌ها را به صورت مقابل بدست آوریم

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$$

واریانس کل را می توان برای K جمعیت (Kمتناهی) نیز به همین صورت تعمیم داد.

مثال: در یک مجموعه با ۱۰ داده، واریانس ۷ و در مجموعه دیگری با ۱۵ داده واریانس ٤ است. واریانس کل داده‌ها چقدر است؟

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2} = \frac{(۱۰ \times ۷) + (۱۵ \times ٤)}{۱۰ + ۱۵} = \frac{۱۳۰}{۲۵} = ۵/۲$$

ازآنجا که واریانس توان دوم انحرافات از میانگین داد ها را نشان میدهد، انحراف معیار می تواند به عنوان یک معیار پایه‌ای برای تعیین میزان پراکندگی داده ها به کار رود

مثلا اگر واحد داده‌ها سانتیمتر باشد واحد واریانس سانتیمتر مربع است.

انحراف معیار

انحراف معیار عبارت است از جذر مثبت واریانس و آن را با S نشان میدهیم و مینویسیم:

$$S = \sqrt{S^{۲}}$$

مثال: انحراف معیار مثال قبل را به دست آورید

$$S = \sqrt{۱۹/۵۷} = ٤/٤۲$$

از آنجایی که واریانس به واحد اندازه‌گیری بستگی دارد به طوری که اگر واحد اندازه‌گیری طول یک نقطه را از سانتیمتر به اینچ تغییر دهیم، واریانس آن تغییرخواهد کرد. بنابراین از معیار دیگری که به واحد اندازه‌گیری بستگی نداشته باشد، استفاده می‌کنیم

ضریب تغییرات در داده هایی با میانگین صفر، تعریف نمی شود.

ضریب تغییرات

ضریب تغییرات داده ها که آن را با C.V نشان می دهیم از نسبت انحراف معیار به میانگین به دست می آید و به صورت زیر می نویسیم.

$$C.V = \frac{S}{\bar{x}} \times ۱۰۰$$

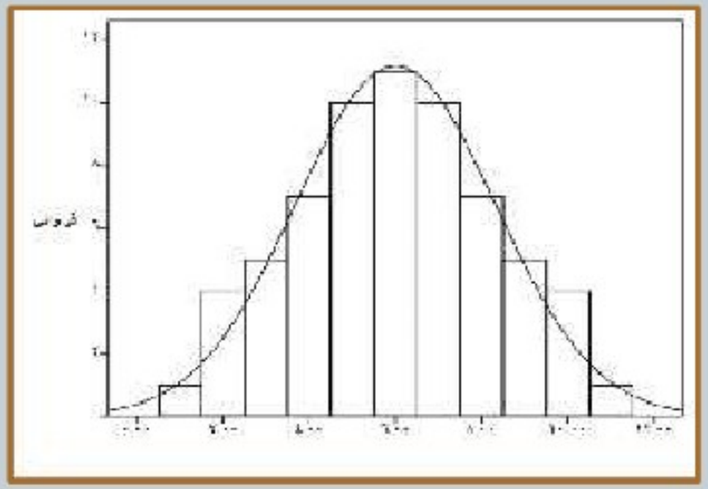
مثال: اگر در یک نمونه، میانگین داده‌ها ۵ و واریانس داده ها ٤ باشد؛ ضریب تغییرات داده‌ها به صورت زیر به دست می آید:

$$C.V = \frac{۲}{۵} \times ۱۰۰ = ٤۰$$

تقارن و چولگی

ممکن است دو مجموعه داده، از لحاظ گرایش به مرکز و پراکندگی تفاوت چندانی با هم نداشته باشند ولی از نظر تقارن با یکدیگر متفاوت باشند. از این رو لازم است تا شاخص های دیگری برای اندازه گیری میزان تقارن یا عدم تقارن داده ها معرفی شوند.

تقارن در توزیع داده ها

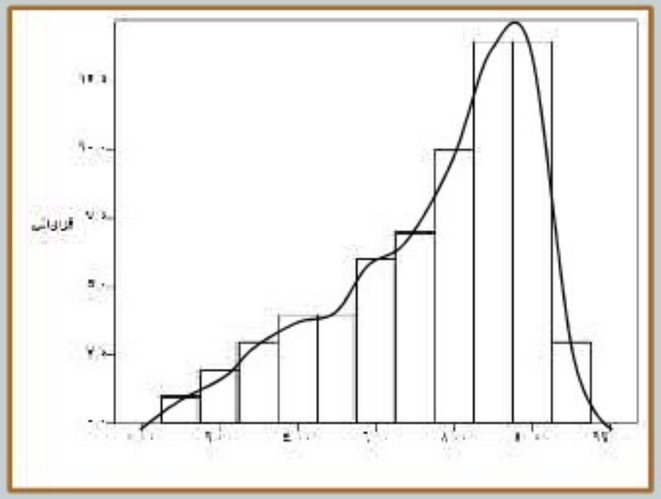


توجه داشته باشید:

در توزیع‌های متقارن تقریبا میانگین، میانه و نما بر هم منطبق اند.

$$Mo = \bar{x} = Md$$

عدم تقارن (چولگی)



ضریب چولگی بر حسب چارک ها

اگر چارک های اول، دوم و سوم یک توزیع را در اختیار داشته باشیم، ضریب چولگی چارکها را به شکل زیر می نویسیم:

$$SK = \frac{(Q_۳ - Q_۲) - (Q_۲ - Q_۱)}{(Q_۳ - Q_۲) + (Q_۲ - Q_۱)} = \frac{Q_۳ + Q_۱ - ۲Q_۲}{Q_۳ - Q_۱}$$

مثال: اگر در یک مجموعه داده، چارک های اول، دوم و سوم به ترتیب $Q_۱ = ۵$ و $Q_۲ = ۹$ و $Q_3 = 11$ باشند، ضریب چولکی برحسب چارکها را بدست آورید.

$$SK = \frac{۱۱ + ۵ - ۱۸}{۱۱ - ۵} = -\frac{۲}{۶} = -\frac{۱}{۳}$$

ضریب چولگی پیرسن

یکی دیگر شاخص های اندازه گیری میزان تقارن و عدم تقارن داده ها، ضریب چولگی پیرسن است. اگر $\overline{x}$ میانگین، Mo نما، Md میانه و S انحراف معیار یک مجموعه داده باشند؛ دو ضریب چولگی که هر دو منسوب به پیرسن است، از رابطه های زیر بدست می‌آیند:

ضریب چولگی نوع اول

$$SK = \frac{(\overline{x} - Mo)}{S}$$

ضریب چولگی نوع دوم

$$SK = \frac{۳(\overline{x} - Md)}{S}$$

شاخص های پراکندگی

مثال: در داده‌های جدول زیر، ضریب چولگی پیرسن و ضریب چولگی برحسب چارک ها را بدست آورده ایم.

برای محاسبه ضریب چولگی پیرسن، ابتدا سه شاخص میانگین، انحراف معیار و میانه داده ها را به دست می آوریم.

[برای محاسبات به این ستونها نیاز داریم]

حدود طبقات	f_i	x'_i	$f_i.x'_i$	$f_i.x_i^2$	F_i
7 -11	2	9	18	162	2
11-15	3	13	39	507	5
15-19	5	17	85	1445	10
19-23	10	21	210	4410	20
23-27	2	25	50	1250	22
	22		402	7774	

با توجه به مقدار ضریب چولگی، توزیع داده ها چوله به چپ است.

$$\bar{x} = \frac{۴۰۲}{۲۲} = ۱۸/۲۷$$

$$\overline{x^۲} = \frac{۷۷۷۴}{۲۲} = ۳۵۳/۳۶$$

$$\Rightarrow S^۲ = ۲۵۳/۳۶ - (۱۸/۲۷)^۲ = ۱۹/۵۷$$

$$S = \sqrt{۱۹/۵۷} = ۴/۴۲$$

$$_i + \frac{n\,p - F_{i-۱}}{f_i} \times C$$

$$Md = ۱۹ + \frac{۱۱ - ۱۰}{۱۰} \times \Rightarrow ۱۹ \cdot ۴ \quad ۱۹/۴$$

$$Sk = \frac{۳(\bar{x} - Md)}{S} = \frac{۳(۱۸/۲۷ - ۱۹/۴)}{۴/۴۲} = -۰/۷۷$$

ضریب چولگی پیرسن

اینک برای محاسبه ضریب چولگی برحسب چارک ها، باید چارک ها را بدست می‌آوریم

$$Q_1 = \begin{cases} 1)\ \frac{1}{4} \times 22 = 5/5 \quad 6 \\ 3)\ Q_1 = L_i + \frac{np - F_i}{f_i} \times C \end{cases} \Rightarrow Q_1 = 15 + \frac{5/5 - 5}{5} \times 4 \Rightarrow Q_1 = 15 + 0/4 = 15/4$$

$$Q_3 = \begin{cases} 1)\ \frac{3}{4} \times 22 = 16/5 \quad 17 \\ 3)\ Q_3 = 19 + \frac{16/5 - 10}{10} \times 4 = 19 + 2/6 = 21/6 \end{cases}$$

با توجه به مقدار ضریب چولگی، توزیع داده ها چوله به چپ است.

ضریب چولگی بر حسب چارک ها

$$S_k = \frac{(21/6 - 19/4) - (19/4 - 15)}{(21/6 - 19/4) + (19/4 + 15)} = \frac{-2/8}{7/2} = -0/39$$

شاخص های پراکندگی

گشتاورها

به $M_r(a)$ گشتاور مرتبه r ام، حول مبدا a می گوییم و به صورت مقابل تعریف می کنیم:

$$M_r(a) = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^r$$

بدیهی است اگر داده ها فراوانی داشته باشند این گشتاور به صورت مقابل به دست خواهد آمد:

$$M_r(a) = \frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - a)^r$$

میانگین داده ها، گشتاور مرتبه اول حول مبدا صفر است

$$M_1(0) = \frac{1}{n}\sum_{i=1}^{n}(x_i) = \bar{x}$$

واریانس، گشتاور مرتبه دوم حول میانگین داده ها است

$$M_2(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2$$

شاخص های پراکندگی

ضریب چولگی گشتاوری

یک روش اصولی برای تعیین میزان تقارن و چولگی داده ها، استفاده از ضریب چولگی بر حسب گشتاورها است که آن را به صورت مقابل معرفی می کنیم:

$$SK = \frac{M_{۳}(\bar{x})}{S^{۳}} = \frac{\frac{۱}{n}\sum_{i=۱}^{n}(x_i - \bar{x})^{۳}}{S^{۳}}$$

در این فرمول، S انحراف معیار داده ها است و به این دلیل آن را در فرمول بالا گنجانده اند تا شاخص ضریب چولگی گشتاورها به واحد اندازه گیری داده ها بستگی نداشته باشد.

مثال: در ۵۰ داده آماری با میانگین ۱۵ و واریانس ٤ داریم:

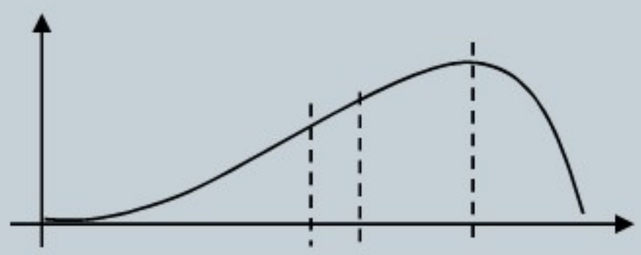
$$M_{۳}(\bar{x}) = \frac{۱}{۵۰}\sum_{i=۱}^{۵۰}(x_i - ۱۵)^{۳} = ۲٤$$

ضریب چولگی گشتاوری داده ها به صورت زیر به دست می آید:

$$SK = \frac{\frac{۱}{n}\sum_{i=۱}^{n}(x_i - \bar{x})^{۳}}{s^{۳}} = \frac{\frac{۲۴}{۵۰}}{۲^{۳}} = ۰/۰۶$$

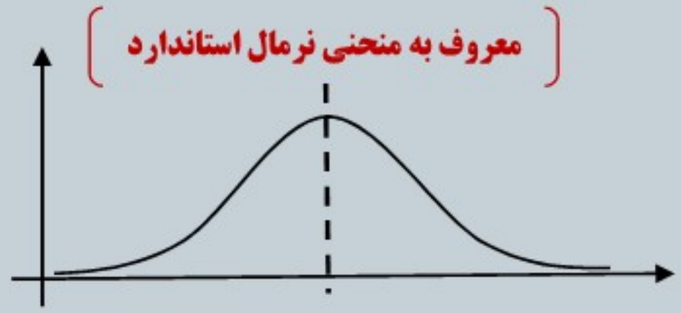
تفسیر چولگی

ضریب چولگی ممکن است مقداری منفی، مثبت یا صفر باشد. بر این اساس وضعیت قرار گرفتن سه شاخص مهم مرکزی، میانگین، میانه و نما را در حالت های مختلف ببینید.



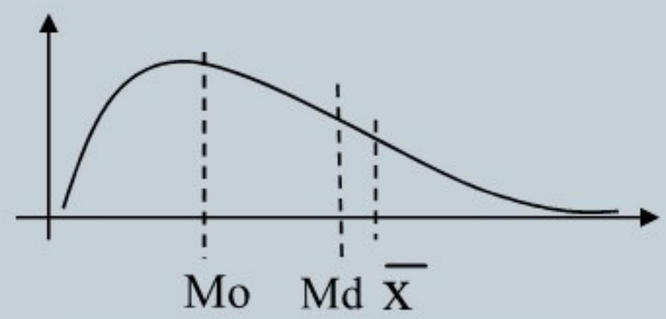
$SK \leq 0$

در این صورت توزیع داده ها نا متقارن است و مي گوييم داده ها چوله به چپ هستند



$SK = 0$

در این صورت توزیع داده ها متقارن است و شاخص های میانه، میانگین و نما بر هم منطبق اند.



$SK \geq 0$

در این صورت توزیع داده ها نا متقارن است و مي گوييم داده ها چوله به راست هستند

روابط تجربی بین میانگین و انحراف معیار

اگر داده ها تقریباً متقارن باشند، در صورتی که میانگین و انحراف معیار داده ها را در اختیار داشته باشیم، می توان به طور تجربی، فاصله های زیر را برای داده ها در نظر گرفت:

در حدود 68 درصد داده‌ها در فاصله $(\bar{x} - s \ , \ \bar{x} + s)$ قرار دارند.

در حدود 95 درصد داده‌ها در فاصله $(\bar{x} - ۲s \ , \ \bar{x} + ۲s)$ قرار دارند.

در حدود 99 درصد داده‌ها در فاصله $(\bar{x} - ۳s \ , \ \bar{x} + ۳s)$ قرار دارند.

مثال: توزیع نمرات یک درس دانشجویان تقریبا متقارن است. اگر میانگین نمرات آنها ۱۵ و واریانس نمرات ۲/۲۵ باشد، فاصله ای را که شامل ٦٨ درصد نمره های دانشجویان است تعیین می کنیم
تقریباً ٦٨ درصد نمرات در فاصله (۵/۱٦ و ۵/۱۳) قرار دارند، زیرا:

$$(\bar{x} - S \ , \ \bar{x} + S) = (۱۵ - ۱/۵ \ , ۱۵ + ۱/۵) = (۱۳/۵ \ , ۱٦/۵)$$

آمار توصیفی

در عصر حاضر کسی نمی‌تواند منکر این واقعیت باشد که آمار نقشی لاینفک در زندگی روزمره ما بازی می کند. اخبار روزانه رسانه‌های گروهی با گزارشی از وضع هوا به پایان می‌رسندو در طول اخبار، به جریانهای بازار بورس و سهام اشاره می‌شود و روزنامه‌ها خبر از افزایش نرخ اجناس می‌دهندو...

آمار به عنوان پایه یک روش و راه موثر در بررسی مسائل موجود، در بسیاری از زمینه‌های علمی از جمله جامعه شناسی، کشاورزی، فیزیک و....به کار گرفته می‌شود. در دانش امروزی، معمولا سعی می‌شود که اطلاعات موجود در یک زمینه خاص، در قالب اعداد نمایش داده شود تا به هنگام تجزیه و تحلیل اطلاعات، فهم بهتری از پدیده مورد مطالعه به دست آمده و امکان مقایسه فراهم گردد. در یک جمله آمار مجموعه‌ای از روشهای جمع آوری، تهیه وتنظیم و تجزیه و تحلیل اطلاعات است که برای کسب یک یا چند نتیجه به خدمت گرفته می‌شود.

برای اینکه نتایج مناسب و مطلوب از اطلاعات که در آمار گیری‌ها جمع آوری می کنیم، به دست آید باید:

- ✓ اعداد نماینده واقعی مشاهدات بوده و غیرواقع یا غلط نباشند
- ✓ به نحو مفیدی تهیه و تنظیم شوند
- ✓ به نحو صحیح تجزیه و تحلیل گردند
- ✓ قابل نتیجه گیری صحیح باشند

به طور کلی، روش‌هایی را که به وسیله آنها می‌توان اطلاعات جمع آوری شده را تنظیم کرده و خلاصه نمود، آمار توصیفی می‌نامیم و در یک کلام آمار توصیفی عبارت از مجموعه روشهایی است که پردازش داده‌ها را فراهم می‌سازد. اطلاع از اصطلاحات زیر در آمار ضروری است.

مجموعه افراد یا اشیایی را که می‌خواهیم یک یا چند خصوصیت مشترک آنها را مورد بررسی قرار دهیم، جمعیت یا جمعیت آماری می‌نامیم.

TESTDOWNLOAD.IR

اندازه قد یا وزن دانشجویان بیست ساله یک شهر، تعداد لامپهای سالم و یا ناسالم تولید شده در یک کارخانه و در یک روز معین، مثالهایی از جمعیتهای آماری هستند.

معمولا مطالعه ویژگی‌های مورد نظر، به هنگامی که جمعیت آماری بسیار گسترده باشد، مستلزم صرف هزینه و وقت زیادی می‌باشد و در بسیاری از مواقع، این امر اصولا امکان پذیر نیست. بنابراین در چنین موردی، برای مطالعه ویژگی مورد نظر، به قسمتی از جمعیت آماری اکتفا می کنیم.

قسمتی از جمعیت را که طبق قاعده و ضوابط خاصی، برای مطالعه خصوصیتی از جمعیت انتخاب می‌شود، یک نمونه از جمعیت می‌نامیم.

این نمونه وقتی مفید و قابل قبول خواهد بود که بتواند نماینده خوبی برای کل جمعیت مورد مطالعه باشد. با توجه به اهمیت این موضوع شاخه‌ای از آمار تحت عنوان نظریه نمونه گیری با بررسی نمونه‌ای به این امر مهم می‌پردازد. در بسیاری از موارد، معمولا نمونه تصادفی ساده را در نظر می گیرند.

برای بررسی اندازه قد دانشجویان بیست ساله یک شهر، انتخاب مثلا ۱۵۰ نفر از بین این جمعیت به طور تصادفی، یا انتخاب ۱۰۰ لامپ به تصادف از لامپهای تولیدی یک کارخانه در یک روز معین، برای تعیین کیفیت لامپهای تولیدی این کارخانه مثالهایی از نمونه تصادفی هستند.

خصوصیت مورد مطالعه، از فردی به فرد دیگر، یا از شی به شی دیگر

درجمعیت آماری تغییر می کند، که آن را اصطلاحاً متغیر می‌نامیم.

معمولا دو نوع متغیر در آمار مورد نظر هستند:

متغیرهای گروهی، نظیر رنگ، نژاد، شغل و گروه خونی که شامل چند گروه یا طبقه می‌باشند.

متغیرهای عددی که ممکن است نتیجه شمارش باشد، مانند تعداد احشام هر خانوار در یک روستا،تعداد حوادث در یک کارخانه در روزهای مختلف، و یا نتیجه اندازه‌گیری باشد، مثل قد دانشجویان بیست ساله در یک شهر، حجم شربت مولتی ویتامین با استاندارد خاص.

انواع متغیر

متغیرهای گسسته: متغیرهای گروهی، متغیرهای عددی که از راه شمارش به‌دست آمده اند

متغیرهای پیوسته: متغیرهایی را که از طریق اندازه‌گیری به دست آمده باشند

در بسیار از مسائل پیش‌رو، اندازه‌گیری ویژگی یک متغیر مستلزم آگاهی و شناخت خاصی است. به طور کلی چهار نوع مقیاس برای اندازه گیری وجود دارد:

- مقیاس اسمی
- مقیاس ترتیبی
- مقیاس فاصله‌ای
- مقیاس نسبتی

مقیاس اسمی: این نوع مقیاس اندازه‌گیری عمدتا برای طبقه بندی داده‌ها به کار می‌رود و منظور از آن اتلاق یک عدد طبیعی به داده‌های متفاوت است.

مثل: اختصاص اعداد ۱ تا ۴ به گروه‌های خونی A,B, AB, O.

این اعداد را نمی‌توان برای مقایسه یا چهار عمل اصلی به کار برد

مقیاس ترتیبی: این نوع مقیاس اندازه‌گیری عموما برای طبقه بندی داده‌ها به منظور یک نوع برتری به کار می‌رود.

TEST DOWNLOAD.IR

مثل: در یک کارخانه ممکن است کارگران را به سه دسته ساده، نیمه ماهر و ماهر تقسیم بندی کنیم. اتلاق به ترتیب اعداد ۱ تا ۳ به این سه دسته یک مقیاس ترتیبی است.

این اعداد تنها برای مقایسه به کار می‌روند و نمی‌توان با آنها چهار عمل اصلی را انجام داد.

مقیاس فاصله ای: این نوع مقیاس اندارزه‌گیری عموما در زمینه‌های که علاوه بر حفظ ترتیب به نحوی فاصله بین ویژگی‌ها را نیز حفظ می‌کند. به عبارت دیگر در چنین مقیاسی نسبت تفاضلها ثابت می‌ماند.

مثل: اندازه‌گیری ضریب هوشی دانش آموزان کلاس اول دبستان در شهر اصفهان.

در این نوع مقیاس، عدد صفر یک مفهوم قراردادی است.

مقیاس نسبتی: این نوع مقیاس اندازه‌گیری علاوه بر حفظ فاصله، نسبت را نیز حفظ می‌کند. به عبارت دیگر در این نوع اندازه‌گیری نسبت دو مقدار بستگی به واحد اندازه‌گیری ندارد.

اطلاعاتی که از مطالعه یک متغیر به دست می‌آیند، معمولا شامل انبوهی عدد یا علامت می‌باشند که آنها را داده می نامیم. داده‌ها را نسبت به نوع متغیری که اندازه گیری می‌کنیم به دو دسته داده گسسته و داده‌های پیوسته تقسیم می کنیم.

معمولا به داده‌های جمع آوری شده که انبوهی عدد است و هیچ نوع پردازشی روی آنها انجام نشده است داده خام می‌گویند.

مواردی که در ارتباط با یک مجموعه از داده‌های می‌بایستی مد نظر قرار داد،عبارت‌اند از:

خلاصه کردن و توضیح داده‌ها به وسیله تنظیم جداول و رسم نمودارها.

TESTDOWNLOAD.IR

محاسبه مقادیر عددی، برای دست یافتن به معیارهایی که تمرکز و یا پراکندگی داده‌ها را نشان دهد.

در آمار،برای اینکه از داده‌های خام واقعیتهای موجود را استخراج کنیم،آنها را به نحوی مناسب دسته‌بندی کرده و جدولهایی به نام جدولهای آماری تهیه می‌نماییم. متداولترین جدول در آمار، جدول فراوانی است.

پیش از آنکه نحوه تنظیم جدول فراوانی را بیان نماییم،اطلاع از اصطلاحات زیر ضروری است.

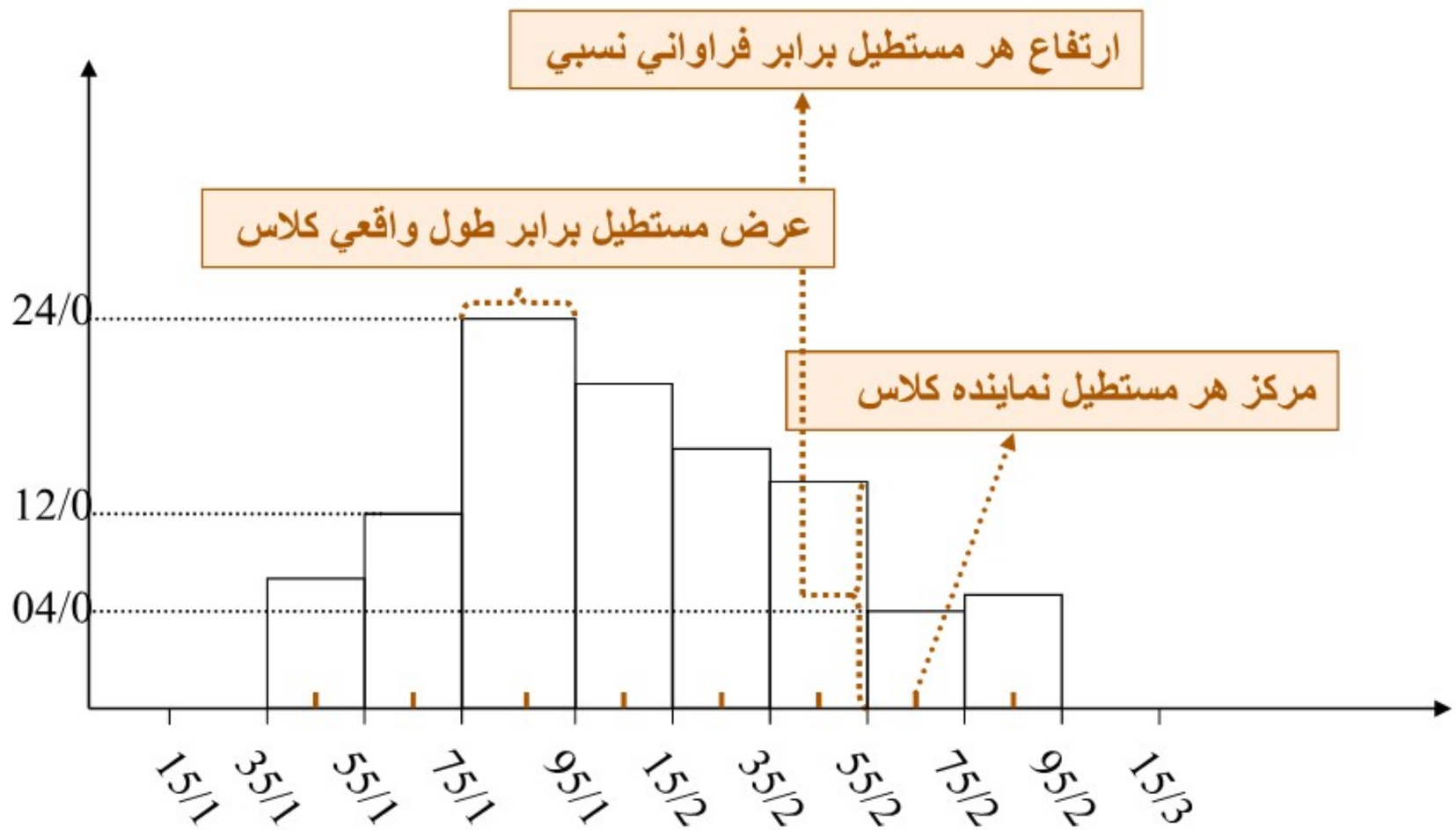
{ین جزوه توسط وب سایت تست دانلود عرضه شده است، testdownload.ir}

معمولا داده‌ها را با نمودارهای مختلف نمایش می‌دهند. عموما این نمودارها در ارتباط با داده‌های پیو.سته به کار گرفته می شود و منظور از نمایش آنها، تجسم عینی اطلاعات نهفته در داده‌ها است. در این بخش به معرفی چند نمودار معروف اکتفا می کنیم:

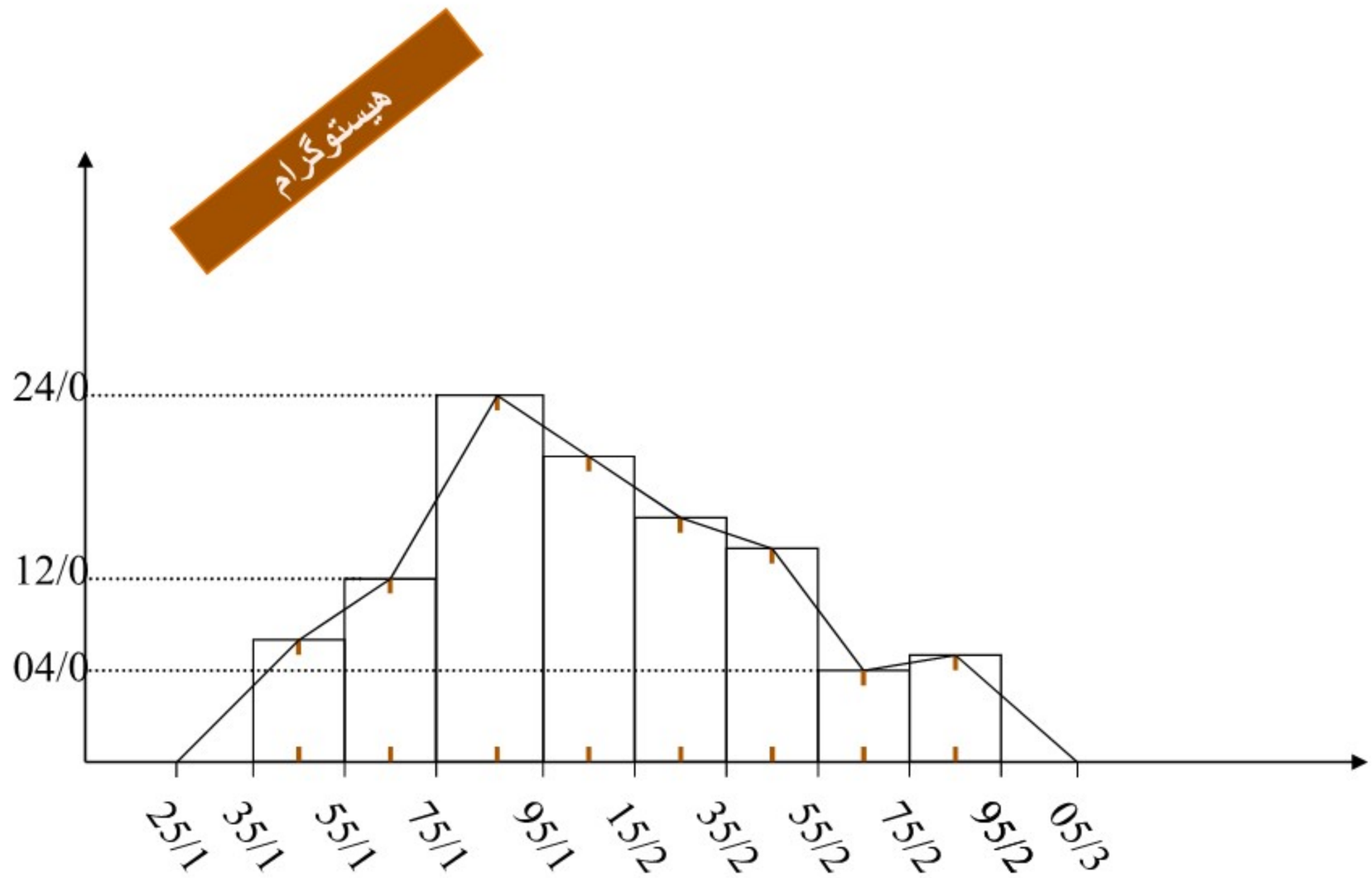
- هیستوگرام
- چندبر فراوانی
- چندبر فراوانی تجمعی
- منحنیهای فراوانی و فراوانی تجمعی
- نمایش نمودار تنه و شاخه
- نمودار جعبه‌ای

TESTDOWNLOAD.IR

نمودارهای آماری:



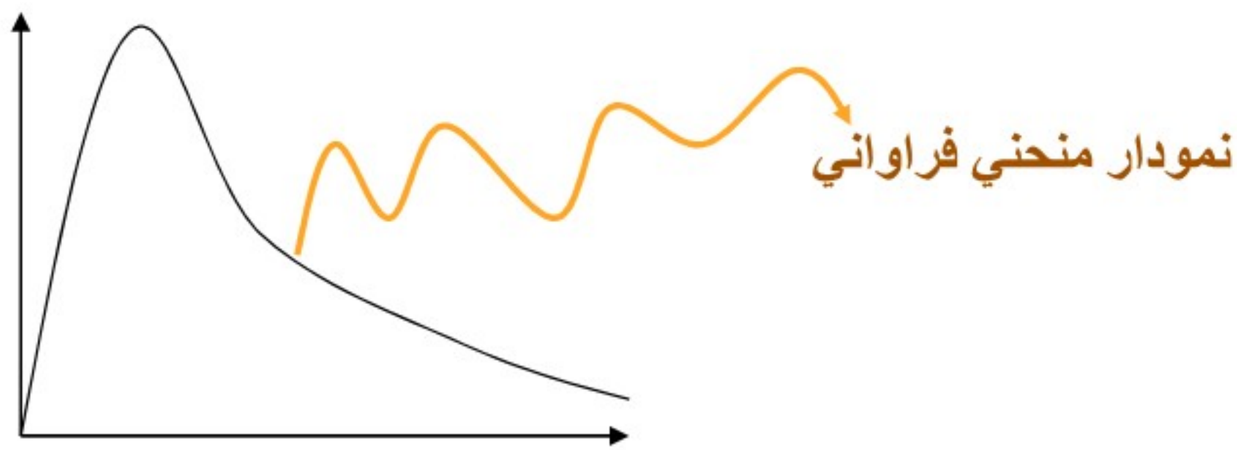
نمودارهای آماری:

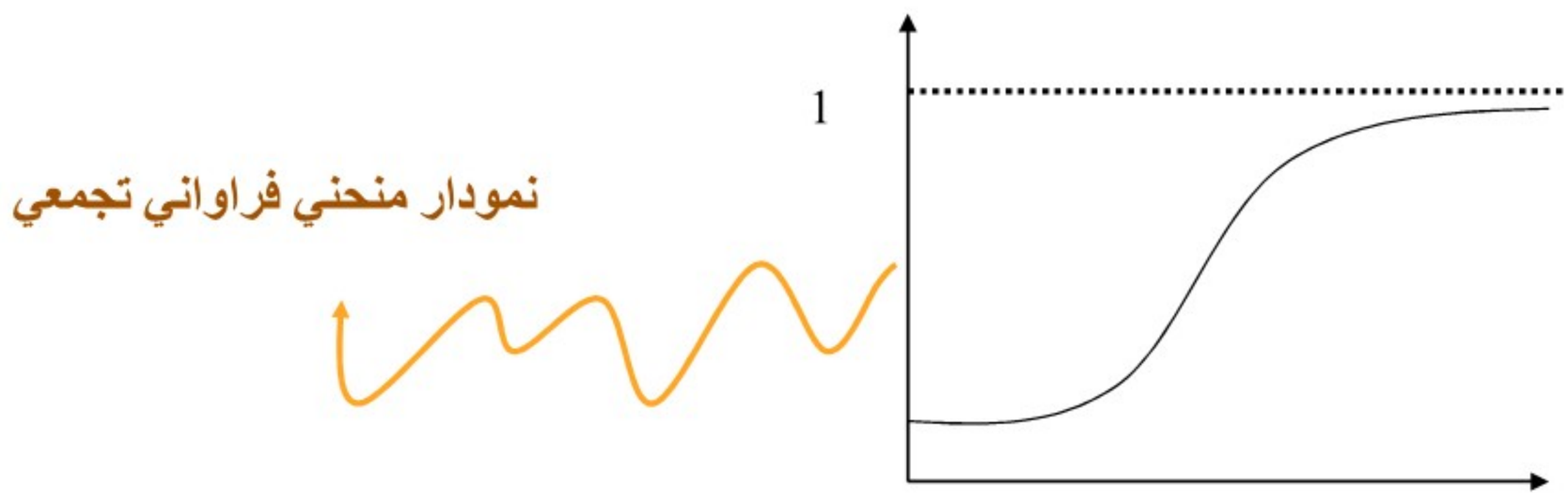


نمودارهای آماری:

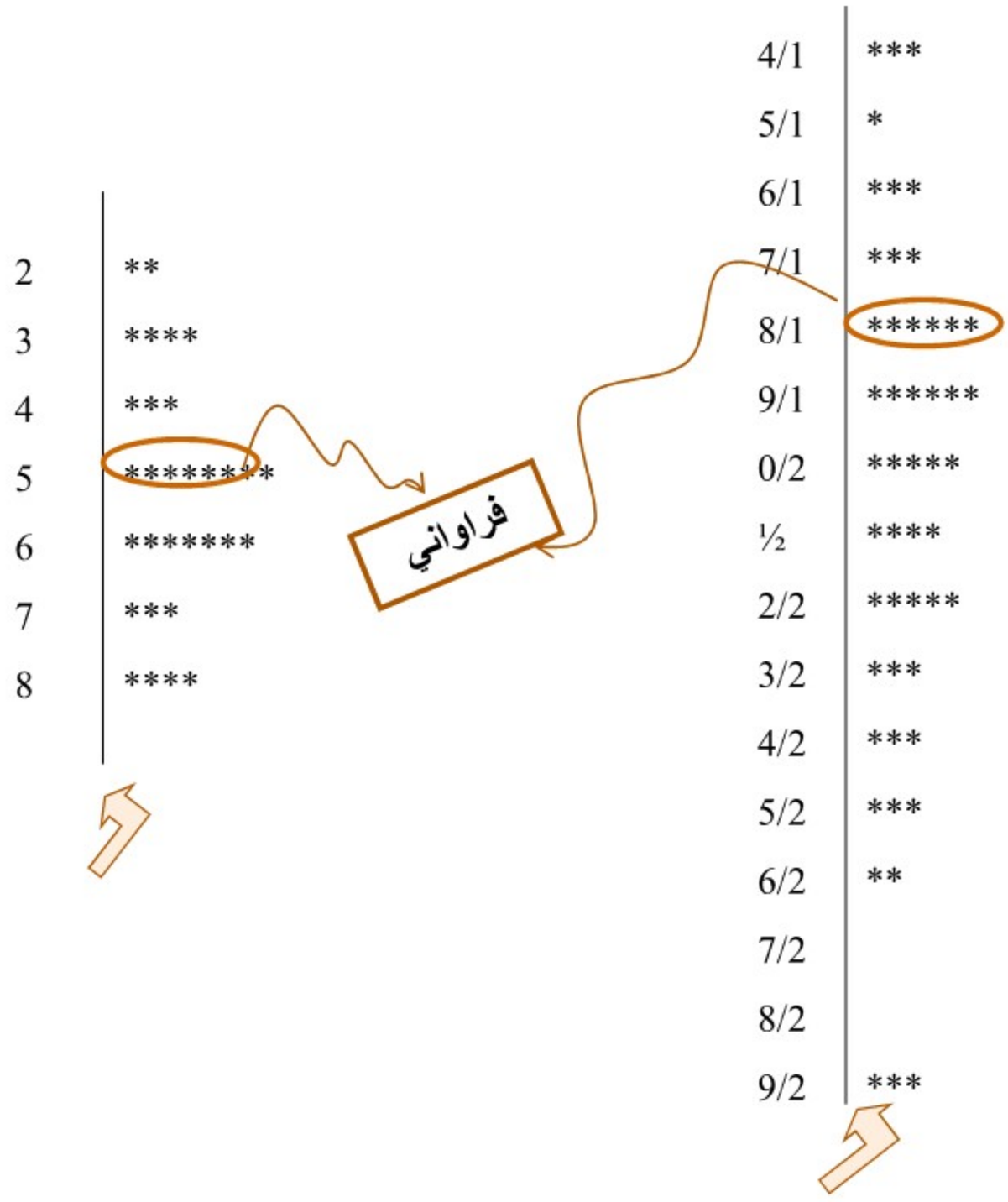
از اتصال نقاطی که طول آنها مرز کلاس و عرض آنها فراوانی نسبی تجمعی تا آن مرز ، یک خط شکسته به دست می‌آید که باشد آن را چندبر فراوانی تجمعی می‌نامند

نمودارهای آماری:





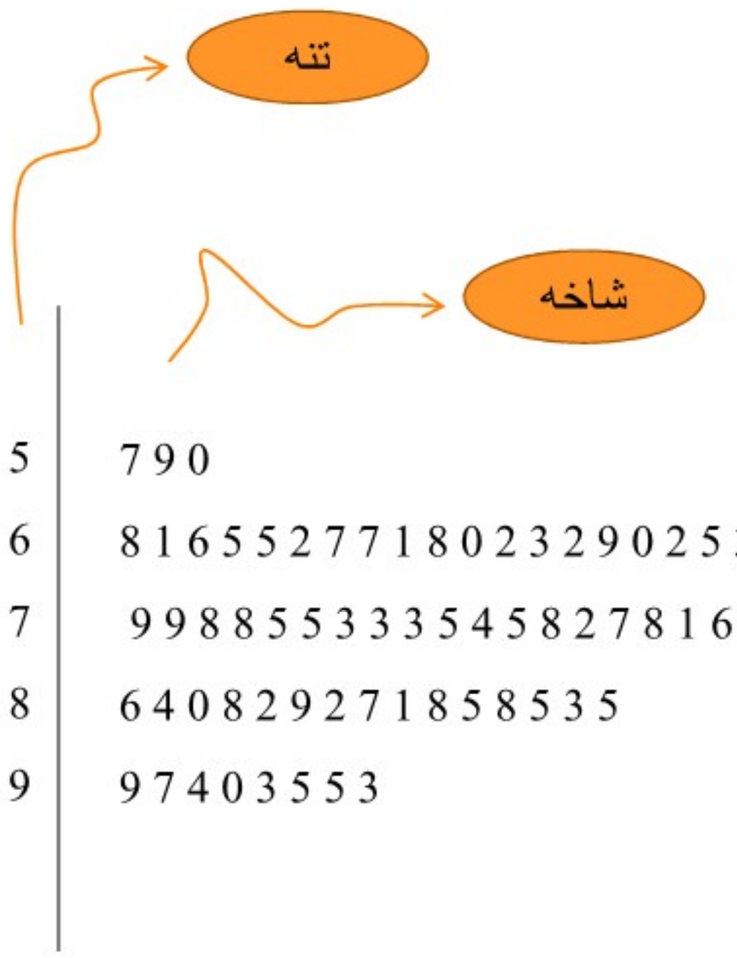
نمودارهای آماری:



نمودارهای آماری:

نمرات ۸۰ دانشجو در امتحانات نهایی درس احتمال و آمار به شرح زیر است:

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
99	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	50	74
86	67	73	81	72	63	76	75	85	77



نمودارهای آماری:

مرتب کرد، پس از ساختن نمودار اولیه معمولا بهتر است مقادیر هر شاخه را از کوچک به بزرگ، با تعداد دفعات تکرار، به صورت زیر:

5	0 7 9
6	0 0 0 1 1 2 2 2 2 3 3 5 5 5 6 7 8 8 9
7	1 1 1 2 2 3 3 3 3 4 4 4 5 5 5 5 5 5 5 6 6 6 6 7 7 8 8 8 8 8 9 9 9
8	0 1 2 2 3 4 5 5 5 6 7 8 8 8 9
9	0 3 3 4 5 5 7 9

معیارهای مرکزی:

با استفاده از جدول فراوانی و رسم نمودارها می‌توانیم داده‌ها را به نحو مطلوبی **تنظیم کرده و اطلاعات نهفته** بهتر است آنها را در یک یا چند **را تا حدودی مشخص کنیم**. با این حال برای ارایه یک گزارش مناسب، ، میانه عدد مناسب نیز خلاصه کنیم. چنین عددی می‌تواند معیار مرکزی باشد. مهمترین معیارهای مرکزی میانگین و نما است که در بخش این به شرح هر یک از آنها خواهیم پرداخت.

هرگاه داده $y_1, y_2, \ldots, y_n$ تعدادهاي $f_1, f_2, \ldots, f_k$ از k نوع $x_1, x_2, \ldots, x_k$ ، با فرض $2 \le k \le n$ ، به ترتیب با تشکیل شده باشند، آنگاه f_i را فراواني x_i مي‌گوييم.

ميانگين حسابي	$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n}$
ميانگين وزني	$\bar{x}_w = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_k}$
ميانگين هندسي **كليه داده‌ها بزرگتر از صفر باشند**	$G = \left(\prod_{i=1}^{n} x_i^{f_i}\right)^{\frac{1}{n}}$

معیارهای مرکزی:

اگر داده‌ها را از کوچک به بزرگ مرتب نماییم،عدد mرا میانه این داده‌ها می‌نامیم،اگر نصف داده‌ها در سمت چپ و نصف داده در سمت راست این عدد قرار گیرد

محاسبه میانه براي داده‌هاي گسسته

فرض کنید $y_1, y_2, \ldots, y_n$ داده‌هاي ما باشند و شکل مرتب شده آنها را با $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ نمایش دهیم آنگاه

$$M = \begin{cases} y_{(\frac{n+1}{2})} & \text{اگر n فرد باشد} \\ \frac{1}{2}[y_{(\frac{n}{2})} + y_{(\frac{n}{2}+1)}] & \text{اگر n زوج باشد} \end{cases}$$

معیارهای مرکزی:

محاسبه میانه براي داده‌هاي پيوسته

کلاس	x_i	f_i	F_i
35/1_55/1	45/1	4	4
55/1_75/1	65/1	6	10
75/1_95/1	85/1	12	22
95/1_15/2	05/2	9	31
15/2_35/2	25/2	8	39
35/2_55/2	2/45	6	45
55/2_75/2	65/2	2	47
75/2_95/2	85/2	3	50
جمع		50	_

$n/2 = 25$

$$m = L_m + \left(\frac{\frac{n}{2} - F_b}{f_m}\right) w$$

طول هر رده

معیارهای مرکزی:

چندك يك معيار كلي‌تر از ميانه است و درعنوان حالت خاص ميانه را نيز در بر مي‌گيرد. اگر pيك عدد حقيقي بين صفر و يك باشد،آنگاه عدد Q_p را چندك مرتبه p مي‌ناميم هر گاه p %100 داده‌ها سمت چپ و (1- p) 100% داده‌ها سمت راست Q_p باشند.

چندكهاي معروف عبارتند از :

چاركها

Q_p چاركها به ازاي 75/0 ، 5/0 ، 25/0 =p به دست مي‌آيند و آنها را به ترتيب با نماد Q_1 (چارك اول)، Q_2 (چارك دوم) و Q_3 (چارك سوم)نشان مي‌دهند.

دهكها

دهكها به ازاي 9/0، .....،2/0 ،1/0=p به دست مي‌آيند و آنها را به ترتيب با نماد D_1 (دهك اول)، D_2 (دهك دوم)،..... و D_9 (دهك نهم) نشان مي‌دهند.

صدكها

صدكها به ازاي 99/0،.....02/0، 01/0=p به دست مي‌آيند و آنها را به ترتيب با نماد P_1(صدك اول)، P_2 (صدك دوم)،.....وP_{99} (صدك نود و نهم) نشان مي‌دهند.

معیارهای مرکزی:

محاسبه چندك براي داده‌هاي گسسته

فرض کنید $y_1, y_2, \ldots, y_n$ داده‌هاي ما باشند و شکل مرتب شده آنها را با $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ نمایش دهیم. براي محاسبه چندك

$$r = (n+1)p \xrightarrow{\text{صحیح باشد}} Q_p = y_{(r)}$$

$$(n+1)p$$

$$r = [(n+1)p], \quad \omega = (n+1)p - r \xrightarrow{\text{صحیح نباشد}} Q_p = (1-\omega)\, y_{(r)} + \omega\, y_{(r+1)}$$

نمودارهای آماری:

محاسبه چندك براي داده‌هاي پيوسته

با توجه به ستون فراواني تجمعي در جدول فراواني، كلاسي را كه چندك در آن قرار دارد مشهص مي‌كنيم.

كلاس	x_i	f_i	F_i
35/1_55/1	45/1	4	4
55/1_75/1	65/1	6	10
75/1_95/1	85/1	12	22
95/1_15/2	05/2	9	31
15/2_35/2	25/2	8	39
35/2_55/2	2/45	6	45
55/2_75/2	65/2	2	47
75/2_95/2	85/2	3	50
جمع		50	_

(p) (n)

$0.25 \times 50 = 12.5$

$$Q_p = L_{Q_p} + \left(\frac{np - F_b}{f_{Q_p}}\right) w$$

$$Q_p = 1.75 + \left(\frac{50 \times 0.25 - 10}{12}\right) \times 0.2$$

نمودارهای آماری:

محاسبه نما براي داده‌هاي گسسته

داده‌ای که فراوانی آن نسبت به دیگر داده‌ها بیشتر باشد، نما یا مد نامیده می‌شود و آن را با نماد Mنمایش می‌دهیم. برای به دست آوردن نما، نخست فراوانی داده‌ها را پیدا می‌کنیم و داده‌ای را که فراوانی آن بیشتر باشد، به عنوان نما اختیار می‌کنیم و اگر دو داده، دارای فراوانی یکسان و بیش از دیگر فراوانی‌ها باشند، هر دو را به عنوان نما اختیار می‌کنیم و داده‌ها را دو نمایی می‌گوییم، به شرط آن که این دو داده در یک صف غیر نزولی،کنار هم نباشند. در صورتی که این دو داده در یک صف غیر نزولی،کنار هم باشند نصف مجموع آنها را به عنوان نما اختیار می‌کنیم. اگر تمام داده دارای فراوانی یکسان باشند،می‌گوییم داده‌ها بدون نما هستند. به یاد داشته باشید که نما، به عنوان یک معیار تمرکز در داده‌های گروهی به کار گرفته می‌شود.

{ین جزوه توسط وب سایت تست دانلود عرضه شده است، testdownload.ir}

نمودارهای آماری:

مثال: براي داده‌هاي 2، 2، 5، 7، 9، 9، 9، 10، 10، 11، 12و 18 نما برابر M=9 است، زيرا فراواني داده 9 بيش از فراواني ديگر داده‌ها است.

مثال: براي داده‌ها 2، 3، 4، 4، 4، 5، 5، 7، 7، 7و 9، دو داده 4 و 7 به عنوان نما اختيار مي‌شوند، زيرا فراواني اين دو داده، بيش از فراواني داده‌هاي ديگر است.

مثال: براي داده‌هاي 3، 5، 8، 10، 12، 15و 16، نما وجود ندارد، زيرا تمام داده‌ها داراي فراواني يكسان هستند.

مثال: براي داده‌ها 2، 3، 4، 4، 4، 5، 5، 5، 7، 7و 9دو داده 4و 5 را كه داراي بيشترين فراواني هستند به عنوان نما بر مي‌گزينيم،اما از آنجا كه اين دو داده در يك صف غير نزولي در كنار يكديگر قرار دادند،نصف مجموف دو داده به عنوان نما اختيار مي‌شود، يعني M=5/4.

نمودارهای آماری:

محاسبه تما براي داده‌هاي پيوسته

کلاس	x_i	f_i	r_i
35/1_55/1	45/1	4	08/0
55/1_75/1	65/1	6	12/0
75/1_95/1	85/1	12	24/0
95/1_15/2	05/2	9	18/0
15/2_35/2	25/2	8	16/0
35/2_55/2	2/45	6	12/0
55/2_75/2	65/2	2	04/0
75/2_95/2	85/2	3	06/0
جمع		50	00/1

از روي جدول ملاحظه مي‌شود كه فراواني رده 75/1_95/1 داراي بيشترين فراواني است بنابراين به عنوان رده نما در نظر مي‌گيريم.

$$M = L_M + \frac{d_1}{d_1 + d_2} \times \omega$$

24/0-12/0

24/0-18/0

$$M = 1.75 + \frac{0.12}{0.12 + 0.06} \times 0.2$$

معیارهای پراکندگی:

با وجود این که در بسیاری از موارد، میانگین توصیف نسبتا کاملی از مجموعه داده‌ها ارائه می‌دهد، اما گاهی وجود اطلاعات بیشتر در مورد داده‌ها ضروری است. یک مفهوم مهم در ارتباط با داده‌های آماری، میزان تغییرات آنهاست،بدین معنی که اندازه‌گیریها تا چه اندازه از فردی به فرد دیگر یا شیی به شیی دیگر تغییر می‌کنند. در این بخش، به بررسی و محاسبه میزان تغیرات به عنوان معیارهای پراکندگی خواهیم پرداخت. مهمترین معیارهای پراکندگی عبارتند از **دامنه، میانگین انحراف ها از میانگین یا از میانه، میان دامنه چارکها، دامنه صدکی، واریانس و انحراف معیار است. علاوه بر مطالب فوق، در این بخش داده‌های استاندارد و ضریب تغیرات را نیز معرفی خواهیم کرد.**

{ین جزوه توسط وب سایت تست دانلود عرضه شده است، testdownload.ir}

معیارهاي پراکندگي:

اگرچه دامنه یک وسیله ساده برای اندازه‌گیری اختلاف و پراکندگی در یک سری از داده‌ها است، اما در بیشتر موارد رضایتبخش نیست. داده‌های بسیار بزرگ یا بسیار کوچک مانع از آنند که دامنه، معرف واقعی میزان انحراف باشد. در چنین مواردی، واریانس یک معیار مورد قبول همگان به

با توجه به اینکه در محاسبه واریانس داده‌ها را مربع می‌کنیم، بدین جهت ریشه دوم مثبت آن را که انحراف معیار یا انحراف استاندارد می‌نامیم، به عنوان یک معیار پراکندگی بر مبنای مقیاس اندازه‌گیری به کار می‌بریم.

ضریب تغییر عبارت است از اندازه نسبی انحراف معیار در مقایسه با میانگین. ضریب همبستگی به واحد اندازه‌گیری وابسته نیست و برای مقایسه جمعیتهای یکسان به کار می‌رود. در مقایسه هر اندازه که ضریب تغییر ویژگی جمعیتی کمتر باشد،ویژگی آن جمعیت بهتر ارزیابی می‌شود.

دامنه	$y_{(1)}$ کوچکتر $y_{(n)}$
میانگین انحرافها از میانگین	$n \sum_{i=1}$
میانگین انحرافها از م	$MD_m = \sum_{i=1}^{k} f_i \lvert x_i$
	$IQR = Q_3 - Q_1$
	$IPR = P_{90} - P_{10}$
	$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2$
انحراف معیار	$s = \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}$
ضریب تغییرات	$v = \frac{s}{\bar{x}}$

معیارهای پراکندگی:

اگر $y_1, y_2, \ldots, y_n$ نمایانگر داده‌هاي خام باشند، براساس جدول فراواني f_1 تا از y_i ها برابر x_1 ، f_2 تا برابر x_2 ،.....و f_k تا برابر x_k است. مي‌دانيم كه $\bar{x}$ و s به ترتيب ميانگين و انحراف معيار داده است. اگر از هر داده $\bar{x}$ را كم و بر s تقسيم كنيم ، يعني

داده‌هاي استاندارد

$$z_i = \frac{x_i - \bar{x}}{s} \quad , \quad i=1,\ldots,k$$

آنگاه $z_1, z_2, \ldots, z_k$ با فراواني‌هاي به ترتيب $f_1, f_2, \ldots, f_k$ را داده‌هاي استاندارد مي‌ناميم. به سادگي مي‌توان نشان داد كه داده‌هاي استاندارد داراي ميانگين برابر با صفر و واريانس برابر با يك 1 هستند و به واحد اندازه‌گيري بستگي ندارند.

چون $V_1 > V_2$ ، بنابراين نتيجه مي‌گيريم كه دانشجويان در امتحان دوم نمرات مطلوبتري را كسب كرده‌اند.

ب) براي مقاسه، ابتدا نمرات دانشجو را استاندارد مي‌كنيم

$$z_1 = \frac{65-60}{6} = \frac{5}{6}, \qquad z_1 \quad \frac{720-700}{7} = \frac{20}{7}$$

چون $z_1 < z_2$ ، بنابراين نمره آزمون دوم دانشجو د رمقايسه از موقعيت بهتري برخوردار است.

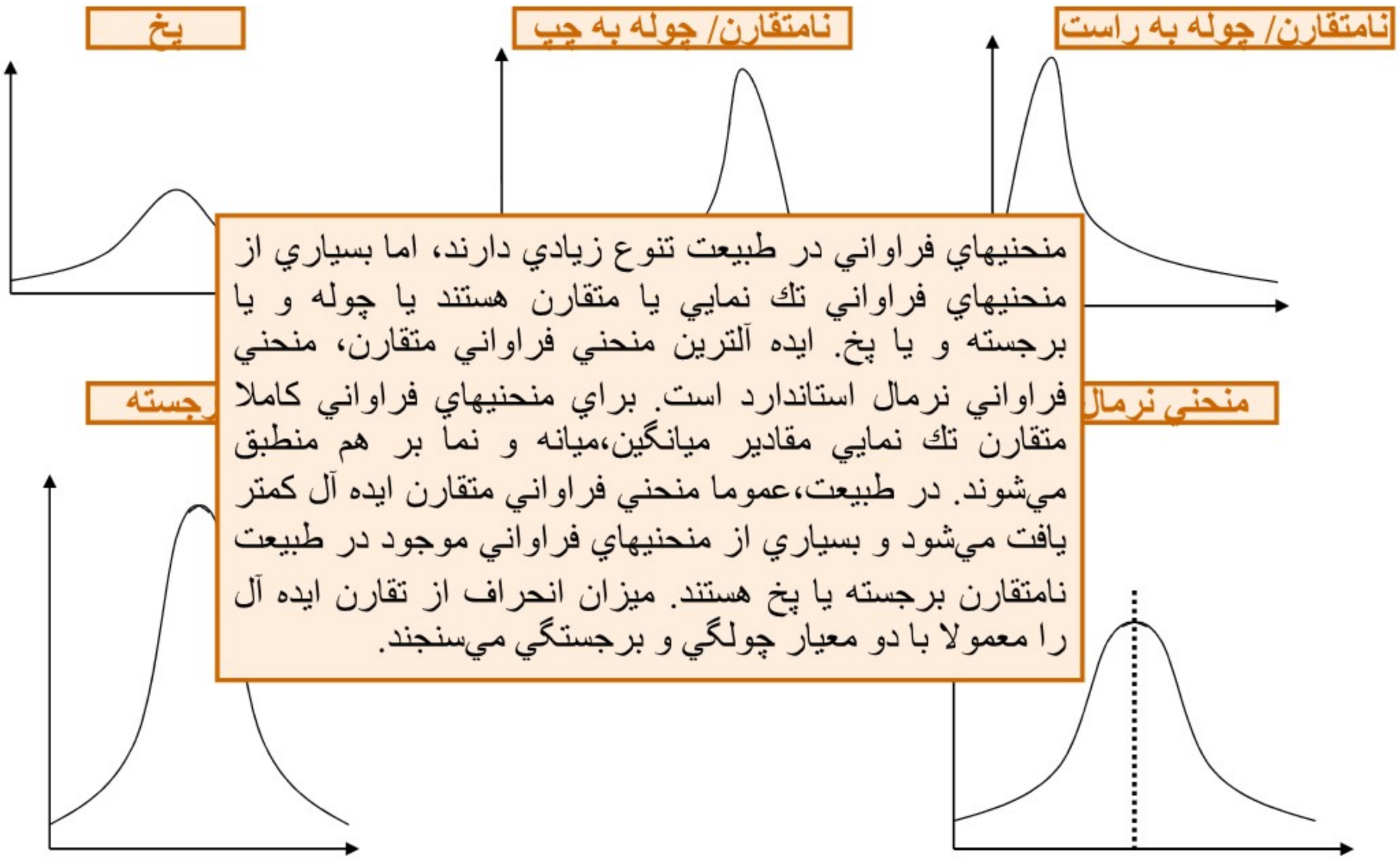
معیارهای پراکندگی

- فرض کنید یک دسته از دانشجویان در دو امتحان شرکت کرده‌اند و خلاصه نتاسج آزمونها به شرح زیر است.
- آزمون اول: میانگین نمرات برابر ۶۰، انحراف معیار برابر ۶، ماکزیمم نمره از ۱۰۰
- آنمون د.م: میانگین نمرات ۷۰۰، انحراف معیار برابر ۷، ماکزیمم نمره از ۱۰۰۰
- الف- چگونه این دو نتیجه را با هم مقیسه و ارزیابی می کنید؟
- ب- اگر دانشجویی در آزمون اول نمره ۶۵ و در آزمون دوم نمره ۷۲۰ را کسب کرده باشد، وضعیت دانشجو در کدام آزمون مطلوبتر است؟
- حل: الف) با محاسبه ضریب تغییر دو آزمون معلوم می‌شود که

$$v_1 = \frac{s_1}{x_1} = \frac{6}{60} = \frac{1}{10} = \%10$$

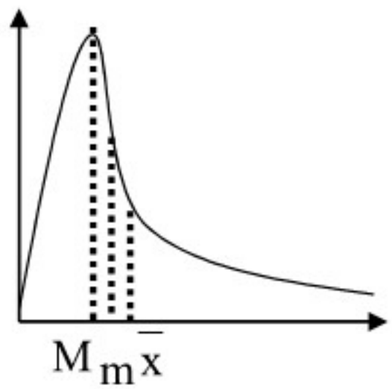
$$v_2 = \frac{s_1}{x_2} = \frac{7}{700} = \frac{1}{100} = \%1$$

منحنیهای فراوانی:



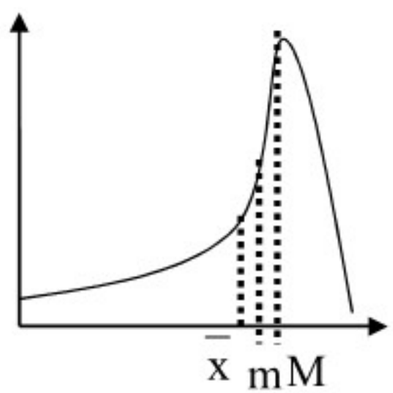
منحنيهاي فراواني در طبيعت تنوع زيادي دارند، اما بسياري از منحنيهاي فراواني تك نمايي يا متقارن هستند يا چوله و يا برجسته و يا پخ. ايده آلترين منحني فراواني متقارن، منحني فراواني نرمال استاندارد است. براي منحنيهاي فراواني كاملا متقارن تك نمايي مقادير ميانگين،ميانه و نما بر هم منطبق مي‌شوند. در طبيعت،عموما منحني فراواني متقارن ايده آل كمتر يافت مي‌شود و بسياري از منحنيهاي فراواني موجود در طبيعت نامتقارن برجسته يا پخ هستند. ميزان انحراف از تقارن ايده آل را معمولا با دو معيار چولگي و برجستگي مي‌سنجند.

منحنیهای فراوانی:



$M < m < \bar{x}$ به راست یا مثبت

چولگي:



$M > m > \bar{x}$ به چپ یا منفي

معیار اندازه گیری چولگی:

$$\text{ضریب چولگی اول پیرسن} = \frac{\bar{x} - M}{s}$$

منحنیهای فراوانی:

میزان کشیدگی یا پخی منحنی فراوانی را یسبت به منحنی نرمال استاندارد، برجستگی منحنی فراوانی می نامیم. فرمول زیر را می توان به عنوان معیار برجستگی به کار برد.

$$\text{ضریب برحستگی صدکی} = k = \frac{(Q_3 - Q_1)/2}{P_{90} - P_{10}}$$

نشان داده شده است که برای منحنی فراوانی نرمال استاندارد $k=0.263$ ، بنابراین معمولا ضریب برجستگی را به صورت زیر تعریف می کنند:

$$\text{ضریب برحستگی صدکی} = k = \frac{(Q_3 - Q_1)/2}{P_{90} - P_{10}} - 0.263$$

برحسب آن که این مقدار مثبت یا منفی باشد گوییم منحنی فراوانی برجسته یا پخ است.

نمودار جعبه ای:

همان گونه که گفته شد، روشهای نموداری و خلاصه کردن داده ها به صورت مقادیر عددی موضوعی اساسی در تجزیه و تحلیلهای آماری است. پیش از این دیدیم که چگونه نمایش نمودار تنه و شاخه را می توان به عنوان ابزاری ساده و مهم در نمایش و استنباط از داده ها به کار گیریم که چنین نموداری بسیار همانند نمودار هستوگرام بود.

نموار با ارزش دیگری که برخی از امتیازهای دیگر را در مقایسه با نمودار تنه و شاخه دارد نمودار جعبه ای است که برخی از امتیازهای دیگر را در مقایسه با نمودار تنه وشاخه دارد .

نمایش نمودار جعبه ای بر پایه داده های مرتب شده از کوچک به بزرگ و تعیین میانه، چارک اول و چارک سوم است.

نمودار جعبه ای:

گام اول نمایش نمودار تنه و شاخه،

9	0 2
10	0 0 9
11	0 5 6 6 7 7 8 9
12	5 9 9 9 9 9
13	0 0
14	0
15	0 4

گام دوم تعیین مکان میان، چارک اول و چارک سوم،

مکان میانه = $\frac{n+1}{2} = \frac{24+1}{2} = 12.5$

با توجه به مقدار به دست امده میانگین داده های دوازدهم و سیزدهم را به عنوان میانه در نظر می گیریم، یعنی

$$m = \frac{118+119}{2} = 118.5$$

نمودار جعبه ای:

برای تعیین مکان چارک اول و سوم به صورت زیر عمل می کنیم

مکان چارکها $= \frac{\text{مکان میانه} + 1}{2} = \frac{125 + 1}{2} = \frac{12 + 1}{2} \quad 6.5$

با توجه به مقدار به دست آمده، میانگین داده های ششم و هفتم از پایین به بالا را به ترتیب به عنوان چارک اول و چارک سوم در نظر می گیریم، یعنی

چارک اول $Q_1 = \frac{110+115}{2} = 112.5$

چارک سوم $Q_2 = \frac{125+129}{2} = 127$

در صورتی که مقادیر به دست آمده در مکانها اعداد صحیح باشند، داده همان مرتبه به عنوان میانه، چارک اول و چارک سوم در نظر گرفته می شود.

{ین جزوه توسط وب سایت تست دانلود عرضه شده است، testdownload.ir}

میانه و چارکها به دست آمده در ارایه نمایشی برای نمودار جعبه ای با روش بیان شده در بخش های قبل فرق دارد. در حقیقت معیارهای به دست آ»ده از این روش را هینج می نامند که کمی با معیارهای گفته شده متفاوت است.

نمودار جعبه ای:

گام سوم **تعیین دو فاصله به عنوان حصارهای درونی و بیرونی است،**

نخست دامنه چارکها را محاسبه می کنیم ،

$$IQR = Q_3 - Q_1 = 125 - 112.5 = 12.5$$

کرانهای حصار درونی را به صورت زیر تعریف می کنیم:

کران پایین حصار درونی

$$LIF = Q_1 - 1.5\,IQR$$

کران بالای حصار درونی

$$UIF = Q_3 + 1.5\,IQR$$

بنابراین حصار درونی در این مثال به صورت زیر تعریف می شود

$$(LIF, UIF) = (93.75, 145.75)$$

نمودار جعبه ای:

کرانهای حصار بیرونی را به صورت زیر تعریف می کنیم:

$$LOF = Q - 3\,IQR$$
$$UOF = Q + 3\,IQR$$

در نتیجه فاصله زیر، حصار بیرونی در این مثال است،

$$LOF \quad UOF \quad \ni \quad (75,164.5)$$

گام چهارم

تعیین مقادیری از داده ها که در همسایگی کرانهای حصار درونی است.

در حقیقت مقادیر این داده ها در حصار درونی قرار دارد و کمینه بیشینه مقدار ممکن از داده ها در حصار درونی است که نزدیک به کران بالا و پایین حصار درونی است. همسایگی کران پایین را با نماد LA و همسایگی کران بالا را با نماد UA نمایش می دهیم. بنابراین در این مثال،

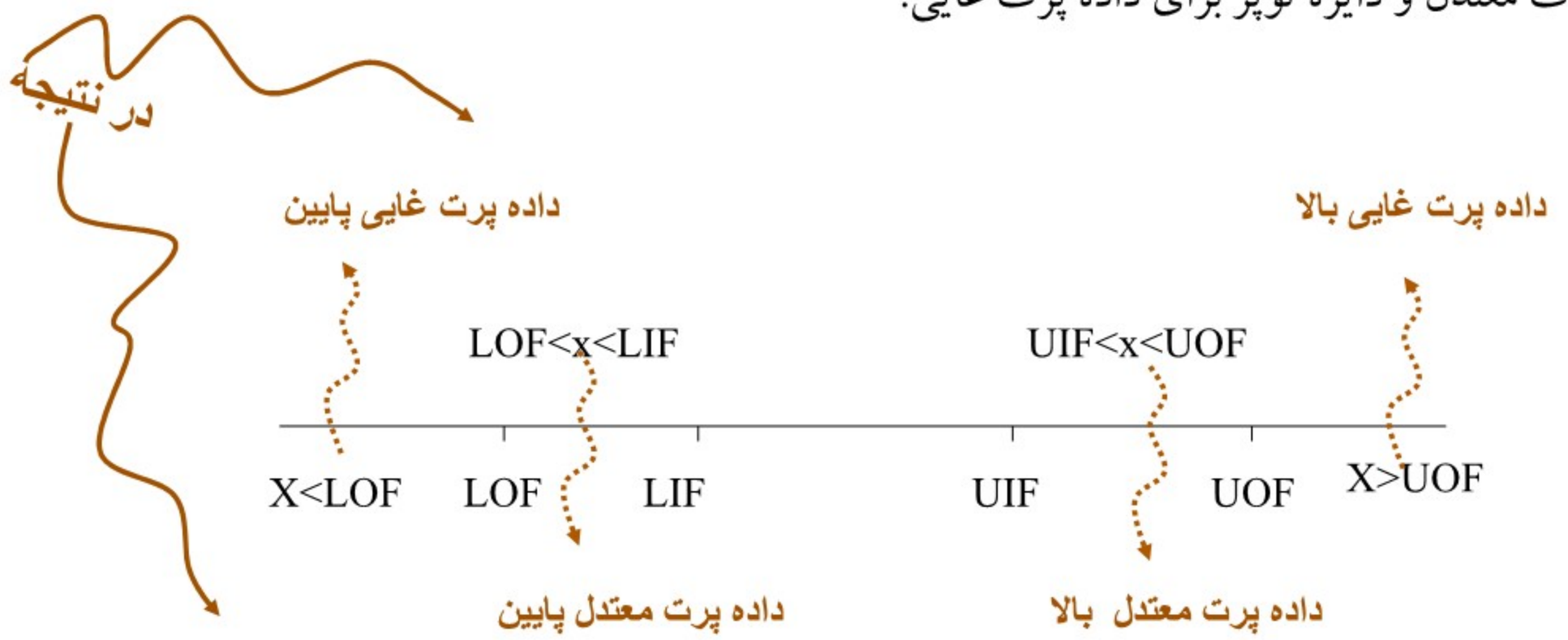
$$LA = 100$$
$$UA = 140$$

نمودار جعبه ای:

گام پنجم **تعیین داده های پرت،**

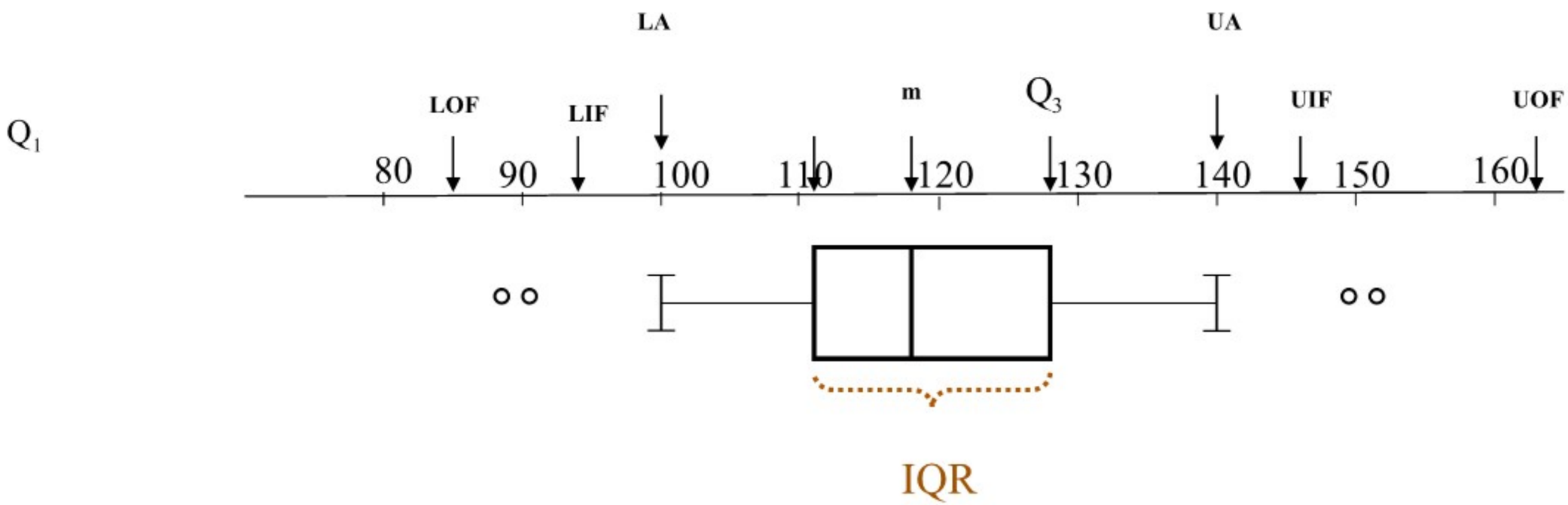
هر داده بیرون از حصار درونی را داده پرت می نامیم. در صورتی که این داده ها بیرون از حصار بیرونی نباشد، آن را داده پرت معتدل و به جز این صورت آن را داده پرت غایی می نامیم.

نمادی که برای نمایش داده های پرت در نمودار جعبه ای به کار خواهیم برد، دایره توخالی برای داد پرت معتدل و دایره توپر برای داده پرت غایی.



نمودار جعبه ای:

بنابراین در این مثال داده های پرت عبارتند از: ۱۵۴، ۱۵۰، ۹۲و ۹۰ و داده پرت غایی نداریم.



انواع ضریب همبستگی

تعریف همبستگی

یکی از تعاریف اساسی در علم آمار تعریف همبستگی و رابطه بین دو متغیر می باشد. بطور کلی شدت وابستگی دو متغیر به یکدیگر را همبستگی تعریف می کنیم. و ممکن علاوه بر شدت همبستگی جهت همبستگی نیز مورد نیاز پژوهشگر باشد. در آمار انواع زیادی از ضرایب همبستگی متفاوت وجود دارند که هر کدام همبستگی بین دو متغیر را با توجه به نوع داده‌ها و شرایط متغیرها اندازه‌گیری می‌کنند.

بطور کلی ضرایب همبستگی بین ۱- تا ۱ تغییر می کنند و رابطه بین دو متغیر می تواند مثبت یا منفی باشد مانند میزان درآمد و قیمت ماشین شخص یا میزان تسلط ماشین نویس و تعداد غلط های تایپی او در یک صفحه.

نکات قابل توجه‌ای در ضریب همبستگی وجود دارد که ذکر آنها مهم می باشد. ضریب همبستگی یک رابطه متقارن می‌باشد، هر چه ضریب همبستگی به یک نزدیکتر باشد میزان وابستگی دو متغیر بیشتر است، اما توجه کنید که این وابستگی به معنای رابطه علت و معلولی نیست وضریب همبستگی حرفی از اینکه کدام علت و کدام معلول است به میان نمی آورد. اما اگر متغیر های دیگری نیز بر روی متغیر وابسته تاثیر داشته باشد آنگاه ممکن است هر کوواریانسی که با متغیر مستقل به اشتراک گذاشته‌اند، تاثیر غلطی را بر ضریب همبستگی با متغیر مستقل داسته باشد. همچنین برای تعمیم ضریب همبستگی می توان وجود رابطه غیرخطی بین دو متغیر همبسته را در حالی که ضریب همبستگی به غلط آن را نشان نمی‌دهد، بررسی کرد. ضریب همبستگی را می‌توان برای سنجش میزان خطای موجود در داده‌ها نیز استفاده کرد، از جمله زیر فاصله‌ها و یا برش‌های مصنوعی دامنه داده‌ها را

نام برد. در صورتی که رابطه بین داده ها با افزایش متغیر افزایش یابد ضریب همبستگی می‌تواند به عنوان معیار عدم هم واریانسی استفاده شود.

انواع ضرایب همبستگی

محاسبه ضرایب همبستگی تا حدود زیادی متاثر از مقیاس اندازه گیری متغیر ها است، بعنوان مثال برای متغیرهای اسمی جهت رابطه اصلا معنی ندارد، بین جنس و معدل تنها می‌توان گفت که شدت وابستگی چه مقدار است اما افزایش یا کاهش جنس معنی ندارد.

با توجه به نوع متغیر ها ضریب همبستگی می‌تواند یکی از حالت‌های زیر را داشته باشد.

۱- دو متغیر اسمی

۲- دو متغیر رتبه‌ای

۳- دو متغیر فاصله‌ای- نسبی

۴- متغیر اسمی و متغیر رتبه ای

۵- متغیر اسمی و متغیر فاصله‌ای - نسبی

۶- متغیر رتبه‌ای و متغیر فاصله‌ای – نسبی

۱- حالتی که دو متغیر اسمی و یا یکی اسمی و دیگری رتبه‌ای باشد

أ. ضریب همبستگی کرامر

ب. ضریب توافقی C

ج. ضریب همبستگی لاندا

د. ضریب همبستگی تاو گودمن و کروسکال

أ. ضریب همبستگی کرامر و فی

این نوع ضریب همبستگی بیشتر در جداول توافقی و بررسی میزان رابطه بین دو متغیر اسمی بکار می‌رود. برای مثال می خواهیم بررسی کنیم که آیا بین جنس و گرایشات سیاسی رابطه وجود دارد یا خیر. ضریب همبستگی کرامر با استفاده از فرمول زیرمحاسبه می‌شود:

$$v = \sqrt{\frac{\chi^2}{n(l-1)}}$$

که در آن χ^2 آماره آزمون خی دو ، n تعداد حجم نمونه، l تعداد ردیف‌های جدول یا ستون‌های آن(هر کدام که تعداد کمتری دارند). البته این فرمول برای حالتی که $l = 2$ باشد به ضریب همبستگی فی معروف است.

ب. ضریب توافقی C

یکی دیگر از معیارهای همبستگی برای دو متغیر با مقیاس اسمی یا یکی اسمی و یکی رتبه‌ای ضریب توافقی C می‌باشد. این شاخص با استفاده از فرمول زیر محاسبه می‌شود:

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

در این رابطه χ^2 آماره آزمون و n حجم نمونه است.

ج. ضریب همبستگی لاندا

یکی دیگر از ضرایب همبستگی که برای دو متغیر اسمی و یا یکی اسمی و یکی رتبه‌ای استفاده می‌شود، ضریب همبستگی لاندا می‌باشد. ضریب همبستگی لاندا تصویر روشن تری را از رابطه بین دو متغیر را نشان می‌دهد. ضریب همبستگی لاندا را با استفاده از فرمول زیر محاسبه می‌کنند:

$$\lambda = \frac{e_1 - e_2}{e1}$$

که در آن e_1 اشتباه گروه بندی در موقیتی که متغیر اول متغیر مستقل باشد و e_2 اشتباه گروه بندی در موقعیتی که متغیر دوم مستقل است می‌باشد.

د. ضریب همبستگی تاو گودمن کروسکال:

این نوع ضریب همبستگی نیز برای ارزیابی شدت رابطه بین متغیرهایی که هر دو اسمی یا یکی اسمی و دیگری رتبه‌ای باشد بکار می‌رود.

۲- ضریب همبستگی در حالتی که دو متغیر دارای مقیاس رتبه‌ای باشند

برای سنجش رابطه بین متغیرهای رتبه‌ای از ضرایب همبستگی زیر استفاده می‌شود:

الف- ضریب همبستگی گاما

ب- ضریب همبستگی تاو کندال b

ج- ضریب تاو کندال c

د- ضریب d سامرز

سعی می‌شود مفهمو هر کدام از ضرایب همبستگی بالا با ارائه مثال واضح و روشن شود.

در ابتدا به مثالی جالب که معنای رابطه را در این حالت مشخص می کند توجه کنید:

فرضیه "بین تحصیلات کارکنان و مسئولیت پذیری آنها رابطه مثبت و معنی داری وجود دارد" را در نظر بگیرید. این فرضیه دو متغیر دارد، تحصیلات کارکنان(با مقوله‌های دیپلم و کمتر، فوق دیپلم، لیسانس و بالاتر) و مسئولیت‌پذیری(با مقوله‌های کم، متوسط، زیاد). رخلاف متغیر های اسمی که جهت رابطه در آنها مفهوم پیدا نمی‌کرد در این متغیر ها بنا به ماهیتشان جهت رابطه مفهوم دارد. لذا قبل از هرچیز بایستی بررسی کرد که رابطه در اینحالت به چه معنی است.

فرض کنید نمره ۶ دانشجو را در دو درس داریم، پس هر فرد دو نمره دارد. حال چنانچه نمره یک فرد با فرد دیگر مقایسه شود، می‌توان بیان داشت که این دو فرد(دو زوج) نسبت به هم یک زوج معکوس را تشکیل می‌دهند یا یک زوج هماهنگ.

فرد	نمره درس ۱	نمره درس ۲
۱	۱۵	۱۰
۲	۱۶	۱۰
۳	۱۶	۱۲
۴	۱۵	۱۱
۵	۱۴	۱۲
۶	۱۵	۱۰

اگر با افزایش یک متغیر، متغیر دیگر افزایش یابد به آن زوج هماهنگ(۱) گویند و اگر با افزایش یکی دیگری کاهش یابد زوج معکوس(۲) گویند و چنانچه با افزایش یا کاهش یک متغیر متغیر دیگر تغییری نکند به آن دو زوج گره خورده گویند(به روابط فرد سوم و اول، پنجم و اول، دوم و اول در بالا توجه کنید). تعداد کل زوج‌های محاسبه شده در این حالت برابر n(n-1)/2 می باشد. بنابراین محاسبه این زوج ها را در زمانی که متغیر ها زیاد و جداول پیچیدگی زیادی دارند مشکل می‌باشد.

حال با توضیح بالا می‌توان به بررسی ضرایب همبستگی در حالتی که هر دو متغیر رتبه‌ای باشند پرداخت.

الف. ضریب همبستگی گاما

این ضریب همبستگی از تعامل زوج‌های هماهنگ و معکوس بدست می‌آید. واضح است چنانچه زوج‌های هماهنگ بیشتر از زوج‌های معکوس باشد، رابطه بین دو متغیر مورد اشاره مثبت است و چنانچه زوج‌های معکوس و هماهنگ با همدیگر برابر باشد، هیچ رابطه‌ای بین آن دو وجود ندارد.

ضریب همبستگی گاما از رابطه زیر محاسبه می‌شود:

$$\gamma = \frac{n_s - n_d}{n_s + n_d}$$

که در آن n_s تعداد زوج‌های هماهنگ، n_d تعداد زوج‌های معکوس می‌باشد.

ب. ضریب همبستگی تاو کندال b

همانطو که مشاهده کردید ضریب همبستگی گاما زوج‌های گره خورده را نادیده می‌گیرد، اما در ضریب همبستگی تاو کندال زوج‌های گره خورده نیز در محاسبه آن دخالت دارند. ضریب همبستگی تاو کندال b بصورت زیر محاسبه می‌شود:

$$T_b = \frac{n_s - n_d}{(n_s + n_d + t_y).(n_s + n_d + t_x)}$$

n_s : تعداد زوج‌های هماهنگ

n_d : تعداد زوج‌های معکوس

t_y : تعداد زوج‌هایی که در متغیر y گره خورده‌اند

t_x : تعداد زوج‌هایی که در متغیر x گره خورده‌اند

ج. ضریب همبستگی تاو کندال C

این ضریب مبتنی بر تعداد زوج‌های معکوس و هماهنگ عمل می‌کند و از رابطه زیر بدست می‌آید:

$$T_c = \frac{2m(n_s - n_d)}{n^2(m-1)}$$

که در آن n_s تعداد زوج های هماهنگ، n_d تعداد زوج‌های معکوس، m کوچکترین سطر یا ستون و n تعداد افراد نمونه است.

د. ضریب همبستگی d سامرز

ضریب همبستگی d سامرز رابطه زیادی با ضریب همبستگی گاما دارد. ضرایب همبستگی گاما، تاو-کندال b و تاو کندال c همگی قرینه هستند و در صورتی که در محاسبه آنها متغیر مستقل و وابسته معلوم باشد، تغییری ایجاد نمی‌شود و ضریب همبستگی ثابت خواهد بود.اما ضریب همبستگی d سامرز شکل خاصی از ضریب همبستگی گاما است که یکی از متغیرها به عنوان متغیر وابسته در نظر گرفته می‌شود. تنها تفاوت آن در مخرج کسر است که علاوه بر زوج‌های هماهنگ و معکوس زوج‌های گره-خورده در متغیر مستقل نیز در مخرج اضافه می‌شود.

$$d = \frac{n_s - n_d}{n_s + n_d + t_y}$$

۳. ضریب همبستگی در حالتی که دو متغیر فاصله‌ای نسبی باشد

در این حالت هر دو متغیر کمیت پذیرند. در این قسمت به بررسی دو ضریب همبستگی معروف پیرسون در حالت پارامتریک و اسپیرمن در حالت ناپارامتریک می‌پردازیم:

الف. ضریب همبستگی پیرسون

این ضریب همبستگی مبتنی بر کوواریانس دو متغیر و انحراف معیار های آنها می‌باشد که می‌توان از برآوردهای آنها برای محاسبه ضریب همبستگی پیرسون استفاده کرد.

$$r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

ب. ضریب همبستگی اسپیرمن

این ضریب همبستگی معادل ضریب همبستگی پیرسون در داده‌های کمی می‌باشد و مبتنی بر رتبه داده-ها محاسبه می‌شود. اما توجه کنید که شرایطی برای محاسبه این ضریب همبستگی نیاز است که در صورت رعایت نکردن ممکن است انحرافات اساسی و جدی را پدید آورد. ازجمله این شرایط باید گفت که هر دو متغیر بایستی مقوله‌های زیادی داشته باشند رتبه آنها مفهوم پیدا کند

۴. متغیرهایی با مقیاس اسمی– رتبه‌ای و متغیرهای با مقیاس فاصله‌ای–نسبی

هنگامی که یک متغیر دارای مقیاس اسمی و رتبه‌ای (مثل جنس، نژاد، میزان رضایت و ...) و متغیر دیگر مقیاس فاصله‌ای یا نسبی داشته باشد مانند درآمد، معدل، اندازه قد و ... آنگاه بایستی شاخصی انتخاب شود که از روی یک متغیر بتوان متغیر دیگر را پیش بینی کرد. از جمله این شاخص‌ها شاخص نسبت همبستگی است که آن را با نماد η^2 نشان می‌دهند.

الف. ضریب همبستگی η^2 (مجذور اتا)

برای فهم بهتر این ضریب همیستگی به فرضیه " آیا بین جنس و معدل داتشجویان رابطه وجود دارد" توجه کنید. در اینجا جنس یک متغیر کیفی یا اسمی است و معدل یک مقیاس فاصله‌ای یا نسبی است(

ب. ضریب همبستگی چند رشته‌ای

همبستگی چند رشته‌ای که توسط اولسون و سایرین(۱۹۸۲) معرفی شد، زمانی مورد استفاده قرار می‌گیرد که میخواهیم همبستگی بین یک متغیر فاصله‌ای را با متغیر دو حالتی یا ترتیبی(که فرض شده است که متغیر اساسا پیوسته ای را منعکس می‌کند) بررسی کنیم. این ضریب همبستگی ر می‌توان تا حدود زیادی مانند ضریب همبستگی پیرسون تعبیر کنیم، در آماره χ^2 آن فرض نرمال دو متغیره بودن داده ها لازم است.

۵. سایر ضرایب همبستگی

در این قسمت به معرفی ضریب همبستگی کاپای کوهن می‌پردازیم.

الف. ضریب همبستگی کاپای کوهن

فرض کنید محققی درصدد باشد که میزان توافق بین مدیران و معاونان را در خصوص میزان اهمیت مشتری در سازمان ارزیابی کند. هر فرد(اعم از مدیر یا معاون) می‌تواند نظر خود را بصورت زیاد، و کم بیان کند. در این حالت که هر دو متغیر اسمی دوتایی می‌باشند می‌توان از ضریب همبستگی کاپای کوهن که بطور معادل در بعضی مواقع ضریب همبستگی کاپا نیز نامیده می‌شود، استفاده می‌شود.

ب. ضریب همبستگی چند حالتی

ضریب همبستگی چند حالتی زمانی مورد استفاده قرار می‌گیرد که هر دو متغیر دو حالتی یا هر دو متغیر ترتیبی باشند، البته توجه کنید که مانند ضریب همبستگی چند رشته‌ای در هر دو متغیر فرض شده است که این متغیر ها اساسا تغییرات متغیر پیوسته ای را منعکس می‌کنند، علی الخصوص زمانی که از مقیاس لیکرت استفاده می‌شود بایستی از این ضریب همبستگی استفاده کرد.

☜ علم آمار

آمار :

روش و چگونگی جمع آوری اطلاعات و بیان آنها در قالب عدد

احتمال :

تجزیه و تحلیل اطلاعات عددی جمع آوری شده

استنباط :

قضاوت درباره کل اطلاعات وقتی بخشی از آن رؤیت شده

جمعیت :

مجموعه تمام عناصری که دارای یک یا چند ویژگی مشترک باشند

نمونه:

بخشی از جمعیت میباشد

متغیر:

یک ویژگی که از فردی به فرد دیگر در یک جمعیت (نمونه) تغییر میکند

۱- متغیر کمی (وزن یا قد)

۲- متغیر توصیفی (مهارت یا هوش)

مقیاس سازی

۱- مقیاس اسمی (تهران ۰۲۱ - شیراز ۰۷۱۱)

۲- مقیاس ترتیبی (ضعیف ۰ – متوسط ۱ – خوب ۲)

۳- مقیاس فاصله ای (x=at+b که b<>0 مثلاً f=9/5(c+32)) در این مقیاس صفر بمعنی هیچ نیست

۴- مقیاس نسبتی (x=at) صفر بمعنی هیچ است این مقیاس خوبی است

داده :

۱- داده گسسته (تعداد فرزند)

۲- داده پیوسته (طول قد)

مشخص کننده مرکزی

۱- میانگین (حسابی – وزنی – هندسی – توافقی و ...)

۲- میانه (وسط صف منظم داده ها)

۳- نما (داده ای که بیشتر تکرار شده)

۴- میانگین و واریانس (میانگین و میزان انحراف داده ها از میانگین)

مشخصه های پراکندگی :

۱- دامنه تغییرات (تفاضل بزرگترین داده از کوچکترین داده)

۲- انحراف متوسط (میانگین قدر مطلق انحراف ها از میانگین داده ها)

۳- انحراف معیار (جذر واریانس که واریانس میانگین توان دوم انحراف ها از میانگین داده ها)

فراوانی :

اگر n شیء متمایز در دسته های مختلفی (n1, n2 , ..) را به تعداد مختلف (w1,w2,..) داشته باشیم w ها را فراوانی گویند

۱- فراوانی نسبی نسبت فراوانی هر مجموعه به کل تعداد آن مجموعه ri=wi/n

۲- فراوانی انباشته (تجمعی) حاصل جمع فراوانی تا موقعیت مورد نظر Fj

۳- فراوانی انباشته نسبی : حاصل جمع Wi ها تقسیم بر n

میانگین

$$\overline{X} = \frac{1}{n}\sum_{h=1}^{n} x_h$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

انحراف معیار = σ

$$v(x) = \sigma^2 = \frac{1}{n}(\sum (x_i - \mu)^2$$

یا

$$\sigma^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n}$$

در حالت داشتن فراوانی

$$\mu = \frac{1}{\sum f_i}\sum_{i=1}^{n} x_i f_i$$

$$v(x) = \sigma^2 = \frac{1}{\sum f_i}(\sum f_i(x_i - \mu)^2$$

ثابت شده است که در طبیعت آمار بصورت نرمال توزیع شده است که معمولاً ۶۸٪ داده ها بین $\mu \pm \sigma$ و ۹۶٪ داده ها بین $\mu \pm 2\sigma$ خواهد بود

۱) مثال

برای داده های xi و فراوانی fi میانگین و واریانس را حساب کنید آیا این داده ها نرمال هستند

	xi	*fi*	*xifi*	*ri*	*Fi*		
	10.4	*1*	*10.4*	*0.02*	*1*		
	11.3	*6*	*67.8*	*0.12*	*7*		
	12.2	*7*	*85.4*	*0.14*	*14*		
	13.1	*12*	*157.2*	*0.24*	*26*		
	14	*12*	*168*	*0.24*	*38*		
	14.9	*9*	*134.1*	*0.18*	*47*		
	15.8	*3*	*47.4*	*0.06*	*50*		
جمع		*50*	*670.3*				
			13.406			***1.80547***	
			میانگین			*واریانس نمونه*	

$s^2 = 1.805$

$s = 1.34$

$\mu \pm \sigma = 12.06 \quad 14.74$

$\mu \pm 2\sigma = 10.71 \quad 16.09$

$\mu \pm 3\sigma = 09.37 \quad 16.43$

داده بین اعداد	12.06-14.74	10.71-16.09	09.37-16.43
تعداد	31	49	50
درصد	62	98	100

رگرسیون و همبستگی خطی

n تا زوج مرتب داریم بهترین رابطه خطی بین دو متغیر زوج پیدا کنیم

$y = a + bx$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \quad , \quad a = \bar{y} - b\bar{x} \quad \Rightarrow y = a + bx$$

در رگرسیون باید ببینیم متغیر مستقل کدام است مثلاً اگر جدولی از طول قامت پدران و پسرانشان داشته باشیم متغیر مستقل پدر است زیرا بتبع قامت پدر قامت پسر تغییر میکند

نکته مهم : در رابطه فوق در مخرج باید متغیر مستقل باشد

مثال :

معادله رگرسیون بر حسب y برای زوجهای مرتب زیر بدست آورید

x = 1,3,4,6,8,9,11,14
y = 1,2,4,4,5,7,8,9
y = a+bx
b = 7/11
a = 5-(7/11)*7=6/11
y = 6/11+(7/11)x

اگر بخواهیم معادله خط رگرسیون x بدست آوریم در تمام مخرج رابطه فوق (رابطه b) بجای x میتوان y قرار داد

x = a+by
a = -0.5
b = 1.5
x = -0.5+1.5y

۲) مثال

در جدول زیر بر حسب اینچ طول قد پدر و پسر ملاحظه میکنید

معادله خط رگرسیون را بنویسید؟

اگر قامت پدری ۷۵ شد پیش بینی میکنید فرزند او دارای چه قدی باشد؟

پدر	65	63	67	64	68	62	70	66	68	67	69	71
پسر	68	66	68	65	69	66	68	65	71	67	68	70

حل :

پدر را X و پسر را Y مینامیم زیرا قد پسر از پدر تبعیت میکند پس پدر متغیر مستقل است

$$\sum x_i = 65+63+.....+71 = 800 \qquad \sum y_i = 68+66+.....+70 = 811$$

$$\sum x_i^2 = 65^2+63^2+.....+71^2 = 53418 \qquad \sum x_i y_i = 65*68+63*66+....+71*70 = 54107$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{54107 - \frac{800*811}{12}}{53418 - \frac{(800)^2}{12}} = 0.476 \quad , \bar{x} = 800/12 \quad \bar{y} = 811/12$$

$$a = \bar{y} - b\bar{x} \qquad a = (811/12) - (0.476)(800/12) = 35.85 \Rightarrow$$

$$y = a + bx$$

$$y = 35.85 + 0.476x$$

$$y = 35.85 + (0.476*75) = 71.55$$

۳) مثال

در جدول زیر طول قد لوبیا (به سانتیمتر) بر حسب سن (به هفته) ملاحظه میکنید

معادله خط رگرسیون را بنویسید؟

اگر هفته دهم شد پیش بینی میکنید طول لوبیا دارای چه قدی باشد؟

سن	1	2	3	4	5	6	7			10		
طول	5	13	16	23	33	38	40			????		

حل :

سن را X و طول را Y مینامیم زیرا طول از سن تبعیت میکند پس سن متغیر مستقل است

$$\sum x_i = 1+2+.....+7 = 28 \qquad \sum y_i = 5+13+.....+40 = 168$$

$$\sum x_i^2 = 1^2+2^2+.....+7^2 = 140 \qquad \sum x_i y_i = 1*5+2*13+....+7*40 = 844$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{844 - \frac{28*168}{7}}{140 - \frac{(28)^2}{7}} = 6.143 \quad , \bar{x} = 28/7 \quad \bar{y} = 168/7$$

$$a = \bar{y} - b\bar{x} \qquad a = (168/7) - (6.143)(28/7) = -0.572 \Rightarrow$$

$$y = a + bx$$

$$y = -0.572 + 6.143x$$

$$y = -0.572 + (6.143*10) = 60.85$$

۱) تمرین :

طی آماری طول قامت ۵۰ کودک بشرح ذیل است

70(6)	75(9)	80(5)	85(7)	90(4)	95(7)
100(2)	105(2)	115(6)	120(1)	125(1)	

میانگین و واریانس و انحراف معیار را حساب کنید

☜ شمارش - ترتیب و ترکیب:

شمارش

اصل جمع : اگر عمل A به n طریق و عمل B به m طریق بتوان انجام داد
چنانچه عمل L منوط به انجام عمل A یا عمل B باشد
در اینصورت تعداد طرق انجام عمل L عبارت است از

n+m

اصل ضرب : اگر عمل A به n طریق و عمل B به m طریق بتوان انجام داد
چنانچه عمل L منوط به انجام عمل A و عمل B توام باشد
در اینصورت تعداد طرق انجام عمل L عبارت است از

n*m

٤) مثال

از منزل تا محل کار با پنج اتوبوس و سه مینی بوس و دو تاکسی میتوان رفت به چند طریق میتوان به مقصد رسید

5+3+2=10

٥) مثال

از منزل تا محل دو ایستگاه است تا ایستگاه اول با پنج اتوبوس و از ایستگاه اول تا محل کار با سه مینی بوس میتوان رفت به چند طریق میتوان به مقصد رسید

5*3=15

٦) مثال

از بین ۴ قاضی و ۶ دبیر و ۳ بازرگان میخواهیم کمیته ۲ نفری با مشاغل مختلف تشکیل دهیم به چند طریق میشود

قاضی و دبیر + قاضی و بازرگان + دبیر و بازرگان 4*6+4*3+6*3=54

$$\binom{4}{1}\binom{6}{1}+\binom{4}{1}\binom{3}{1}+\binom{6}{1}\binom{3}{1}$$

۲) تمرین

در یک شهر شرکت مخابرات شماره های تلفن همراه یازده رقمی که رقم سمت چپ آنها با ۰۹۱۱۷۱۱و ۰۹۱۳۷۱۳و۰۹۱۱۷۱۵ شروع شود میتواند سرویس بدهد - کل تلفنهایی که میتواند واگذار شود چقدر است.

ترتیب

ترتیب n شئ (ترتیب انتخاب n شئ از بین n شئ) = n! (مثلاً با سه رقم متفاوت چند عدد سه رقمی)

تعداد ترتیب k شئ از بین n شئ = $P_k^n=\frac{n!}{(n-k)!}$ (مثلاً با سه رقم متفاوت چند عدد دو رقمی)

ترکیب = ترتیب اهمیت ندارد

ترکیب

تعداد ترکیب n شئ (ترکیب انتخاب n شئ از بین n شئ) (۱)

تعداد ترکیب k شئ از بین n شئ = $C_k^n=\frac{n!}{(n-k)!\,k!}$ (مثلاً انتخاب سه محصول ازبین ۱۰ محصول)

اگر n شئ داشته باشیم n1 تا از نوع اول و n2 تا از نوع دوم و .... باشد آنگاه تعداد ترتیب

$$\frac{n!}{n_1!\,n_2!\ldots.n_k!}$$

اگر n شئ داشته باشیم که در یک محیط دایره قرار گرفته باشند تعداد ترتیب

$$\frac{n!}{n}=(n-1)!$$

اگر k شئ از بین n شئ بخواهیم بخواهیم انتخاب کنیم چون ترتیب در این حالت مهم نیست ترکیب گویند

$$\binom{n}{k}=C_k^n=\frac{n!}{(n-k)!\,k!}$$

۷) مثال

- n شئ از بین n شئ انتخاب کنید

$$C_n^n=\binom{n}{n}=\frac{n!}{(n-n)!\,n!}=1$$

۸) مثال

- به چند طریق میتوان دو شئ را از بین سه شئ انتخاب کرد

اگر ترتیب مهم باشد

$$P_k^n=\frac{n!}{(n-k)!}\quad =\frac{3!}{(3-2)!}=6$$

اگر ترتیب مهم نباشد

$$C_k^n=\frac{n!}{(n-k)!\,k!}=\frac{3!}{2!(3-2)!}=3$$

۹) مثال

به چند طریق میتوان سه شئ را از بین سه شئ انتخاب کرد

اگر ترتیب مهم باشد

$$P_k^n=\frac{n!}{(n-k)!}\quad =\frac{3!}{(3-3)!}=6$$

اگر ترتیب مهم نباشد

$$C_k^n=\frac{n!}{(n-k)!\,k!}\quad =\frac{3!}{3!(3-3)!}=1$$

۱۰) مثال

چند عدد دو رقمی با ارقام ۹ و ۸ و ۶ و ۲ و ۱ میتوان نوشت

یکان دهگان

٤ ٥ بدون تکرار ارقام = ٤*٥=٢٠

$$P_2^5=\frac{5!}{(5-2)!}=20$$

روش دوم

چون ترتیب مهم است

$$= \frac{5!}{(5-2)!} = 4*5 = 20 \quad P_k^n = \frac{n!}{(n-k)!}$$

۱۱) مثال

چند عدد زوج سه رقمی با ارقام ۸ و ۵و ۶ و ۱ و ۲ بدون تکرار ارقام میتوان نوشت

بدون تکرار

روش اول - تعداد حالاتی که در یک موقعیت میتوان ارقام را بدون تکرار داشت

یکان	دهگان	صدگان
۳	٤	۳

بدون تکرار ۳*٤*۳=۳۶

روش دوم - کل اعداد سه رقمی که با پنج رقم میتوان نوشت

$$P_3^5 = \frac{5!}{(5-3)!} = 60$$

کل اعداد دو رقمی که با چپا رقم میتوان نوشت (یعنی کلیه سه رقمی های دارای سمت راست رقم فرد ۵)

$$P_2^4 = \frac{4!}{(4-2)!} = 12$$

کل اعداد دو رقمی که با چپا رقم میتوان نوشت (یهنی کلیه سه رقمی های دارای سمت راست رقم فرد ۱)

$$P_2^4 = \frac{4!}{(4-2)!} = 12$$

بنابراین کل اعداد زوج سه رقمی از پنج رقم

60-12-12=36

با تکرار ارقام

تعداد حالاتی که در یک موقعیت میتوان ارقام را با تکرار داشت

یکان	دهگان	صدگان
۵	۵	۳

با تکرار ۵*۵*۳=۷۵

۱۲) مثال

نمرات کلاسی بصورت P و F است بیست نفر نمره P گرفته اند به چند طریق دو نفر میتوان از بین آنها انتخاب کرد که نمره P گرفته باشند

$$C_k^n = \frac{20!}{2!(20-2)!}$$

۱۳) مثال

بیست نفر در یک کلاس نمره های مختلفی گرفته اند دو نفر چگونه میتوان از بین آنها انتخاب کرد

در این حالت ترتیب مهم است

$$P_2^{20} = \frac{20!}{(20-2)!}$$

۱٤) مثال

تعداد ۱۰ نفر دور میز دایره ای به چند طریق امکان نشستن دارند

(n-1)!=(10-1) !

۱۵) مثال

۵ لامپ سبز و ۲ لامپ زرد و ۳ لامپ قرمز داریم به چند طریق میتوان ده عدد سر پیچ مشخص را روی لامپها نصب کرد

اول	دوم	سوم
سبز	زرد	قرمز

$$\binom{10-5-2}{3}\binom{10-5}{2}\binom{10}{5}$$

۱۶) مثال

٤ کتاب ریاضی و ۳ کتاب شیمی و ۲ کتاب تاریخ و یک کتاب زبان (همه متفاوت) داریم این ده عدد کتاب را میخواهیم در قفسه مرتب کنیم بطوریکه کتابهای ریاضی کنارهم و کتابهای تاریخ کنارهم و ..... باشد

کل حالاتی که ٤ کتاب ریاضی میتواند کنار هم باشد = !٤ و کل حالاتی که ۲ کتاب تاریخ میتواند کنار هم باشد = !۲ و .....

و چهار گروه ریاضی و شیمی و تاریخ و زبان میتواند کنار هم قرار گیرد !٤

بنابراین کل حالات

4! * (4 ! * 3! * 2! * 1!)

۱۷) مثال

٤ کتاب ریاضی و ۳ کتاب شیمی (مشخص) داریم به چند طریق میتوان ۳ کتاب را داشت که ۲ تا ریاضی و یکی شیمی در آن باشد

$$\binom{3}{1}\binom{4}{2}$$

۱۸) مثال

به چند طریق میتوان یک کمیته ۳ نفری شامل ۲ پزشک عمومی و یک چشم پزشک از بین ٤ پزشک عمومی و ۳ چشم پزشک داشت

$$\binom{3}{1}\binom{4}{2}$$

۱۹) مثال

به چند طریق میتوان از بین ۵ خانم و ۷ آقا کمیته ای شامل ۲ خانم و ۳ آقا داشت

$$\binom{5}{2}\binom{7}{3}$$

۲۰) مثال

به چند طریق میتوان از بین ۵ خانم و ۷ آقا کمیته ای شامل ۲ خانم و ۳ آقا داشت ولی چون ۲ آقای مشخص با هم اختلاف دارند نباید باهم باشند

تعدادحالات کمیته سه نفره آقایان که آن دو نفر نباشند $A=\binom{2}{0}\binom{5}{3}$

تعدادحالات کمیته سه نفره آقایان که یکی از آن دو نفر باشد $A=\binom{2}{1}\binom{5}{3}$

تعداد حالات کمیته دو نفره از بین ۵ خانم $A=\binom{5}{2}$

تعداد کل حالت مورد نظر = (A + B) * C

۲۱) مثال

معادله $x1+x2+x3+ \quad +xk=n$ الف -چند جواب صحیح مثبت دارد ب - چند جواب مثبت و صفر دارد

حالت اول - ابتدا بصورت ساده زیر در نظر میگیریم

$x1+x2+x3=8$

مثل این است که بگوییم دو چوب خط چگونه روی ۸ گو میتوان قرار داد

$000 \mid 000 \mid 00 \quad \Rightarrow \quad \binom{8-1}{3-1}$

کل تعداد حالات

$\binom{n-1}{k-1}$ تعداد جواب صحیح

برای وقتی جواب صفر هم مد نظر باشد به دو طرف معادله اصلی به تعداد k تا عدد یک اضافه میکنیم

$(x1+1)+(x2+1)+ \ldots\ldots+ (xk+1)=n+1+1\ldots\ldots1$

$y1+ \quad y2 + \ldots\ldots \quad yk \ = n+k$

مثل حالت اول حل میشود

$\binom{n+k-1}{k-1}$ تعداد جواب صحیح و صفر

۲۲) مثال

چهار پروژه داریم و میتوانیم روی هر کدام مضارب صحیح از هزار تومان خرج کنیم

الف - تعداد حالات که بخواهیم ۲۰ هزار تومان را مصرف کنیم

ب - تعداد حالات که مقداری از ۲۰ هزار تومان را خرج کنیم

حالت الف -

$x1+x2+x3+x4= 20$

$\binom{20+4-1}{4-1}$

حالت ب - یک پروژه فرضی پنجم اضافه میکنیم

$x1+x2+x3+x4 <= 20$

$x1+x2+x3+x4+x5 = 20$

$\binom{20+5-1}{5-1}$

☜ احتمالات

تعداد کل حالات ممکن

تعداد کل حالات مطلوب

احتمال

$$\text{احتمال} = \frac{\text{کل حالات مطلوب}}{\text{کل حالات ممکن}}$$

احتمال شرطی

$$P(M|E) = \frac{P(M \cap E)}{P(E)}$$

احتمال توام دو پیش آمد

$$P(M \cap E) = P(E).P(M|E)$$

احتمال توام دو پیش آمد مستقل

$$P(M \cap E) = P(E).P(M)$$

همیشه احتمال بین صفر و یک است

قضیه بیز

اگر بتعداد Ei پیش آمد مستقل داشته باشیم اگر A یک پیش آمد دلخواه باشد

$$P(E_r \mid A) = \frac{P(E_r)P(A|E_r)}{\sum_{i=1}^{n} P(E_i)P(A|E_i)}$$

۲۳) مثال

سکه ای دوبار پرتاب میکنیم احتمال حداقل یک شیر آمدن چقدر است

S= { HH , HT , TH , TT } فضای کلی

¼ ¼ ¼ ¼ احتمال هر حالت

A= { HH,HT,TH } فضای مورد نظر

P(A)=3/4 احتمال

۲٤) مثال

تاسی ناریب که احتمال عدد زوج آن دو برابر فرد است داریم با یک بار پرتاب احتمال اینکه عدد کمتر از ۴ بیاید چقدر است

S = { 1 , 2 , 3 , 4 , 5 , 6 }
w 2w w 2w w 2w
w+2w+w+2w+w+2w=1
9w=1 w = 1/9
S = { 1 , 2 , 3 , 4 , 5 , 6 }
1/9 2/9 1/9 2/9 1/9 2/9
A = { 1 , 2 , 3 }
1/9 2/9 1/9
P(A) = 1/9+2/9+1/9=4/9

۲۵) مثال

در یک ظرف ۶ توپ سفید و ۵ توپ سیاه داریم دو توپ را بیرون میاوریم احتمال اینکه یکی سفید و یکی سیاه باشد چقدر است ؟. احتمال

اولی سفید دومی سیاه

بدون جایگزینی

روش اول

تعداد کل حالات ممکن

تعداد کل حالات دو توپ = 11 * 10 = 110 حالت توپ دوم =10 حالت توپ اول = 11

تعداد کل حالات مطلوب

اولی سیاه دومی سفید = 5 * 6 = 30

اولی سفید دومی سیاه = 6 * 5 = 30

30+ 30 = 60

احتمال مورد نظر P (A) = 60 / 110

روش دوم $\frac{\text{مطلوب}}{\text{ممکن}} = \frac{\binom{5}{1}\binom{6}{1}}{\binom{11}{2}} = \frac{60}{110}$

قسمت دوم P(wb)=P(w1).P(w1b2)=6/11+5/10=30/110

روش دوم $\frac{\binom{5}{0}\binom{6}{1}}{\binom{11}{1}} \cdot \frac{\binom{5}{1}\binom{5}{0}}{\binom{10}{1}} = \frac{30}{110}$

۲۶) مثال

بیست نفر داریم احتمال اینکه حداقل دو نفر روز تولد یکسان داشته باشند چقدر است

تعداد کل فضا = 365 * 365... * 365 = (365) ^ 20

احتمال تولد روزهای مختلف = 365 * 364 * 363... * 346

P(A)= (365 * 364 * 363 ....* 346) / (365^20)= 0.5886

P(B) = 1-P(A)=0.4114

۲۷) مثال

مطابق جدول ۹۰۰ نفر داریم یک فرد انتخاب میکنیم متوجه میشویم که استخدامی است احتمال مرد بودن چقدر است

	مستخدم	بیکار	
مرد	460	40	500
زن	140	260	400
	600	300	900

$$P(M \downarrow E) = \frac{P(M \cap E)}{P(E)} = \frac{460/900}{600/900} = 460/600$$

۲۸) مثال

در جعبه ای ۲۰ فیوز داریم و میدانیم که ۵ عدد از آنها خراب است دو فیوز انتخاب میکنیم احتمال هر دو خراب چقدر است؟

روش اول

P(A ∩B) = P (A) . P (B) = (5/20) . (4 / 19)

روش دوم

$$P(Z)=\frac{\binom{15}{0}\binom{5}{2}}{\binom{20}{2}}$$

۲۹) مثال

۷۰ در صد دانشجویان مرتب و ۳۰ در صد نا مرتب هستند دانشجوی مرتب ۹۰٪ احتمال موفقیت و دانشجوی نا مرتب ۴۰ در صد احتمال موفقیت دارند یک دانشجو بتصادف انتخاب میکنیم ملاحظه شد قبول شده احتمال مرتب بودن او جقدر است .

A=موفق

B1=مرتب

B2=نا مرتب

$$P(B1\downarrow A)=\frac{P(B1).P(A\downarrow B1)}{P(B1).P(A\downarrow B1)+P(B2).P(A\downarrow B2)}$$

$$\frac{0.7*0.9}{0.7*0.9+0.3*0.4}=\frac{63}{75}$$

۳۰) مثال

شخصی سکه‌ای در دست دارد که ۷۰٪ حدس میزنیم سکه اش دو شیری باشد سکه را پرتاب میکند ملاحظه میشود شیرآمده است احتمال دو شیری بودن سکه چقدر است؟

A=شیر P(A)= 50%

B1=یک شیری بودن P(B1)=30%

B2=دو شیری بودن P(B2)=70%

$$P(B2\downarrow A)=\frac{P(B2).P(A\downarrow B2)}{P(B1).P(A\downarrow B1)+P(B2).P(A\downarrow B2)}$$

$$\frac{0.7*1}{0.3*0.5+0.7*1}=\frac{0.7}{0.85}=0.81=81\%$$

۳۱) مثال

آزمایش تشخیص سرطان در مورد ۹۵٪ بیماران سرطانی جواب مثبت میدهد

آزمایش تشخیص سرطان در مورد ۱۰٪ بیماران غیر سرطانی جواب مثبت میدهد

از بین بیماران یک بیمارستان که ۳٪ آنها سرطانی هستند بیماری بتصادف انتخاب میکنیم

آزمایش تشخیص سرطان در مورد این بیمار سرطانی جواب مثبت میدهد احتمال اینکه واقعاً سرطانی باشد چقدر است.

A=پاسخ مثبت

B1=سرطانی باشد P(B1)=3%

B2= سرطانی نباشد P(B2)=97%

$$P(B1\downarrow A)=\frac{P(B1).P(A\downarrow B1)}{P(B1).P(A\downarrow B1)+P(B2).P(A\downarrow B2)}$$

$$\frac{0.03*0.95}{0.03*0.95+0.97*0.10}=$$

متغیر تصادفی *Random Variable*

تابعی از فضای نمونه – زیر مجموعه از اعداد حقیقی معمولا با حروف بزرگ مثل X نمایش میدهند
برد تابع را تکیه گاه متغیر تصادفی میگویند و با Sx نمایش میدهند

تابع توزیع *Distribution Function*

اگر X متغیر تصادفی باشد آنگاه تابع توزیع X بصورت $F_X(x)$ نشان میدهند دامنه تابع R میباشد و بصورت زیر تعریف میشود

$$F_X(x) = P(X \leq x)$$

خواص تابع توزیع

۱- $0 \leq F_X(x) \leq 1$

۲-

$$\lim_{x \to -\infty} F_X(x) = F_x(-\infty) = 0$$

$$\lim_{x \to +\infty} F_X(x) = F_x(+\infty) = 1$$

۳- حداقل از راست در هر نقطه پیوسته است

۴- غیر نزولی است یعنی

$$x_1 < x_2 \rightarrow F_X(x_1) < F_X(x_2)$$

نکته – هر تابع حقیقی که دارای چهار خاصیت فوق باشد تابع توزیع است

نکته –

$$\lim_{x \to a^+} F_X(x) = F_X(a^+)$$

تابع چگالی احتمال :

۱- گسسته : اگر X یک متغیر تصادفی گسسته باشد آنگاه تابع چگالی احتمال که با $f_X(x)$ نشان داده میشود بصورت زیر تعریف میشود

$$f_X(x) = P(X = x)$$

که دامنه آن مجموعه اعداد حقیقی است و دارای دو خاصیت زیر است

$$f_X(x) \geq 0$$

$$\sum_x f_X(x) = 1$$

۲- پیوسته : میزان فشردگی احتمال روی خط اعداد حقیقی تشکیل تابعی حقیقی میدهد بنام چگالی احتمال $f_X(x)$ که دارای دو خاصیت زیر است

$$f_X(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

رابطه تابع توزیع با تابع چگالی احتمال

۱- تبدیل در گسسته (با داشتن تابع چگالی تابع توزیع بدست آورید و بالعکس)

$$F_X(x) = \sum_{t \le x} f_X(t)$$

$$f_X(x) = F_X(x) - F_X(x^-)$$

۲- تبدیل در پیوسته (با داشتن تابع چگالی تابع توزیع بدست آورید و بالعکس)

$$F_X(x) = \int_{-\infty}^{x} f_x(t)\,dt$$

$$f_X(x) = F'_X(x_0)$$

۳۲) مثال

سکه ای سه بار پرتاب میکنیم تعداد شیرها را بررسی نمایید

S={HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}

تعداد شیر	0	1	1	2	1	2	2	3
احتمال هر مورد	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

3/8= احتمال دو شیر

1/8= احتمال سه شیر

3/8= احتمال یک شیر

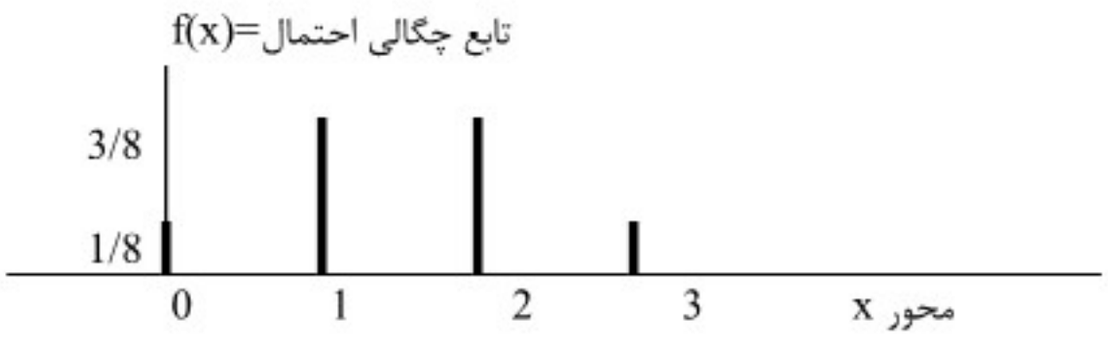
1/8= احتمال اصلا شیر نیاد

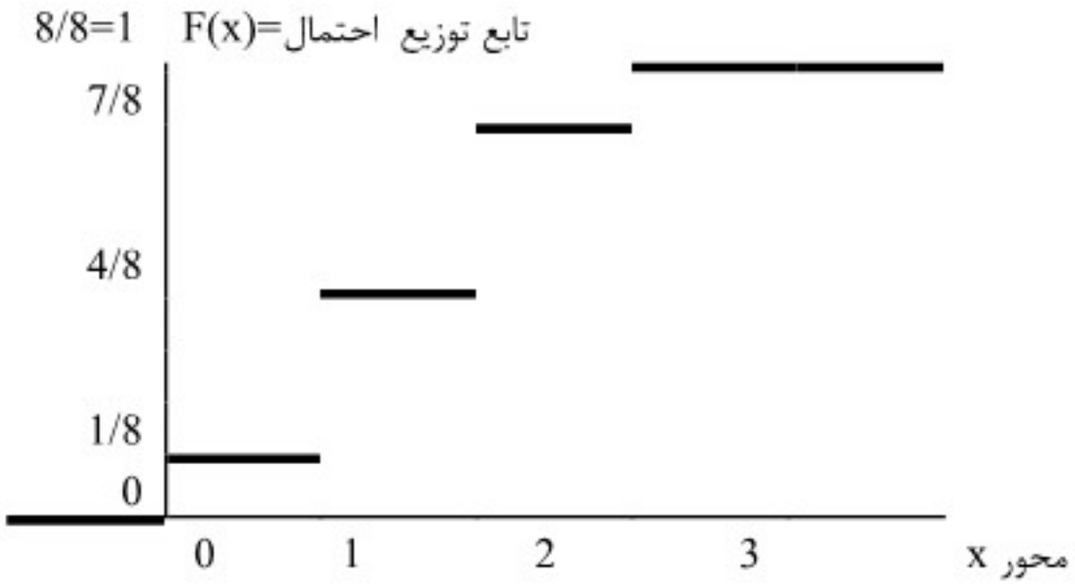
=Xتعداد شیر	0	1	2	3
چگالی احتمالP(X=x) = f(x)تابع احتمال	1/8	3/8	3/8	1/8

Rx= { 0 , 1 , 2 , 3) x برد

$$F(x) = \begin{Bmatrix} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 4/8 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & 3 \le x \end{Bmatrix}$$

توزیع جمعی احتمال





۳۳) مثال

شرکتی ۸ کامپیوتر دارد که ۳ عدد آن خراب است ۲ کامپیوتر خرید میکنیم احتمال خرابی چقدر است

$$\frac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}}$$ احتمال هر دو خراب

$$\frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}}$$ احتمال یکی خراب

$$\frac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}}$$ احتمال هر دو سالم

x= کامپیوتر خراب

R(x)= { 0,1,2 }

f(x)= توزیع احتمال f (0)= 10/28 f(1)=15/28 f(2)=3/28

$$F(x)=\begin{Bmatrix} 0 & x<0 \\ 10/28 & 0\le x<1 \\ 25/28 & 1\le x<2 \\ 1 & 2\le x \end{Bmatrix}$$ توزیع تجمعی احتمال

توزیع‌های مهم

آزمایش برنولی

آزمایشی که دارای دو نتیجه باشد (موفقیت S و شکست F) احتمال موفقیت را P و احتمال شکست 1-P میگردد.

توزیع دوجمله‌ای

اگر آزمایش برنولی را n بار بطور مستقل تکرار میکنیم و تعداد پیروزی‌ها را X بنامیم

برد تابع (تکیه گاه متغیر تصادفی)

متغیر تصادفی X دارای توزیع دوجمله ای بصورت زیر مینویسند

$$X \sim B(n, p)$$

تابع چگالی باین صورت است

$$f_x(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,2,\ldots,n \\ 0 & \text{سایرین} \end{cases}$$

توزیع پواسن

اگر یک آزمایش برنولی مستقلاً تکرار کنیم و تعداد پیروزیها در یک فاصله زمانی مشخص X بنامیم آنگاه X دارای توزیع پواسن با پارامتر لاندا است که لاندا متوسط پیروزها است

$$X \sim p(\lambda)$$

$$f_x(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & x = 0,1,2,\ldots,n \\ 0 & \text{سایرین} \end{cases}$$

توزیع نرمال

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

$$f_x(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} & -\infty < x < +\infty \quad , \ \mu = \text{میانگین} \qquad \text{واریانس} \ \sigma > 0 \\ 0 & \text{سایرین} \end{cases}$$

اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد آنگاه متغیر تصادفی $Z = \dfrac{X-\mu}{\sigma}$ دارای توزیع نرمال استاندارد است $Z \sim N(0,1)$

$$X \sim N(\mu, \sigma^2)$$
$$Z \sim N(0,1)$$

و احتمال آن چنین خواهد بود

$$P(x < b)$$

$$p(\frac{x-\mu}{\sigma} < \frac{b-\mu}{\sigma}) = p(z < \frac{b-\mu}{\sigma}) \Rightarrow$$

حالا از روی جدول نرمال استاندارد احتمال حاصل میشود

توزیع نرمال استاندارد

اگر در توزیع نرمال میانگین مساوی صفر و واریانس یک باشد آنگاه توزیع نرمال استاندارد نامیده میشود

$$Z \sim N(0,1)$$

$$f_z(z) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{\frac{-z}{2}} & -\infty < x < +\infty \\ 0 & \text{سایرین} \end{cases}$$

ثابت شده است که در طبیعت آمار بصورت نرمال توزیع شده است که معمولاً ۶۸٪ بین $\mu \pm \sigma$ و ۹۶٪ بین $\mu \pm 2\sigma$ خواهد بود

۳۴) مثال

تیراندازی ۲۰ بار بطرف یک هدف شلیک میکند احتمال به هدف اصابت کردن هر شلیک ۰,۴ است

الف) تابع چگالی احتمال را بنویسید

ب) احتمال اینکه از این ۲۰ بار دقیقا ۱۲ بار بهدف بخورد چقدر است

ج) احتمال اینکه از این ۲۰ بار حداکثر ۱۲ بار بهدف بخورد چقدر است

توجه شود

۱- که در آزمایش برنولی زمان و مکان مطرح نیست

۲- چون یا به هدف میخورد یا نمیخورد پس برنولی است

$$f_x(x)=\begin{cases}\binom{20}{x}0.4^x0.6^{20-x} & x=0,1,2,....20\\ 0 & \text{سایرین}\end{cases}$$

$$P(x=12)=\binom{20}{12}0.4^{12}0.6^{8}$$

$$P(x\le 12)=\sum_{x=0}^{12}\binom{20}{x}0.4^x0.6^{20-x}$$

ضمنا اگر جدول کوچکتر یا مساوی در دسترس باشد میتوان نوشت

$$P(x=a)=F_x(a)-F_x(a^-)$$
$$P(x=12)=P(x\le 12)-P(x\le 11)$$

در جدول برای موارد ذیل داریم

$$n=20,\ p=0.4,\ r\le 12$$
$$P(x\le 12)=0.97$$
$$P(x\le 11)=0.943$$
$$P(x=12)=0.97-0.943=0.0355$$
$$P(x\ge 12)=1-P(x\le 12)=1-0.943=0.0565$$

۳۵) مثال

در سال گذشته بطور متوسط ۱۰ زلزله در شمال کشور بوقوع پیوسته است

احتمال اینکه در سه ماه آینده دقیقا ۳ زلزله بوقوع بپیوندد چقدر است

احتمال اینکه در سه ماه آینده حداکثر ۶ زلزله بوقوع بپیوندد چقدر است

در ۱۲ ماه ۱۰ لرزه بنابراین در ۳ ماه با یک تناسب ساده ۲/۵ زلزله بطور متوسط خواهیم داشت

(e = 2.7183)

$$p(x=3),\ x\sim p(2.5)$$
$$p(x=3)=\frac{e^{-2.5}2.5^{3}}{3!}$$
$$p(x\le 6)=\sum_{x=0}^{6}\frac{e^{-2.5}2.5^{x}}{x!}$$

۳۶) مثال

نمرات امتحان کلاسی دارای توزیع نرمال و میانگین ۶۵ و انحراف معیار ٤ میباشد ، از این کلاس ٤٠ نفری ٤ نفر مـردود شدند ، کمترین نمره قبولی را مشخص کنید

$$v=\sigma^2=4^2=16$$
$$X\approx N(\mu,\sigma^2)=(65,16)$$

برای تقریب به توزیع نرمال چون حدنهایی جمعیت و نمونه یکی است و مشخص است تقسیم بر σ و گرنه تقسیم بر $\sqrt{\frac{\sigma^2}{n}}$

$$p(x<a)=\frac{4}{40}=0.1$$
$$p(\frac{x-\mu}{\sigma}<\frac{a-65}{4})=0.1$$
$$p(z<\frac{a-65}{4})=0.1$$

از روی جدول عدد ۱.۲۸ - حاصل میشود

$$\frac{a-65}{4}=-1.28\Rightarrow a=59.88$$

☜ توزیع *t-student*

این توزیع برای نمونه های کوچکتر از ۳۰ میباشد و با یک درجه آزادی همراه است هرچه درجه آزادی بیشـتر باشـد نمـودار تـابع چگـالی احتمال به نمودار نرمال استاندارد نزدیک است با درجه آزادی ۳۰ و بیشتر تابع توزیع t بر تابع توزیع نرمال دقیقاً منطبق میشود

$$x\sim t(n)$$

توزیع هندسی

توزیع فوق هندسی

توزیع تی t

توزیع کای مربع

توزیع F

قضیه حد مرکزی

۱- اگر از یک جمعیت n عضوی با میانگین معلوم μ و واریانس معلوم σ^2 یک نمونه تصادفی n تایی $(n \succ 30)$ یکی یکی و با عمل جایگزینی انتخاب کنیم آنگاه میانگین نمونه ای یعنی $\bar{x}$ تقریبا دارای توزیع نرمال با میانگین $\mu_{\bar{x}} = \mu$ و واریانس $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ و انحراف معیار $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$ خواهد بود

و متغیر تصادفی $z = \frac{\bar{x} - \mu_{\bar{x}}}{\sqrt{\sigma^2/n}}$ دارای توزیع نرمال استاندارد میباشد

۲- اگر از یک جمعیت نرمال با واریانس مجهول یک نمونه تصادفی n تایی با $n<30$ انتخاب کنیم آنگاه متغیر تصادفی $t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$ دارای توزیع t با n-1 درجه آزادی است که در آن $\bar{x}$ میانگین نمونه و μ میانگین جمعیت و S انحراف معیار نمونه میباشد.

۳- اگر در یک جمعیت دوجمله ای با احتمال پیروزی p یک نمونه تصادفی n تایی $(n \succ 30)$ یکی یکی انتخاب کنیم (در دو جمله ای همیشه با عمل جایگزینی است) آنگاه احتمال پیروزی در نمونه یعنی $\hat{p}$ تقریبا دارای توزیع نرمال با میانگین $\mu_{\hat{p}} = E(\hat{p}) = p$ و واریانس $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$ خواهد بود

نکته : چون یک توزیع پیوسته به گسسته تقریب زده میشود در موارد کوچکتر نصف واحد یعنی $\frac{1}{2n}$ اضافه و در موارد بزرگتر نصف واحد یعنی $\frac{1}{2n}$ اضافه کم میشود

نکته : آزمایش برنولی مستقل از هم است یعنی با جایگزینی است

۳۷) مثال

کارخانه لامپ سازی ادعا میکند که طول عمر لامپهایش ۱۲۸۰ با انحراف معیار ۱۵۰ است از این کارخانه یک نمونه ۱۰۰ تایی لامپ انتخاب میکنیم احتمال اینکه این لامپها بیشتر از ۱۳۰۰ عمر کند چقدر است؟

از جمعیتی با میانگین $\mu = 1280$ و $\sigma = 150$ یک نمونه تصادفی $n = 100$ تایی انتخاب میکنیم احتمال اینکه $p(\bar{x} > 1300)$ باشد چقدر است

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

$$\bar{x} \sim N(1280, 150^2/100)$$

$$p(\bar{x} > 1300) = p\left(\frac{\bar{x} - \mu_{\bar{x}}}{\sqrt{\sigma^2/n}} > \frac{1300 - 1280}{\sqrt{150^2/100}}\right)$$

$$p(z > 1.33) = 1 - p(z \le 1.33) = 1 - 0.9082 = 0.0918$$

۳۸) مثال

کارخانه لامپ سازی ادعا میکند که طول عمر لامپهایش ۱۲۸۰ با انحراف معیار ۱۵۰ است از این کارخانه یک نمونه ۲۵ تایی لامپ انتخاب میکنیم احتمال اینکه این لامپها بیشتر از ۱۳۰۰ عمر کند چقدر است؟

از جمعیتی با میانگین $\mu = 1280$ و $s = 150$ یک نمونه تصادفی $n = 25$ تایی انتخاب میکنیم احتمال اینکه $p(\bar{x} > 1300)$ باشد چقدر است

$$\frac{\bar{x}-\mu}{\sqrt{s^2/n}} \sim t(24)$$

$$p(\frac{\bar{x}-\mu}{\sqrt{s^2/n}} > \frac{1300-1280}{\sqrt{150^2/25}})$$

$$p(t > 0.67) = 1 - p(z \le 0.67) \text{ از جدول}$$

$$t_p(24) = 0.67 \quad \Rightarrow p = 0.74 \quad \Rightarrow p(t > 0.67) = 1 - 0.74 = 0.26$$

موارد فوق ازجدول توزیع t با n-1=25-1=24 درجه آزادی و از داخل جدول برای 0.67 و در سطر اول جدول عدد احتمال 0.74 حاصل شد

۳۹) مثال

کارخانه تلوزیون سازی ۲٪ از تولیداتش معیوب است اگر یک محموله ۴۰۰ تایی از این کارخانه خریداری شود

الف) احتمال اینکه ۳٪ یا بیشتر از تولیداتش معیوب باشد چقدر است

ب) احتمال اینکه ۲٪ یا کمتر از تولیداتش معیوب باشد چقدر است

$$\frac{1}{2n} = \frac{1}{2*400} = 0.00125$$

$$P(\hat{p} \ge 0.03) = P(\hat{p} \ge 0.03 - 0.00125)$$

$$P(\hat{p} > 0.02875) =$$

$$P(\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.02875-0.02}{\sqrt{\frac{0.02*0.98}{400}}})$$

$$p(z > 1.25)$$

$$P(\hat{p} \le 0.02) = P(\hat{p} < 0.02 + 0.00125) = P(\hat{p} < 0.02125)$$

$$P(z < \frac{0.02125-0.02}{\sqrt{\frac{0.02*0.98}{400}}}) = P(z < 0.18) = 0.5714$$

برآورد فاصله ای

سطح زیر منحنی ۱ یک است اگر $\alpha/2$ در سمت راست منحنی و $\alpha/2$ در سمت چپ منحنی باشد

بنابراین $1-\alpha$ درصد مطمئن هستیم که z بین $Z_{\alpha/2}$ و $-Z_{\alpha/2}$ است

$$P(-|Z_{\alpha/2}| < Z < |Z_{\alpha/2}|) = 1-\alpha$$

$$P(-|Z_{\alpha/2}| < \frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} < |Z_{\alpha/2}|) = 1-\alpha$$

$$P(\bar{x} - |Z_{\alpha/2}|\sqrt{\sigma^2/n} < \mu < \bar{x} + |Z_{\alpha/2}|\sqrt{\sigma^2/n}) = 1-\alpha$$

بنابراین $1-\alpha$ درصد مطمئن هستیم میانگین جمعیت بین $\bar{x} - Z_{\alpha/2}\sqrt{\sigma^2/n}$ و $\bar{x} + Z_{\alpha/2}\sqrt{\sigma^2/n}$ است

مثلاً برای جمعیت دو جمله ای

$$P(\hat{p} - |Z_{\alpha/2}|\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + |Z_{\alpha/2}|\sqrt{\hat{p}(1-\hat{p})/n}) = 1-\alpha$$

نکته : اگر در مواردی σ معلوم نبود از S استفاده میکنیم

۴۰) مثال

یک نمونه ۱۰۰ نفری از رای دهندگان انتخابات ریاست جمهوری انتخاب شدند ۵۹ نفر اظهار داشتند که بنفع کاندید A رای میدهند فاصله اطمینان ۹۵٪ برای نسبت افرادی که به کاندید A رای میدهند حساب کنید

$$1-\alpha = 0.95 \qquad \alpha/2 = 0.025$$

$$Z_{0.025} = 1.96 \qquad n = 100 \qquad \hat{p} = 59/100 = 0.59$$

$$P(\hat{p} - |Z_{\alpha/2}|\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + |Z_{\alpha/2}|\sqrt{\hat{p}(1-\hat{p})/n}) = 1-\alpha$$

$$0.59 - 1.96\sqrt{0.59*0.41/100} \quad \text{و} \quad 0.59 + 1.96\sqrt{0.59*0.41/100}$$

$$0.496 \quad \text{و} \quad 0.684$$

بنابراین میتوان با اطمینان ۹۵٪ اظهار داشت که بیش از ۵۰٪ مردم به کاندید A رای میدهند

امید ریاضی

اگر X یک متغیر تصادفی گسسته یا پیوسته باشد آنگاه امید ریاضی X که با علامت E(X) نشان داده میشود بصورت زیر تعریف میشود

$$E(X) = \sum_{x} x f_X(x)$$

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx$$

۴۱) مثال

صاحب دستگاهی هستیم که دارای نمایش سه سکه است

هر بار فشردن دسته کل نمایش را تغییر میدهد

اگر بازیگری دکمه فشار دهد بازاء هر تعداد شیر باید ۱۰۰ ریال بپردازیم

چند ریال بگیریم که نه سود کنیم و نه ضرر

S={HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}

تعداد شیر	0		100	100	100	200	200	300
احتمال هر مورد	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

ریال پرداختی بازاء تعداد شیر

(3/8=احتمال دو شیر) (1/8=احتمال سه شیر) (3/8= احتمال یک شیر) (1/8= احتمال اصلا شیر نیاد)

X=x	0	100	200	300	سایرین
$f_X(x)$	1/8	3/8	3/8	1/8	0

E(x)=0*1/8+100*3/8+200*3/8+300*1/8=150

که در حقیقت نام دیگرش میانگین است

۴۲) مثال

یک شرکت تولید برق در مقابل آتش سوزی به میزان خسارت کلی یک میلیارد تومان بیمه میشود

مدیر عامل از تجربیات خود میداند که :

احتمال آتش سوزی کلی یک میلیونیم است احتمال آتش سوزی ۵۰٪ یک صدهزارم است

احتمال آتش سوزی ۲۵٪ یک ده هزارم است

حق بیمه سالیانه چقدر باشد تا در سال یک میلیون تومان سود کنیم

1-0.000001-0.00001-0.0001=0.999889

X=x	-1000	-500	-250	+a	میلیون تومان
$f_X(x)$	0.000001	0.00001	0.0001	0.999889	0

$E(X) = \sum_x x f_x(x) = 1$ میلیون تومان

1= (-1000*0.000001)+ (-500*0.00001)+ (-250*0.0001)+ (a*0.999889)

a=1.031114

یعنی اگر در سال یک میلیون سی هزارتومان دریافت کنیم سالی یک میلیون تومان سود میکنیم

توزیع	احتمال	متوسط یا میانگین	واریانس	
برنولی $x \sim B(n,p)$ $n=1$	$p^x(1-p)^{1-x}$	p	pq	$x=0,1$
دوجمله ای $x \sim B(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	npq	$x=0,1,\ldots.n$
پواسن $x \sim P(\lambda)$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	$x=0,1,\ldots.$
نرمال $x \sim N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$	μ	σ^2	$-\infty \le x \le +\infty$
نرمال استاندارد $x \sim N(0,1)$	$\frac{1}{\sqrt{2\pi}}e^{\frac{-z}{2}}$	0	1	$-\infty \le z \le +\infty$

آزمون فرضها :

میخواهیم بحث قبول یا رد فرض صفر را بررسی کنیم

فرض H_α مشابه خواسته مسئله در نظر میگیریم و فرض H_0 را مساوی آن مقدار در مسئله قرار میدهیم

۱- اگر در فرض H_α علامت کوچکتر بود آنگاه اگر $Z < -|Z_\alpha|$ بود فرض H_0 را رد میکنیم و پذیرای فرض مقابل یعنی H_α میشویم وگرنه اعلام میکنیم دلیلی بر رد H_0 نداریم

۲- اگر در فرض H_α علامت بزرگتر بود آنگاه اگر $Z > +|Z_\alpha|$ بود فرض H_0 را رد میکنیم و پذیرای فرض مقابل یعنی H_α میشویم وگرنه اعلام میکنیم دلیلی بر رد H_0 نداریم

٤۳) مثال

در آزمایش طول قامت ۵۰ کودک میانگین 89.2 سانتیمتر و انحراف معیار 15.5 سانتیمتر میباشد آیا میتوان ادعا کرد طول قامت کودکان کمتر از ۹۰ سانتیمتر است (هر وقت آلفا مشخص ندادند مقدارش 0.05 در نظر میگیریم)

$n = 50 \qquad \bar{x} = 89.2 \qquad s = 15.5$

$H_a : \ \mu < 90$

$H_0 : \mu = 90$

$$\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{15.5^2}{50}} \qquad Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{89.2 - 90}{\sqrt{\frac{15.5^2}{50}}} = -0.365$$

$Z_\alpha = Z_{0.05} = m \Leftrightarrow \ p(z \le m) = 1 - 0.05 = 0.95$

$from\ table \Rightarrow m = 1.645 = Z_{0.05} = Z_\alpha$

$(Z = -0.365) < (-Z_\alpha = -1.645)$ غلط

چون $Z < -Z_\alpha$ نیست اعلام میکنیم دلیلی بر رد H_0 نداریم یعنی نمیتوان فرض H_α را پذیرفت پس نمیتوان گفت قامت کوچکتر از ۹۰ است

٤٤) مثال

لامپهای ساخت کارخانه ای با انحراف معیار ۱۴۲ ساعت در یک نمونه ۱۰۰ تایی میانگین عمرشان ۱۲۸۰ ساعت شد صاحب کارخانه ادعا میکند که با همین اطلاعات میانگین عمر لامپهای ای کارخانه از ۱۲۰۰ ساعت بیشتر است آیا میتوان ادعای او را قبول یا رد کرد ؟ (هر وقت آلفا مشخص ندادند مقدارش 0.05 در نظر میگیریم)

$n = 100 \qquad \bar{x} = 1280 \qquad s = 142$

$H_a : \ \mu > 1200$

$H_0 : \mu = 1200$

$$\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{142^2}{100}} \qquad Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{1280 - 1200}{\sqrt{\frac{142^2}{100}}} = 5.63$$

$Z_\alpha = Z_{0.05} = m \Leftrightarrow \ p(z \le m) = 1 - 0.05 = 0.95$

$from\ table \Rightarrow m = 1.645 = Z_{0.05} = Z_\alpha$

$(Z = 5.63) > (+Z_\alpha = +1.645)$ صحیح

چون $Z > +Z_\alpha$ میباشد فرض H_0 را رد میکنیم و پذیرای فرض مقابل یعنی H_α میشویم یعنی بله ادعای صاحب کارخانه میتواند درست باشد

٤٥) مثال

صاحب یک کارخانه داروسازی ادعا میکند که داروی ضد حساسیت این کارخانه در مورد بیش از ۹۰٪ حساس در مدت ۸ ساعت پس از مصرف دارو نتیجه مثبت میدهد

یک نمونه ۲۰۰ نفری آدمهای حساس انتخاب کردیم و دارو را استفاده نمودند و ۱۶۰ نفر بهبود یافتند آیا این داده ها دلیل کافی ارائه میدهد که ادعای صاحب کارخانه غلط است ؟ ($\alpha = 0.01$)

توجه شود که چون صورت مسئله خلاف نظر صاحب کارخانه خواسته علامت کوچکتر استفاده میشود

$n = 200 \quad \hat{p} = 160/200 = 0.8 \qquad s = 200 * 0.8 * (1 - 0.8)$

$H_a : \ \mu < 0.9$

$H_0 : \mu = 0.9$

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.9(1-0.9)}{200}} \qquad Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.8 - 0.9}{\sqrt{\frac{0.9(1-0.9)}{200}}} = -4.72$$

$Z_\alpha = Z_{0.01} = m \Leftrightarrow p(z \le m) = 1 - 0.01 = 0.99$

$from\ table \Rightarrow m = 2.33 = Z_{0.01} = Z_\alpha$

$(Z = -4.72) < (-Z_\alpha = -2.33)$ صحیح

چون $Z < -Z_\alpha$ میباشد فرض H_0 را رد میکنیم و پذیرای فرض مقابل یعنی H_α میشویم یعنی بلـه ادعـای صـاحب کارخانـه میتواند غلط باشد

مدل مهره و جعبه

اگر N جعبه متمایز داشته باشیم و اگر R مهره داشته باشیم و بخواهیم این R مهره درون جعبه قرار دهیم حالتهای زیر بوجود میاید

۱- مهره متمایز مهره مکرر مجاز

$$C(N,R) = N * N * N ..... * N = N^R$$

در این حالت اهمیتی ندارد که R < N باشد یا نباشد

٤٦) مثال

سه ظرف رنگی مشخص سبز و قرمز و آبی داریم و دو مهره A و B به چند طریق میتوان دو مهره درون سه ظرف چید که مهره مکرر مجاز باشد

$$C(3,2) = 3^2 = 9$$

سبز	AB	–	-	A	B			A	B
قرمز	–	AB	–	B	A	A	B		
آبی	-	–	AB	-	-	B	A	B	A
-	1	2	3	4	5	6	7	8	9

۲- مهره متمایز مهره مکرر غیرمجاز

$$C(N,\overline{R}) = N*(N-1)*(N-2).....*(N-R+1) = \frac{N!}{(N-R)!}$$

این حالت همان ترتیب میباشد

در این حالت باید R <= N باشد

٤٧) مثال

سه ظرف رنگی مشخص سبز و قرمز و آبی داریم و دو مهره A و B به چند طریق میتوان دو مهره درون سه ظرف چید که مهره مکرر غیر مجاز باشد

$$C(3,\overline{2}) = P(3,2) = \frac{3!}{(3-2)!} = 6$$

سبز	A	B			A	B			
قرمز	B	A	A	B					
آبی	-	-	B	A	B	A			
-	1	2	3	4	5	6	-	-	-

۳- مهره غیرمتمایز مهره مکرر غیرمجاز

$$C(\overline{N},\overline{R})=\binom{N}{R}=\frac{N!}{(N-R)!R!}$$

این حالت شبیه حالت ترکیب میباشد

در این حالت باید $R <= N$ باشد

۴۸) مثال

سه ظرف رنگی مشخص سبز و قرمز و آبی داریم و دو مهره A و A به چند طریق میتوان دو مهره درون سه ظرف چید که مهره مکرر غیر مجاز باشد

$$C(\overline{3},\overline{2})=\frac{3!}{(3-2)!2!}=3$$

سبز	A		A						
قرمز	A	A							
آبی	-	A	A						
-	1	2	3	-	-	-	-	-	-

٤- مهره غیرمتمایزمهره مکرر مجاز

$$C(\overline{N},R)=\binom{N+R-1}{R}=\binom{N+R-1}{N-1}$$

در این حالت میتواند $R > N$ باشد

شبیه اینکه بگوییم N جعبه بیکدیگر میچسبانیم که N+1 خط عمودی تشکیل میدهد و R مهره مشابه درون این جعبه ها جا دهیم چون خط اول و آخر ثابت است پس N-1 خط عمودی وجود دارد و R مهره مشابه درون آنها بچینیم

۴۹) مثال

سه ظرف رنگی مشخص سبز و قرمز و آبی داریم و دو مهره A و A به چند طریق میتوان دو مهره درون سه ظرف چید که مهره مکرر مجاز باشد

$$C(\overline{3},2)=\binom{3+2-1}{2}=\frac{4!}{(4-2)!2!}=6$$

سبز	A		A	AA					
قرمز	A	A			AA				
آبی	-	A	A			AA			
-	1	2	3	4	5	6	-	-	-

مثالهایی از احتمال شرطی

۵۰) مثال

تاس را پرتاب میکنیم مشاهده میشود عدد زوج است احتمال اینکه بر ۳ بخش پذیر باشد چقدر است

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

$B = \{2,4,6\}$

$A = \{3,6\}$

$A \cap B = \{6\}$

$$p(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

۳) تمرین

تاس را پرتاب میکنیم مشاهده میشود بر ۳ بخش پذیر است احتمال اینکه عدد زوج باشد چقدر است؟(۱/۲)

۵۱) مثال

ظرفی محتوی ۵ گوی قرمز و ۷ گوی سیاه است

از درون ظرف دو گوی را یکی یکی بدون جایگزینی بیرون میاوریم

الف) احتمال اینکه گوی اول قرمز و گوی دوم سیاه باشد چقدر است

ب) احتمال اینکه یک گوی قرمز و یک گوی سیاه باشد چقدر است

الف) ترتیب اهمیت دارد

$$P(R1 \cap B2) = P(R1).P(B2 | R1) = (5/12).(7/11)$$

ب) ترتیب اهمیت ندارد چون در حالت ب دو پیش آمد ناسازگارند با هم جمع میشوند

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

$$P[(R1 \cap B2) \cup (B1 \cap R2)] = P(R1 \cap B2) + P(B1 \cap R2) = P(R \cap B) + P(B \cap R) =$$

$$P(R).P(B|R) + P(B).P(R|B) = [(5/12).(7/11)] + [(7/12).(5/11)]$$

میتوانستیم یک حالت بدست آورده در ۲ ضرب کنیم

ب) برای حالت ب میتوانستیم از فرمول ترکیب استفاده کنیم

$$\frac{\binom{5}{1}\binom{7}{1}}{\binom{12}{2}}$$

۵۲) مثال

ظرفی محتوی ۵ گوی قرمز و ۷ گوی سیاه است

از درون ظرف سه گوی را یکی یکی بدون جایگزینی بیرون میاوریم

الف) احتمال اینکه گوی اول قرمز و گوی دوم سیاه و گوی سوم قرمز باشد چقدر است

ب) احتمال اینکه دو گوی قرمز و یک گوی سیااه باشد چقدر است

الف) ترتیب اهمیت دارد

$$P(R1 \cap B2 \cap R3) = (5/12).(7/11).(4/10)$$

ب) ترتیب اهمیت ندارد چون در حالت ب دو پیش آمد ناسازگارند با هم جمع میشوند

$$P(RBR \text{ or } RRB \text{ or } BRR) = 3 * P(RBR) = 3.[(5/12).(7/11)].(4/10)]$$

ب) برای حالت ب میتوانستیم از فرمول ترکیب استفاده کنیم

$$\frac{\binom{5}{2}\binom{7}{1}}{\binom{12}{3}}$$

۵۳) مثال

یک سکه دو بار پرتاب میکنیم احتمال اینکه دفعه اول شیر و دفعه دوم خط بیاید چقدر است

چون دومی مستقل از اولی است

$$P(H \cap T) = P(H).P(T \mid H) = P(H).P(T) = (1/2).(1/2) = 1/4$$

۵۴) مثال

میزی دارای ۳ کشو است کشو اول ۵ سکه طلا و ۷ سکه نقره کشو دوم دارای ۳ سکه طلا و ۳ سکه نقره کشو سوم دارای ۹ سکه طلا و ۲ سکه نقره میباشد یک کشو بتصادف انتخاب و یک سکه بیرون کشیدیم مشاهده شد طلا است احتمال اینکه از کشو دوم باشد چقدر است

5G,7S 3G,3S 9G,2S

E1=پیش آمد انتخاب کشو اول E2=پیش آمد انتخاب کشو دوم E3=پیش آمد انتخاب کشو سوم

A=پیش آمد سکه طلا

$$P(E2 \downarrow A) = \frac{P(E2).P(A \downarrow E2)}{P(E1).P(A \downarrow E1) + P(E2).P(A \downarrow E2) + P(E3).P(A \downarrow E3)}$$

$$= \frac{(1/3)(3/6)}{(1/3)(5/12) + (1/3)(3/6) + (1/3)(9/11)} =$$

۵۵) مثال

در هفته آینده بدلیل تعمیرات یک واحد از پنج واحد نیروگاهی در مدار است (خروج واحد یعنی خاموشی)

پنج روز هفته آینده مناسبتهایی در سطح شهر است که اهمیت محاسبه احتمال خاموشی را نیاز داریم

در روز شنبه احتمال خروج واحد ۷۰٪ در روز یکشنبه و سه شنبه احتمال خروج واحد ۴۰٪

در روز دوشنبه و چهار شنبه احتمال خروج واحد ۲۰٪ میباشد

در یک روز در هفته آینده مشاهده شد خاموشی بوجود آمده احتمال اینکه این روز دوشنبه باشد چقدر است

$$P(E0) = P(E1) = ... = 1/5 = 0.2$$

$$P(E2 \downarrow A) = \frac{P(E2).P(A \downarrow E2)}{P(E0).P(A \downarrow E0) + P(E1).P(A \downarrow E1) + P(E2).P(A \downarrow E2) + P(E3).P(A \downarrow E3) + P(E4).P(A \downarrow E4)}$$

$$= \frac{(0.2)(0.2)}{(0.2)(0.7) + (0.2)(0.4) + (0.2)(0.2) + (0.2)(0.4) + (0.2)(0.2)} =$$

کامپیوتر و توابع اکسل و آمار و احتمالات

تابع میانگین حسابی	=AVERAGE(محدوده)
ترکیب	=COMBIN(تعداد موفقیت,تعدادآزمایش)
تابع محاسبه دو جمله ای	=BINOMDIST(مقدار منطقی, احتمال موفقیت, تعداد آزمایش , تعدادموفقیت)

۵۶) مثال :

پرتاب یک سکه دو احتمال ۵۰٪ را بدنبال دارد اگر این سکه ۳۰ بار پرتاب شود

کل تعداد حالات که ۱۲ شیر حاصل شود چقدر است

$$\binom{30}{12} = \frac{30!}{(30-12)!12!} = 86493225$$

=COMBIN(30,12)= 86493225

احتمال اینکه ۱۲ بار شیر بیاید چقدر است (مقدار منطقی FALSE)

$$P(x=12) = \binom{30}{12} 0.5^{12} 0.5^{30-12} = 0.08$$

=BINOMDIST(12,30,0.5,FALSE)=0.08

احتمال اینکه حداکثر ۱۲ بار شیر بیاید چقدر است(مقدار منطقی TRUE)

$$P(x \le 12) = \sum_{x=0}^{12} \binom{30}{x} 0.5^{x} 0.5^{30-x} = 0.18$$

=BINOMDIST(12,30,0.5,TRUE)=0.18

☜

تابع محاسبه ضریب همبستگی	=intercept(مجموعه وابسته , مجموعه مستقل)

۵۷) مثال

ضریب همبستگی بین دو سری اعداد زیر بدست آورید

x = 1,3,4,6,8,9,11,14
y = 1,2,4,4,5,7, 8, 9
$y = a + bx$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \quad , \quad a = \bar{y} - b\bar{x} \quad \Rightarrow y = a + bx$$

y = a+bx
b = 7/11=0.636
Xavg=7
Yavg=5
a = 5-(7/11)*7=6/11
y = 6/11+(7/11)x
y=0.545+0.637x

اگر سری مستقل x را از a1 تا a8 نوشته و سری وابسته y را از b1 تا b8 آنگاه

=intercept(b1:b8; a1:a8)=0.545

تابع محاسبه فاصله میانگین جامعه	=confidence(آلفا , انحراف معیار , اندازه نمونه)

۵۸) مثال

با آلفا ۵٪ و انحراف معیار ۲,۵ و اندازه نمونه ۵۰ فاصله اطمینان میانگین جامعه را حساب کنید
یک منهای آلفا بیانگر ۹۵٪ اطمینان است

=confidence(0.05,2.5,50)=0.69

تابع محاسبه توزیع دو جمله ای تجمعی مشروط	=critbinom(تعداد تلاشها , احتمال موفقیت , آلفا)

۵۹) مثال

سکه ای را ۱۰ بار پرتاب میکنیم احتمال شیر در هر بار ۵۰٪ میباشد برای اینکه با اطمینان ۷۵٪ بگوییم شیر بیاید حداقل چند پرتاب کنیم

=critbinom (10,0.5,0.75)=6

تابع محاسبه واریانس نمونه	=var (محدوده)

1,3,4,6,8,9,11,14
=var(1,3,4,6,8,9,11,14)=18.85

تابع محاسبه واریانس جامعه	=varp (محدوده)

1,3,4,6,8,9,11,14
=varp(1,3,4,6,8,9,11,14)=16.5

٦٠) مثال کمک مثال ٤٨

ظرفی محتوی پنج گوی ۳ گوی قرمز و ۲ گوی سیاه است تمام گویها شماره گذاری شده است
ازدرون ظرف یک گوی را بیرون میاوریم
الف) احتمال قرمز بودن؟ ب) احتمال قرمز شماره ۲ بودن؟
از درون ظرف دو گوی را یکی یکی بدون جایگزینی بیرون میاوریم
ج)احتمال اینکه گوی اول قرمز و گوی دوم سیاه باشد چقدر است
د) احتمال گوی اول قرمز شماره۲ و گوی دوم سیاه شماره ۱ باشد
ه) احتمال اینکه یک گوی قرمز و یک گوی سیاه باشد
و) احتمال اینکه یک گوی قرمز شماره ۲ و یک گوی سیاه شماره ۱ باشد
از درون ظرف یک گوی بیرون آورده و به آن نگاه کرده بداخل ظرف باز میگردانیم سپس گوی دوم را بیرون میاوریم و نگاه میکنیم (با جایگزینی)
ح)احتمال اینکه گوی اول قرمز و گوی دوم سیاه باشد چقدر است
ط) احتمال گوی اول قرمز شماره۲ و گوی دوم سیاه شماره ۱ باشد
ی) احتمال اینکه یک گوی قرمز و یک گوی سیاه باشد
ک) احتمال اینکه یک گوی قرمز شماره ۲ و یک گوی سیاه شماره ۱ باشد

بدون جایگزینی
الف) احتمال قرمز بودن

$S = \{R1, R2, R3, B1, B2\}$

$A = \{R1, R2, R3\}$

$P = 3/5$

ب)احتمال قرمز بودن شماره ۲

$S = \{R1, R2, R3, B1, B2\}$

$A = \{R2\}$

$P = 1/5$

ج)از درون ظرف دو گوی را یکی یکی بدون جایگزینی بیرون میاوریم

S={G1G2,G1G3,G1B1,G1B2 G2G1,G2G3,G2B1,G2B2 G3G1,G3G2,G3B1,G3B2 B1G1,B1G2,B1G3,B1B2 B2G1,B2G2,B2G3,B2B1 }

S=20

احتمال اینکه گوی اول قرمز و گوی دوم سیاه باشد چقدر است
از روی دو مجموعه فوق کلیه حالتهای **GB**

A={G1B1,G1B2,G2B1,G2B2,G3B1,G3B2}
A=6
P(A)=(A/S)=6/20=3/10

روش دیگر

$S1 = \{R1, R2, R3, B1, B2\}$

$A1 = \{R1, R2, R3\}$

$P(R) = 3/5$

$S2 = \{R1, R2, B1, B2\}$

$A2 = \{B1, B2\}$

$P(B) = 2/4$

$$OR \quad P(B) = P(B|R) = \frac{P(B \cap R)}{P(R)} = \frac{P(B).P(R)}{P(R)} = \frac{(2/4).(3/5)}{3/5} = 2/4$$

P(RandB)=(3/5)(2/4)=6/20=3/10

د) احتمال اولی قرمز شماره ۲ دومی سیاه شماره ۱ باشد

از روی دو مجموعه فوق

A={R2B1}
A=1
P(A)=A/S=1/20=0.05

راه دیگر

P(R2andB1)=(1/5).(1/4)=1/20

ه) احتمال اینکه یک گوی قرمز و یک گوی سیاه باشد

از روی مجموعه فوق

A=12
A={R1B1,R1B2,R2B1,R2B2,R3B1,R3B2,B1G1,B1G2,B1G3,B2G1,B2G2,B2G3}
P(A)=A/S=12/20=0.6

راه دیگر

P(RandB) U P(BandR)=(3/5)(2/4)+(2/5)(3/4)=12/20

روش دیگر (ه)

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

$$P[(R \cap B) \cup (B \cap R)] = P(R \cap B) + P(B \cap R) = P(R \cap B) + P(B \cap R) =$$

$$P(R).P(B|R) + P(B).P(R|B) = [(3/5).(2/4)] + [(2/5).(3/4)]$$

میتوانستیم یک حالت بدست آورده در ۲ ضرب کنیم

روش دیگر (ه) برای حالت ه میتوانستیم از فرمول ترکیب استفاده کنیم

$$\frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}}$$

و) احتمال یک گوی قرمز شماره ۲ و یک گوی سیاه شماره ۱ باشد

از روی مجموعه فوق

A={R2B1,B1R2}
A=2
P(A)=A/S=2/20=0.1

روش دیگر

P(R2andB1) U P(B1andR2) = (1/5).(1/4) + (1/5).(1/4)=2/20=1/10

با جایگزینی

از درون ظرف دو گوی را یکی یکی با جایگزینی بیرون میاوریم

S={ G1G1,G1G2,G1G3,G1B1,G1B2 G2G1,G2G2,G2G3,G2B1,G2B2
G3G1,G3G2,G3G3,G3B1,G3B2 B1G1,B1G2,B1G3,B1B1,B1B2
B2G1,B2G2,B2G3,B2B1,B2B2 }

ز)احتمال اینکه گوی اول قرمز و گوی دوم سیاه باشد چقدر است

از روی مجموعه فوق

A={G1B1,G1B2,G2B1,G2B2,G3B1,G3B2 }
A=6
P(A)=6/25

روش دیگر

$S1 = \{R1, R2, R3, B1, B2\}$

$A1 = \{R1, R2, R3\}$

$P(R) = 3/5$

$S2 = \{R1, R2, R3, B1, B2\}$

$A2 = \{B1, B2\}$

$P(B) = 2/5$

$$OR \quad P(B) = P(B|R) = \frac{P(B \cap R)}{P(R)} = \frac{P(B).P(R)}{P(R)} = \frac{(2/5).(3/5)}{3/5} = 2/5$$

P(RandB)=(3/5)(2/5)=6/25

د) احتمال اولی قرمز شماره ۲ و دومی سیاه شماره ۱ باشد

A={R2B1}
A=1
P(A)=A/S=1/25=0.04

روش دیگر

P(R2andB1)=(1/5).(1/5)=1/25

ه) احتمال اینکه یک گوی قرمز و یک گوی سیاه باشد

A={G1B1,G1B2,G2B1,G2B2,G3B1,G3B2,B1G1,B1G2,B1G3,B2G1,B2G2,B2G3}
A=12
P(A)=A/S=12/25=0.48

روش دیگر

P(RandB) U P(BandR)=(3/5)(2/5)+(2/5)(3/5)=12/25

روش دوم ه

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

$$P[(R \cap B) \cup (B \cap R)] = P(R \cap B) + P(B \cap R) = P(R \cap B) + P(B \cap R) =$$

$$P(R).P(B|R) + P(B).P(R|B) = [(3/5).(2/5)] + [(2/5).(3/5)]$$

میتوانستیم یک حالت بدست آورده در ۲ ضرب کنیم

و) احتمال یک گوی قرمز شماره ۲ یک گوی سیاه شماره ۱ باشد

A={R2B1,B1R2}

A=2
P(A)=A/S=2/25=0.08

روش دیگر

P(R2andB1) U P(B1andR2) = (1/5).(1/5) + (1/5).(1/5)=2/25=0.08

فرمولهای احتمال در مجموعه ها

$P(\Phi) = 0$

$P(A') = 1 - P(A)$

$P(A) = P(A \cap B) + P(A \cap B')$

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

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

$IF \;\; A \cap B = \Phi \quad P(AandB) = P(A).P(B)$

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

$IF \;\; B \subset A \quad P(A - B) = P(A) - P(B)$

پایان

در هر حرفه ای که هستید نه اجازه دهید که به بدبینیهای بیحاصل آلوده شوید و نه بگذارید که بعضی لحظات تاسف بار که برای هر ملتی پیش می آید شما را به یاس و ناامیدی بکشاند. در آرامش حاکم بر آزمایشگاهها و کتابخانه هایتان زندگی کنید .

نخست از خود بپرسید : "***برای یادگیری و خودآموزی چه کرده ام ؟*** "

سپس همچنان که پیشتر میروید بپرسید : "***من برای کشورم چه کرده ام ؟*** "

و این پرسش را آنقدر ادامه دهید تا به این احساس شادیبخش و هیجان انگیز برسید که شاید سهم کوچکی در پیشرفت و اعتلای بشریت داشته اید.

اما هر پاداشی که زندگی به تلاشهایمان بدهد یا ندهد هنگامی که به پایان تلاشهایمان نزدیک میشویم هر کداممان باید حق آن را داشته باشیم که با صدای بلند بگوییم

"***من آنچه در توان داشته ام انجام داده‌ام*** "

لوئی پاستور ۱۸۹۵-۱۸۲۲