روش‌های کمی مقدماتی

مدرس :

سید بابک ابراهیمی

دانش‌آموخته ممتاز دانشگاه صنعتی شریف

مدیر مطالعات اقتصادی سازمان اقتصادی کوثر

مشاور مرکز مالی ایران

بخش اول: مفاهیم آماری

تعریف آمار (Statistics)

روش علمی است که برای جمع‌آوری، تلخیص، تجزیه و تحلیل، تفسیر و بطور کلی برای مطالعه و بررسی مشاهدات بکار گرفته می‌شود.

از فنون آماری در مدیریت برای چه مقاصدی استفاده می‌شود ؟

۱- برای تبدیل داده‌ها به اطلاعات
۲- برای بررسی صحت و سقم فرضیات
۳- برای تعیین اعتبار و پایایی تحقیقات

سیر تحول علم آمار از نظر موضوعی عبارتند از :

روش‌های آماری کلاسیک را می‌توان در دو گروه آمار **توصیفی** و آمار **استنباطی** دسته‌بندی نمود. (در حالت کلی سه دسته)

۱- آمار توصیفی
۲- آمار استنباطی
۳- آمار ناپارامتریک (غیرکلاسیک)

مهم

آمار توصیفی (Descriptive Statistics)

یعنی محاسبه مقادیر و شاخص‌های جامعه آماری با استفاده از **سرشماری** تمامی عناصر آن.
(**توصیف کل جامعه از طریق محاسبه پارامترها**) سرشماره عمومی نفوس و مسکن

آمار استنباطی (Statistical Inference)

آماری که در آن محقق ابتدا **آماره‌ها** را محاسبه و سپس به کمک تخمین و آزمون فرض آماری، آن‌ها را به پارامترهای‌جامعه تعمیم می‌دهد. (محاسبه میانگین قد دانش‌آموزان با استفاده از محاسبه میانگین قد دانش‌آموزان ۲۲ مدرسه از ۲۲ منطقه تهران)

آمار ناپارامتریک

این نوع آمار در مقابل آمار پارامتریک یعنی آمارهای توصیفی و استنباطی دارای توزیع نرمال قرار می‌گیرد و برای مشاهدات **فاقد توزیع آماری** کاربرد دارد. (اغلب متغیرها دارای مقیاس کیفی هستند.)

تعریف جامعه

تعریف اول: جامعه بزرگترین مجموعه از موجودات است که در یک زمان معین، مطلوب ما قرار می‌گیرند.

تعریف دوم: جامعه شامل تمامی اعضای یک مجموعه مشخص می‌باشد.

مثل جامعه فرهنگیان ایران، جامعه شرکت‌های پتروشیمی بورس

جامعه آماری

تعدادی از عناصر مطلوب موردنظر که حداقل دارای یک صفت مشخصه باشند.

صفت مشخصه

صفتی است که بین همه عناصر جامعه آماری مشترک و متمایزکننده جامعه آماری از سایر جوامع باشد.

انواع جامعه آماری

محدود: یعنی جامعه مقادیر از تعداد محدود و ثابتی تشکیل شده و پایان‌پذیر باشد.

نامحدود: یعنی جامعه از یک ردیف بی‌انتهایی از مقادیر تشکیل شده باشد. (مجموعه اعداد صحیح)

تعریف نمونه (Sample)

نمونه عبارت است از تعداد محدودی از آحاد جامعه آماری که بیان‌کننده ویژگی‌های اصلی جامعه باشد. (**نمونه یک زیرمجموعه از جامعه است که برای مطالعه صفت خاصی از جامعه آماری انتخاب می‌شود**).

(شرکت‌های پتروشیمی با سرمایه بالای ۴۰۰۰میلیارد حاضر بورس)

مهم

انواع شاخص‌های آماری

پارامتر: شاخص‌هایی که از طریق سرشماری (اندازه‌گیری تمامی عناصر جامعه آماری) بدست می‌آیند. (نمایش با حرف بزرگ)

(**نکته**: پارامترهای جامعه مقادیری ثابت و پایدار هستند و تا زمانی که خود جامعه تغییر نکند، پارامتر جامعه تغییر **نمی‌کند**.)

آماره: شاخص‌هایی که از طریق داده‌های نمونه‌گیری (اندازه‌گیری بخشی از جامعه) بدست می‌آیند. (نمایش با حرف کوچک یا علامت بار)

(**نکته**: آماره مقداری متغیر و ناپایدار است، بدین معنی که از یک نمونه دیگر تغییر می‌کند و به همین دلیل اگر نمونه موردنظر یک نمونه تصادفی باشد، آماره را یک **متغیر تصادفی** می‌گویند. به آماره برآوردکننده یا تخمین‌زننده نیز می‌گویند.)

مقایسه پارامتر و آماره

شاخص	گروه	نماد کلی	میانگین	واریانس	نسبت
آماره	نمونه	$\hat{\theta}$	$\overline{X}$	S^2	$\overline{P}$
پارامتر	جامعه	θ	μ	σ^2	P

نقش متغیرها در فرضیات

متغیرها، فرضیه‌ها را بصورتی نشان می‌دهند که محققان رفتاری و مدیریتی بتوانند آن‌ها (فرضیه‌ها) را مشاهده و اندازه‌گیری نمایند.

انواع متغیرها

۱- متغیر خصیصه
۲- متغیر مستقل
۳- متغیر وابسته
۴- متغیر تعدیل‌کننده (واسطه‌ای)
۵- متغیر کنترل

متغیر خصیصه

متغیری که مقدار آن از یک فرد به فرد دیگر و یا از یک عضو به عضو دیگر جامعه آماری ممکن است تغییر کند. مثل اندازه سازمان

متغیر مستقل

متغیری است که توسط پژوهش‌گر اندازه‌گیری، دستکاری یا انتخاب می‌گردد تا تاثیر یا رابطه آن با متغیر دیگر (متغیر وابسته) اندازه‌گیری شود.
به متغیر مستقل متغیر درونداد و یا به عبارتی متغیر محرک گفته می‌شود.

مثال: تحقیقی با موضوع بررسی میزان ارتباط نوشیدن چای با ابتلا به سرطان معده را در نظر گرفت. در این تحقیق متغیر مستقل **نوشیدن چای** و متغیر وابسته **سرطان معده** می باشد.

متغیر وابسته

متغیری است که ارزش یا مقدار آن به متغیر مستقل بستگی دارد. متغیر وابسته در اختیار محقق نیست و محقق نمی‌تواند در آن دخل و تصرف یا دستکاری کند.

متغیر وابسته، متغیری است که مشاهده یا اندازه‌گیری می‌شود تا تاثیر متغیر مستقل بر آن معلوم و مشخص شود.

متغیر وابسته، **متغیر پاسخ** و یا **برونداد** اطلاق می‌شود.

متغیر تعدیل‌کننده (واسطه‌ای)

یک متغیر ثانوی است که رابطه بین متغیر مستقل و متغیر وابسته را تحت تأثیر قرار می‌دهد.

متغیر کنترل

چون در یک پژوهش اثرات همه متغیرها قابل بررسی نیست، پژوهشگر اثرات برخی متغیرها را از طریق کنترل آماری خنثی می‌کند. اینگونه متغیرها که اثرات آنها توسط پژوهشگر قابل حذف است را متغیر کنترل گویند

فرق متغیر تعدیل‌کننده با متغیر کنترل

موقع انجام تحقیق، پژوهشگر سعی می‌کند تأثیرات متغیر کنترل را از بین ببرد ولی تأثیرات متغیر تعدیل کننده را مورد بررسی قرار می‌دهد.

مقیاس اندازه‌گیری متغیرها

علاون بر نوع متغیرها، مقیاس اندازه‌گیری متغیرها از اهمیت ویژه‌ای برخوردار است:

۱- مقیاس اسمی (Nominal scale)
۲- مقیاس ترتیبی (Ordinal scale)
۳- مقیاس فاصله‌ای (Interval scale)
۴- مقیاس نسبی (Ratio scale)

مقیاس اسمی یا طبقه‌ای

محققان از این مقیاس، صرفاً برای **طبقه‌بندی** اشیاء، اشخاص و یا خصوصیات استفاده می‌کنند.

این مقیاس عناصر جامعه را فقط گروه‌بندی می‌کند به طوری که تمامی عناصری که صفت مشترک دارند در یک گروه قرار می‌گیرند.

این نوع مقیاس به علت ضعف در اندازه‌گیری، در صفات کیفی استفاده می‌شود.

مثال: مشخص‌کردن سازمان‌ها با اسم‌های A و B و...

A: سازمان‌های دارای مدیریت منابع انسانی

B: سازمان‌های فاقد مدیریت منابع انسانی

مقیاس رتبه‌ای یا ترتیبی

اگر بین اسامی ایجاد شده یا طبقات حاصله ناشی از مقیاس‌بندی اسمی یک نوع رابطه هم وجود داشته باشد پژوهشگران از مقیاس ترتیبی استفاده می‌نمایند.

برای مثال طبقه بندی افراد جامعه به صورت (پردرآمد ـ متوسط ـ کم درآمد)، (قوی ـ متوسط ـ ضعیف)، (بالا ـ وسط ـ پایین)، (بزرگتر ـ مساوی ـ کوچکتر) نشان‌دهنده حالت ترتیبی است.

این نوع مقیاس نیز به علت ضعف در اندازه‌گیری، در **صفات کیفی** مورد استفاده قرار می‌گیرد.

مقیاس فاصله‌ای

اگر در مقیاس ترتیبی، فاصله بین اعداد یا طبقات از یک نظم خاصی پیروی نماید (فواصل یکسان باشند) محققان از مقیاس فاصله‌ای برای اندازه‌گیری متغیرها استفاده می‌نمایند.

این مقیاس علاوه بر دارا بودن ویژگی‌های دو مقیاس اسمی و رتبه‌ای یعنی گروه‌بندی عناصر جامعه و تعیین رتبه آن‌ها، واحدهای عددی اندازه‌گیری را با فاصله‌های یکسان به کار می‌برد.

نسبت هر دو فاصله مستقل از واحد اندازه‌گیری و مستقل از نقطه صفر است.

این مقیاس دارای صفر **قراردادی** است. **مثل**: *سانتی‌گراد و فارنهایت*

مقیاس نسبی یا کسری

مقیاسی است که علاوه بر داشتن همه خصوصیات مقیاس فاصله‌ای، دارای نقطه صفر واقعی نیز هست. **مثل**: پوند،گرم، متر، اینچ

دقیق‌ترین مقیاس برای صفات کمی است.

در مقیاس نسبی، صفر به معنای «هیچ» است وجود صفر مطلق این امکان را فراهم می‌کند که **نسبت** اندازه‌ها معنی‌دار باشد.

برای مثال وزن دو شی مختلف را می‌توان هم با گرم و هم با پوند اندازه‌گیری کرد. اما نسبت آن‌ها با هم فرقی نمی‌کند و به واحد اندازه‌گیری ربطی ندارد.

جدول مقادیر مقیاس‌های چهارگانه

مراتب / مقیاس	ترتیب	فواصل	صفر قراردادی	صفر مطلق (واقعی)
اسمی	ندارد	ندارد	ندارد	ندارد
رتبه‌ای	دارد	ندارد	ندارد	ندارد
فاصله‌ای	دارد	داد	دارد	ندارد
نسبی	دارد	دارد	دارد	دارد

توزیع فراوانی

سازماندهی مشاهدات را در آمار، توزیع فراوانی گویند. به عبارتی دیگر، توزیع فراوانی جدولی خلاصه شده از داده‌های جمع‌آوری شده از جامعه آماری است.

مراحل طبقه‌بندی داده‌ها

۱- مرتب‌کردن داده‌ها و محاسبه **دامنه تغییرات** (R)

۲- مشخص‌کردن **تعداد طبقات** (K)

۳- محاسبه‌نمودن **فاصله طبقات** (I)

۴- سازماندهی طبقات

دامنه تغییرات (R)

به تفاضل بین مقادیر **حداکثر** و **حداقل** مشاهده جامعه (نمونه) دامنه تغییرات گفته می‌شود.

فرمول:
$$R = x_{\max} - x_{\min}$$

فرمول‌های محاسبه تعداد طبقات (K)

۱- فرمول تجربی استورجس $K = 1 + 3/32 \log N$

۲- **روش تقریبی** (رو به بالا گرد می‌کنیم) $K = \sqrt{N}$

(N تعداد مشاهدات می‌باشد.)

تعیین فاصله طبقات (I or C)

فاصله طبقات از تقسیم مقدار R (دامنه تغییرات) بر مقدار محاسبه شده برای تعداد طبقات (K) بدست می‌آید.

$$I = C = \frac{R}{K} = \frac{\text{دامنه تغییرات}}{\text{تعداد طبقات}}$$

سازماندهی داده‌ها

پس از مشخص‌شدن K و I سازماندهی یعنی تعیین نوع جدول و شیوه طبقه‌بندی داده‌ها شروع می‌شود که این بستگی به نوع داده‌های جمع‌آوری شده دارد.

مثال

فرض کنید اعداد زیر نشان‌دهنده بازده یک سهم (بر حسب درصد) در ۱۲سال مختلف است.
این داده‌ها به ترتیب صعودی مرتب شده‌اند (این کار برای تشکیل جداول و نمودارهای بعدی ضروری است):

11.43 و 7.16 و 4.68 و 4.42 و 4.28 و 2.38 و 2.35 و 1.34 و 0.28 و 1.64- و 4.04- و 4.57-

$$R = X_{max} - X_{min} = 11.43 - (-4.57) = 16$$

فرض کنید، می‌خواهیم ۴ بازه (طبقه) مختلف داشته باشیم و اعداد را در آن‌ها دسته‌بندی کنیم. (K=4)

$$\text{طول طبقه} = \frac{۱۶}{۴} = ۴$$

از کوچکترین عدد شروع کرده و هر بار به آن چهار واحد اضافه می‌کنیم تا مرز بازه‌ها (طبقه‌ها) به دست آیند.

-۴,۵۷   -۰,۵۷   ۳,۴۳   ۷,۴۳   ۱۱,۴۳

+۴   +۴   +۴   +۴

ادامه مثال و تعریف دو مفهوم جدید

بازه	مرزهای بازه (طبقه)	فراوانی مطلق
A	$-۴.۵۷ \leq X < -۰.۵۷$	۳
B	$-۰.۵۷ \leq X < ۳.۴۳$	۴
C	$۳.۴۳ \leq X < ۷.۴۳$	۴
D	$۷.۴۳ \leq X \leq ۱۱.۴۳$	۱

$\boxed{F_i}$

فراوانی مطلق : تعداد دفعات تکرار یک داده (داده‌های مربوط به هر طبقه) را در اصطلاح فراوانی مطلق آن داده (یا آن طبقه) گویند و حاصل جمع فراوانی‌های مطلق برابر **حجم جامعه (نمونه)** است.

$$\boxed{\sum F_i = N}$$

ادامه مثال و تعریف دو مفهوم جدید

بازه	فراوانی مطلق	فراوانی نسبی
A	۳	۲۵٪
B	۴	۳۳,۳٪
C	۴	۳۳,۳٪
D	۱	۸,۳٪

f_i

فراوانی نسبی : تعداد دفعات تکرار یک داده (داده‌های مربوط به هر طبقه) را نسبت به تمامی داده‌ها (داده‌های متعلق به تمامی طبقات) به عبارتی دیگر اگر **فراوانی مطلق** هر داده (طبقه) را بر **حجم جامعه** تقسیم کنیم فراوانی نسبی آن داده (طبقه) به دست می‌آید.

$\sum f_i = ۱$

$f_i = \frac{F_i}{N}$

به کمک این فراوانی می‌توان درصد تراکم داده‌ها را در هر طبقه مشخص نمود بعبارتی از فراوانی نسبی جهت یافتن محل تمرکز داده‌ها استفاده می‌شود.

چند نکته کاربردی

- **مرکز یک طبقه** برابر نصف مجموع حد پائین و حد بالای آن طبقه است. که به آن **نماینده طبقه** یا **متوسط طبقه** نیز می‌گویند.

طول طبقه یا **دسته** تفاوت بین حدود بالا یا پایین **دو** طبقه **متوالی** است.

طبقات	۰ ـ ۵	۵ ـ ۱۰	۱۰ ـ ۱۵
فراوانی	۳	۴	۱۳

در طبقه اول ۵ حد بالا و ۰ حد پائین را تشکیل می‌دهد.

از طرفی $\frac{0+5}{2}=2.5$ مرکز طبقه اول، $\frac{5+10}{2}=7.5$ مرکز طبقه دوم، $\frac{10+15}{2}=12.5$ مرکز طبقه سوم است.

طول طبقات ۵ است، چرا که ۵ = ۰ − ۵، ۵ = ۵ − ۱۰ یا ۵ = ۱۰ − ۱۵ می‌باشد.

چند نکته کاربردی

پیوسته‌کردن جدول با طبقات ناپیوسته

در جداول با طبقات گسسته عبارت:

$$\frac{\text{حدبالای طبقه} - \text{حدپایین طبقه بعدی}}{2}$$

محاسبه شده و مقدار محاسبه شده را از حد پایین طبقات کم و به حد بالای طبقات اضافه کنیم.

طبقات	۱۴،۵-۱۹،۵	۹،۵-۱۴،۵	۴،۵-۹،۵
فراوانی	۴	۷	۳

جدول طبقه‌بندی شده پیوسته

طبقات	۱۵-۱۹	۱۰-۱۴	۵-۹
فراوانی	۴	۷	۳

جدول طبقه‌بندی شده ناپیوسته

±۰.۵

فراوانی تجمعی و فراوانی نسبی تجمعی

Cumulative Frequency

فراوانی تجمعی (F_{ci}): فراوانی تجمعی هر داده (طبقه) برابر است با جمع فراوانی مطلق همان طبقه بعلاوه فراوانی‌های مطلق طبقات ماقبل آن:

$$F_{ci} = F_1 + F_2 + \dots + F_i$$

نکته: فراوانی تجمعی طبقه آخر برابر با حجم کل جامعه است.

Cumulative Relative Frequency

فراوانی نسبی تجمعی (f_{ci}) درصد مشاهدات واقع شده بین حد پایین اولین طبقه و حد بالای i امین طبقه را نشان می‌دهد.

$$f_{ci} = \frac{F_{ci}}{N}$$

نکته: فراوانی نسبی تجمعی آخرین طبقه یک است.

مثال(۱)

بازه	فراوانی مطلق	فراوانی نسبی	فراوانی تجمعی	فراوانی نسبی تجمعی
A	۳	۲۵٪	۳	۲۵٪
B	۴	۳۳,۳٪	۴+۳=۷	۲۵٪+۳۳,۳٪=۵۸,۳٪
C	۴	۳۳,۳٪	۷+۴=۱۱	۵۸,۳٪+۳۳,۳٪=۹۱,۷٪
D	۱	۸,۳٪	۱۱+۱=۱۲	۹۱,۷٪+۸,۳٪=۱۰۰٪

عدد ۱۱ در ستون فراوانی تجمعی و در مقابل ردیف C نشان می‌دهد که ۱۱ داده از ۱۲ داده فوق در یکی از بازه‌های A، B و یا C قرار دارند. به بیان دیگر ۱۱ داده از ۱۲ داده فوق در بازه $-۴.۵۷ \leq X \leq ۷.۴۳$ قرار دارند. فراوانی نسبی تجمعی نیز همین مفهوم را به صورت درصد بیان می‌کند: ۹۱,۷٪ از داده‌ها در یکی از سه بازه A، B و یا C قرار دارند یا به بیان دیگر ۹۱,۷٪ از داده‌ها در بازه (۷,۴۳ و ۴,۵۷-] قرار دارند.

مثال(۲)

به جدول زیر توجه کنید؛

C-L	۰_۵	۵_۱۰	۱۰_۱۵	۱۵ _ ۲۰	
F_i	۱۰	۲۰	۳۰	۴۰	$N=\sum_{i=1}^{k} F_i = 100$
F_{ci}	۱۰	۳۰	۶۰	۱۰۰	۱۰۰ = N = حجم کل جامعه = فراوانی تجمعی طبقه آخر
f_i	$\frac{10}{100}$	$\frac{20}{100}$	$\frac{30}{100}$	$\frac{40}{100}$	$\sum_{i=1}^{k} F_i = \frac{10+20+30+40}{100} = 1$
f_{ci}	$\frac{10}{100}$	$\frac{30}{100}$	$\frac{60}{100}$	$\frac{100}{100}$	۱ = فراوانی نسبی تجمعی آخرین طبقه

آشنایی با نمودارها

۱- نمودارهای کمی: مخصوص داده‌هایی با مقیاس فاصله‌ای و نسبی
۲- نمودارهای وصفی: مخصوص داده‌هایی با مقیاس اسمی و یا رتبه‌ای

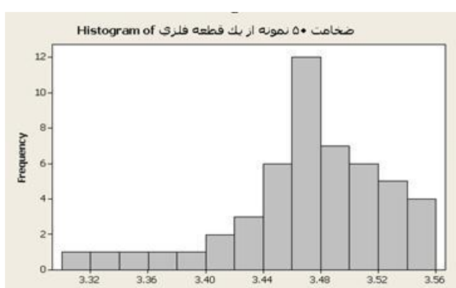
مهم‌ترین نمودارهای کمی

۱- بافت نگار (هیستوگرام)
۲- چندضلعی (پلی‌گون)
۳- فراوانی تجمعی (اُجایو)
۴- تحلیل اکتشافی داده‌ها
۴-۱- نمودار شاخه و برگ
۴-۲- نمودار جعبه‌ای

هیستوگرام یا بافت‌نگار

هیستوگرام نموداری است در دستگاه مختصات که محور افقی آن با **حدود واقعی طبقات** و محور عمودی آن با **فراوانی مطلق یا نسبی** درجه‌بندی می‌شود.

پس از مدرج‌کردن محورها بر روی حدود واقعی (کرانه‌های هر طبقه) مستطیلی عمودی رسم می‌شود که **مساحت** آن مساوی با **فراوانی نسبی آن طبقه** می‌باشد.



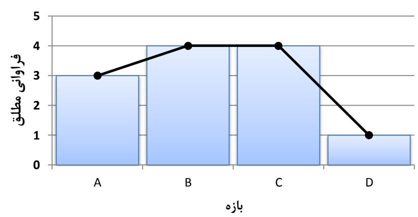
در این نمودار، **مرکزیت** و **پراکندگی داده‌ها** مورد توجه قرار می‌گیرد ولی امکان شناسایی نقاط انفرادی داده‌ها را فراهم نمی‌کند. چون در این نمودار مشاهدات قرار گرفته در یک طبقه از هم قابل تمیز نیستند.

نمودار چندضلعی یا پلی‌گون

نموداری است که نقطه میانی هر طبقه (مرکز هر طبقه) روی محور افقی و فراوانی مطلق یا نسبی روی محور عمودی آن نشان داده می‌شود.

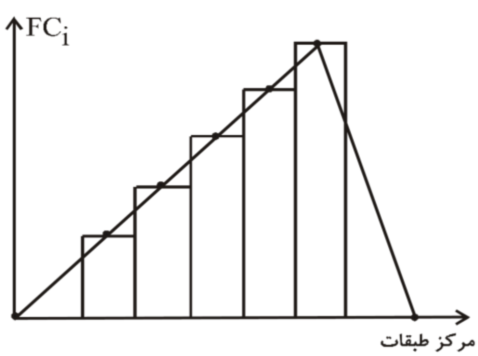
هر یک از رئوس نمودار **چندضلعی** همان نقطه میانی ضلع بالایی مستطیل‌های **هیستوگرام** مربوط به جدولی فراوانی مورد نظر است.

اگر بخواهیم دو یا چند توزیع را به صورت هندسی با هم مقایسه کنیم، رسم نمودار چند ضلعی انتخاب مناسبی است.



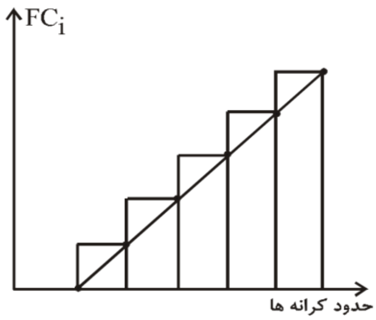
پلی‌گون فراوانی تجمعی

برای رسم این نمودار محور عمودی (yها) براساس فراوانی تجمعی طبقات و محور افقی (Xها) بر اساس مرکز طبقات یا متوسط طبقات مدرج می‌شود. در این نمودار ارزش همه مقادیر هر طبقه یکسان فرض می‌شود. بنابراین نقطه میانی هر طبقه با فراوانی تجمعی آن طبقه در نظر گرفته می‌شود و سپس آن را به هم متصل می‌کند.



منحنی فراوانی تجمعی

تنها فرق این نمودار با نمودار پلی‌گون فراوانی تجمعی در این است که در این نمودار بجای نماینده طبقات از حد بالای کرانه‌ها استفاده می‌شود.



کاربردهای نمودار فراوانی تجمعی

۱- برای محاسبه چندک‌ها (چارک‌ها، دهک‌ها، صدک‌ها)
۲- برای مقایسه پدیده‌ها (مثل میزان رشد تورم در کشورها)

نمودارهای وصفی

این دسته از نمودارها برای نمایش هندسی داده‌های کیفی بکار می‌روند، در این نمودارها هر یک از مقادیر بعنوان یک طبقه در نظر گرفته می‌شوند.

مهم‌ترین نمودارهای وصفی

۱- نمودار ستونی
۲- نمودار دایره‌ای

نمودار ستونی

این نمودار در یک دستگاه مختصات که محور افقی نشان‌دهنده کیفیت مشاهدات و محور عمودی نشان‌دهنده فراوانی مطلق یا نسبی هر گروه است ترسیم می‌شود.

نمودار دایره‌ای

این نمودار ابزار مناسبی برای تجسم مشاهدات بوده و معمولاً بر حسب درصد تهیه می‌شود و به نمودار کلوچه‌ای نیز معروف است.

مراحل تهیه نمودار دایره‌ای

۱- تبدیل فراوانی مطلق به **نسبی**

۲- پیداکردن مساحت هر قطاع از دایره

$$S_i = 360 \times f_i$$

۳- تقسیم مساحت دایره بر حسب S_iها

۴- نوشتن نوع و درصد مشاهدات بر روی دایره

آزمون(۱)

بخش دوم:شاخص‌های مرکزی

آشنایی با پارامترهای مرکزی در جوامع کوچک ($N \leq ۲۰$)

پارامترهای مرکزی

به هر معیار عددی که معرف مرکز مجموعه داده‌ها باشد، پارامتر مرکزی اطلاق می‌شود.

به عبارت دیگر یک مقدار عددی که بتواند ویژگی‌های کلی داده‌های آماری را داشته باشد و معرف مطلوبی برای تمامی داده باشد در اصطلاح یک معیار مرکزی یا یک پارامتر مرکزی می‌باشد که معمولا در حوالی مرکز منحنی فراوانی قرار گرفته‌اند.

۱- میانگین (Mean)
۲- میانه (Median)
۳- مد (نما)
۴- چارک‌ها (Quantiles)

میانگین

- **میانگین حسابی ساده**

این میانگین از تقسیم مجموع مشاهدات بر تعداد آن‌ها بدست می‌آید:

$$\mu = \frac{\sum x_i}{N} = \frac{x_1 + x_2 + \ldots + x_N}{N} \qquad \bar{X} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

میانگین جامعه

میانگین نمونه

میانگین

- **میانگین حسابی موزون**

اگر هر یک از مشاهدات دارای تکرار باشند، در این صورت تعداد تکرارها بعنوان وزن مشاهدات تلقی شده و آن‌ها را با w_i یا Fi نشان می‌دهند.

$$\mu = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{N} = \frac{\sum_{i=1}^{n} F_i x_i}{N} = \sum_{i=1}^{n} f_i x_i$$

برای محاسبه میانگین حسابی در داده‌های طبقه بندی شده نیز این فرمول مهم است.

میانگین

- **میانگین پیراسته**

از این میانگین زمانی استفاده می‌شود که در توزیع مشاهدات، تعداد اندکی از آن‌ها، با بقیه داده‌ها همخوانی و تجانس نداشته باشد (داده‌های پرت).

طرز بدست آوردن میانگین پیراسته

۱- مرتب‌کردن **صعودی** داده‌ها

۲- **حذف** تمام مشاهدات کوچکتر از LN % پایین و بزرگتر از LN% بالا

۳- محاسبه میانگین برای باقیمانده مشاهدات

میانگین پیراسته داده‌های ۳۹، ۳۸، ۱۴، ۱۳، ۱۲، ۱۰، ۲، ۲، ۱ کدام است؟

$$\mu = \frac{\sum X_i}{N} = \frac{۱۰+۱۲+۱۳+۱۴}{۴} = ۱۲/۲۵$$

میانگین

- **میانگین هندسی ساده**

از این میانگین برای محاسبه اندازه‌های نسبی مانند: نسبت‌ها، درصدها، شاخص‌ها و نرخ‌های رشد استفاده می‌شود.

برای اعداد مطلق

$$\mu_G = \sqrt[n]{X_1 \times X_2 \times \cdots \times X_n} = (X_1 \times X_2 \times \cdots \times X_n)^{\frac{1}{n}}$$

میانگین هندسی یک رشته عدد برابر است با ریشه Nام حاصلضرب آن اعداد

اگر داده‌ها تحت عنوان نرخ رشد یا نرخ بازده بر حسب درصد مطرح شده باشند می‌توانیم هر داده دلخواه x_i را به صورت $(1 + \frac{x_i}{100})$ نوشته شوند سپس میانگین هندسی داده‌ها عبارت است از:

بازده سهام

$$\mu_G = \sqrt[N]{\Pi HPR} - 1 = \sqrt[N]{(1+R_1)(1+R_2)\ldots(1+R_N)} - 1$$

میانگین

اگر داده‌های آماری به صورت اندازه‌های غیرنسبی (دارای واحد مانند متر و ریال و ...) مطرح گردند برای محاسبه متوسط نرخ رشد یا متوسط نرخ تورم به صورت زیر عمل می‌کنیم:

$$\bar{x}_G = \sqrt[N-1]{\frac{x_2}{x_1} \cdot \frac{x_3}{x_2} \cdot \frac{x_4}{x_3} \cdot \frac{x_N}{x_{N-1}}} = \sqrt[N-1]{\frac{x_N}{x_1}}$$

متوسط نرخ رشد $= \%\,(\bar{x}_G - 1) \times 100$

اگر بزرگتر از ۱ بود منهای یک می‌کنیم.

میانگین

- **میانگین هارمونیک**

از این نوع میانگین، برای محاسبه میانگین مشاهداتی استفاده می‌شود که از مقیاس‌های ترکیبی همانند «کیلو در ساعت» یا «دور در ثانیه» برخوردار هستند.

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

در صورت تکرار داده‌ها (وزن داشتن آن‌ها) از فرمول زیر استفاده می‌شود:

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

نکات میانگین

در هر جامعه آماری فقط یک میانگین وجود دارد.

مقادیر بزرگ و کوچک به سهم خود در میانگین سهم دارند.

میانگین تنها پارامتری است که اگر به جای کلیه داده‌ها قرار گیرد، مجموع آن ها تغییری نمی‌کند.

مجموع انحرافات از میانگین همیشه صفر است.

$$\sum (x_i - \mu) = 0$$

مجموع مجذور انحرافات از میانگین همیشه می‌نیمم است

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

MIN

$$\mu(\pm a) = \pm a \qquad \mu(x \pm a) = \mu_x \pm a \qquad \mu_{(\pm ax)} = \pm a \mu_{(x)}$$

مد (نما)

به مقداری گفته می‌شود که در میان سایر مقادیر توزیع، بیشترین تکرار را داشته باشد، مد را با Mo نشان می‌دهند.

میانه

برای مجموعه‌ای از داده‌های آماری شاخص مرکزی (عددی) که ۵۰٪ از داده‌ها بالاتر از آن و ۵۰٪ دیگر پایین‌تر از آن قرار دارند میانه گفته می‌شود.

۵۰٪ داده‌های آماری کم‌تر یا مساوی میانه هستند یعنی: $Med \leq$ ۵۰٪ مشاهدات

۵۰٪ داده‌های آماری بیش‌تر از میانه هستند یعنی: $Med >$ ۵۰٪ مشاهدات

میانه

در صورتی که داده‌های آماری بدون جدول فراوانی باشند برای به دست آوردن میانه ابتدا باید داده‌ها را به صورت صعودی یا نزولی مرتب نماییم.

$$Med = \begin{cases} x_{\frac{n+1}{2}} & \text{تعداد داده‌ها فرد} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{تعداد داده‌ها زوج} \end{cases}$$

میانه برای داده‌های طبقه‌بندی شده

۱- از روی جدول، فراوانی تجمعی را حساب می‌کنیم.

۲- $\frac{N}{۲}$ را بدست آورده سپس در جدول اولین طبقه‌ای که مقدار فراوانی تجمعی آن بیشتر یا مساوی $\frac{N}{۲}$ است را به عنوان طبقه میانه‌دار انتخاب می‌کنیم.

$$\text{میانه}: Md = \text{حد پایین طبقه میانه دار} + \left(\frac{\frac{N}{۲} - \text{فراوانی تجمعی طبقه ماقبل میانه دار}}{\text{فراوانی مطلق طبقه میانه دار}}\right) \times \text{طول هر طبقه}$$

نکات میانه

حاصل جمع **قدرمطلق** تفاضل‌های مقادیر از **میانه**، از جمع قدرمطلق تفاضل‌های مقادیر از هر عدد دیگری **کوچکتر** است.

$$\sum |x_i - M_d| = Minimum$$

در هر جامعه آماری فقط یک میانه وجود دارد.

برخلاف میانگین، میانه از اعداد بسیار بزرگ یا بسیار کوچک متاثر نمی‌شود.
(در هنگام وجود داده‌های پرت پارامتر مرکزی مناسب‌تری است)

از لحاظ هندسی میانه طول خط عمودی است بر محور Xها در نمودار بافت نگار(هیستوگرام) که آن را به دو قسمت مساوی تقسیم می‌کند.

مد (نما)

به مقداری گفته می‌شود که در میان سایر مقادیر توزیع، بیشترین تکرار را داشته باشد، مد را با Mo نشان می‌دهند.

چارک

اگر جامعه آماری به چهار قسمت مساوی تقسیم شود، به هر یک از قسمت‌ها یک چارک گفته می‌شود و آن‌ها را با Q نشان می‌دهند.

انواع چارک‌ها

Q_1: مقداری که ۲۵٪ مشاهدات، پایین‌تر از آن است.

Q_2: مقداری که ۵۰٪ مشاهدات، پایین‌تر از آن است. (**میانه**)

Q_3: مقداری که ۷۵٪ مشاهدات، پایین‌تر از آن است.

نحوه محاسبه چارک‌ها

۱- مرتب نمودن صعودی داده‌ها

۳- پیدا نمودن محل چارک موردنظر

۴- تعیین نمودن مقدار چارک موردنظر به کمک محل چارک

چگونگی محاسبه محل چندک‌ها با یک فرمول

$$L_y = (n+1)\frac{y}{100}$$

تنها نکته این فرمول این است که باید هم چندکی را با یک تناسب ساده به صدک تبدیل کنیم.

الف) اگر حاصل عبارت فوق عدد صحیح r بود، قرار می‌دهیم $Q_p = x_r$

ب) اگر برابر عدد صحیح r نبود با فرمول ترکیب خطی زیر آن را محاسبه می‌کنیم.

$$Q_p = x_r + w(x_{r+1} - x_r)$$

آزمون(۲)

بخش سوم: شاخص‌های پراکندگی

آشنایی با پارامترهای پراکندگی در جوامع کوچک ($N \leq 20$)

پارامترهای پراکندگی

شاخص‌هایی هستند که متوسط میزان دوری و نزدیکی داده‌های توزیع را نسبت به میانگین نشان می‌دهند.

در مورد دارایی‌های مالی، شاخص‌های گرایش مرکزی اطلاعاتی در مورد بازده دارایی و شاخص‌های پراکندگی اطلاعاتی در مورد ریسک دارایی به ما ارائه می‌دهند.

۱- دامنه تغییرات
۲- دامنه میان چارکی
۳- انحراف متوسط از میانگین
۴- واریانس
۵- انحراف معیار
۶- نیمه واریانس
۷- ضریب تغییرات

دامنه تغییرات (R)

ساده‌ترین شاخص پراکندگی است و با کم‌کردن کوچکترین مشاهده از بزرگترین آن‌ها در یک سری توزیع بدست می‌آید.

فرمول:
$$R = MAX_{X_i} - MIN_{X_i}$$

دامنه میان چارکی (IQR)

این شاخص ، پراکندگی داده ها را در فاصله چارک اول و چارک سوم نشان می‌دهد و کاری به مقادیر کوچکتر از Q_1 و بزرگتر Q_3 ندارد.

فرمول دامنه میان چارکی

برای محاسبه این شاخص، کافیست که مقادیر Q_1 و Q_3 را بدست آورده و از هم کم کنیم.

$$IQR = Q_3 - Q_1$$

نیمه میان چارکی

برای بدست آوردن این شاخص، که به انحراف چارکی نیز معروف است، کافیست، مقدار دامنه میان چارکی را بر عدد ۲ تقسیم نماییم.

$$SIQR = \frac{Q_3 - Q_1}{2}$$

شاخص‌های مناسب برای توزیع‌های نامتقارن

۱- استفاده از میانه بعنوان بهترین شاخص مرکزی

۲- استفاده از انحراف چارکی بعنوان بهترین شاخص پراکندگی

انحراف متوسط از میانگین

این شاخص از تقسیم مجموع قدر مطلق انحرافات تک تک مشاهدات از میانگین‌شان بر تعداد مشاهدات بدست.

$$A \cdot D_{\mu} = \frac{\sum \left| X_i - \mu_X \right|}{N}$$

محاسن و معایب $A{\cdot}D_{\mu}$

محاسن: در نظر گرفتن تغییرات کل داده‌ها

معایب: ۱- نشان ندادن تأثیر انحرافات بزرگ

۲- بی‌بهره بودن از بعضی از خواص مطلوب میانگین حسابی

واریانس جامعه و نمونه

در این شاخص پراکندگی، بر خلاف شاخص انحراف متوسط از میانگین بجای قدر مطلق از مجذور (توان ۲) انحرافات استفاده می‌شود.

فرمول:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{N}$$

ویژگی‌های واریانس

در یک بیان ساده واریانس بیشتر، نشان‌دهنده پراکندگی بیشتر داده‌ها نسبت به میانگین است. در مورد دارایی‌های مالی، این مساله (واریانس بیشتر) به معنی **ریسک** بیشتر است.

- واریانس عدد ثابت C برابر **صفر** است.
- اگر مقدار ثابت a را به مشاهدات اضافه یا از آنها کم کنیم، واریانس تغییر نمی‌کند.
- اگر مشاهدات در مقدار ثابت k ضرب شوند، واریانس جدید برابر k^2 می‌شود.

انحراف معیار

این شاخص به منظور برطرف کردن عیوب شاخص‌های قبلی است یعنی نشان ندادن تأثیر انحراف بزرگ توسط $A{\cdot}D_\mu$ و افزایش‌دادن (بیش‌نمایی) تأثیر این انحراف توسط δ_X^2

$$\delta_X = \sqrt{\delta_X^2}$$

نیمه واریانس

یعنی متوسط مجذور مقادیر **نامطلوب**:

$$S.V=\frac{\sum_{i=1}^{k}(X_i-\mu_x)^2}{K}$$

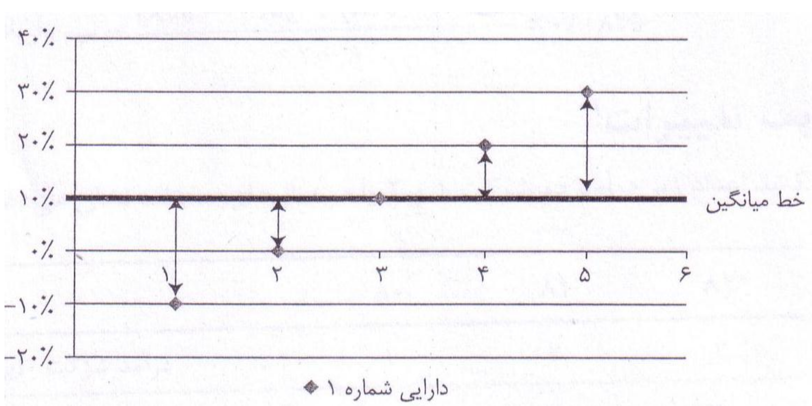
$\mathcal{N}$= تعداد مشاهدات جامعه

$\mathcal{K}$= تعداد مقادیر نامطلوب

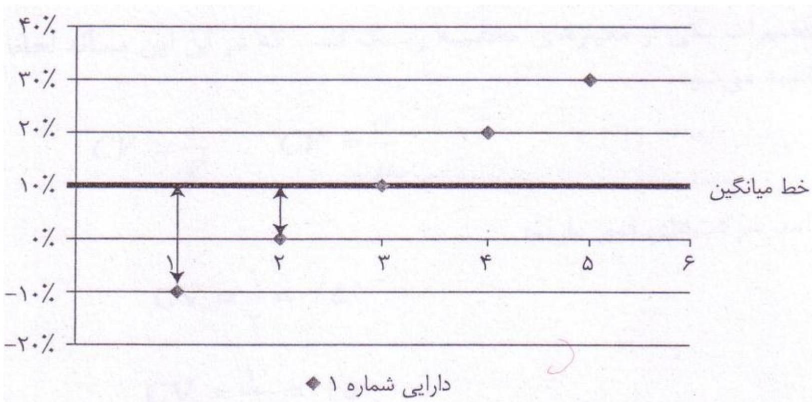
μ_x= میانگین کل مشاهدات جامعه

در داده‌های مربوط به سود و درآمد مقادیر کوچک‌تر از میانگین و در داده‌های مربوط به زیان و هزینه مقادیر بزرگ‌تر از میانگین، **نامطلوب** قلمداد می‌شوند.

نیمه واریانس



نیمه واریانس



ضریب پراکندگی یا ضریب تغییرات

ضریب پراکندگی یکی از معیارهای پراکندگی نسبی است که با فرمول زیر بیان می‌شود:

$$C \cdot V = \frac{\delta_X}{\mu_X}$$

δ_x = انحراف معیار مشاهدات (ریسک)

μ_x = میانگین مشاهدات (بازده)

برای مقایسه دو جامعه در مواردی که:

۱- مقیاس‌ها یکسان نیستند.

۲- مقیاس یکسان ولی تفاوت زیادی در بزرگی مشاهدات وجود دارد.

۳- واریانس‌های جوامع یکسان ولی میانگین‌هایشان متفاوت است.

نسبت شارپ

ضریب پراکندگی یکی از معیارهای پراکندگی نسبی است که با فرمول زیر بیان می‌شود:

$$= \frac{\bar{r}_p - r_f}{\sigma_p}$$

Where:

$\bar{r}_p$ = Expected portfolio return

r_f = Risk free rate

σ_p = Portfolio standard deviation

نسبت شارپ=ریسک سهم /(بازده بدون ریسک – بازده انتظاری سهم)

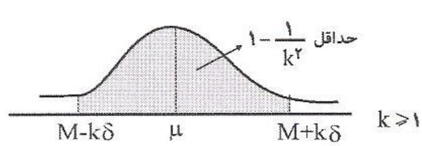
هر مقدار نسبت شارپ یک سرمایه گذاری بیشتر باشد آن سرمایه گذاری از مطلوبیت بالاتری برخوردار است.

منفی بودن نسبت شارپ نشان می‌دهد که بازده انتظاری سهم مورد نظر از بازده بدون ریسک اوراق مشارکت کمتر بوده و لذا سرمایه گذاری در چنین شرکتی توجه‌پذیر به نظر نمی‌رسد.

قاعده چبیشف

به ازای هر ضریب $K > ۱$، فاصله $\bar{X} - KS$ تا $\bar{X} + KS$ شامل **حداقل** $(۱ - \frac{۱}{K^۲})$ از داده‌ها است.

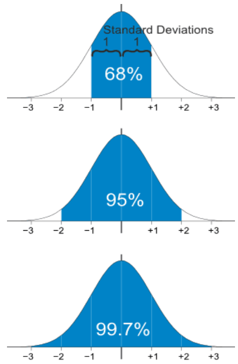
فرض کنید k=۲. در این حالت حداقل $۱ - \frac{۱}{۲^۲} = \frac{۳}{۴}$ از داده‌ها، یعنی حداقل ۷۵٪ از داده‌ها در بازه ی $\bar{X} - ۲S$ تا $\bar{X} + ۲S$ قرار دارند.



قضیه فوق برای همه مجموعه‌های آماری خواه نرمال یا غیرنرمال به کار می‌رود و نقطه ضعف آن کلمه حداقل است که مشخص کننده درصد خاصی نیست.

نکته تکمیلی

وقتی که توزیع مقادیر نمونه، نرمال یا تقریبا نرمال (دارای چولگی ضعیف) باشد بین میانگین و انحراف معیار روابط زیر برقرار است:



- *تقریبا ۶۸٪ مقادیر جامعه در فاصله یک انحراف معیار در طرفین میانگین قرار دارد.*
- *تقریبا ۹۵٪ مقادیر جامعه در فاصله دو انحراف معیار از طرفین میانگین قرار دارد.*
- *تقریبا ۹۹٪ مقادیر جامعه در فاصله انحراف معیار در طرفین میانگین قرار دارد.*

فرمول‌های دیگر واریانس

جامعه: $\delta^2 = \frac{\sum x_i^2}{N} - \mu^2$

نمونه: $s^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)}$

عملیات جبری میانگین و واریانس

اگر جامعه آماری از ترکیب چند جامعه مستقل با میانگین‌ها و واریانس‌های مشخص تشکیل شده باشد، می‌توان میانگین و واریانس جامعه کل را بدست آورد

- **فرمول میانگین حسابی جامعه کل**

N = تعداد مشاهدات هر جامعه

μ_i = میانگین هر جامعه

$$\mu = \frac{N_1\mu_1 + N_2\mu_2 + \dots + N_K\mu_K}{N_1 + N_2 + \dots + N_K} = \frac{\sum N_i\mu_i}{N}$$

- **فرمول واریانس جامعه‌کل**

μ = میانگین کل جامعه

$$\delta^2 = \frac{\sum N_i\delta_i^2}{N} + \frac{\sum N_i(\mu_i - \mu)^2}{N}$$

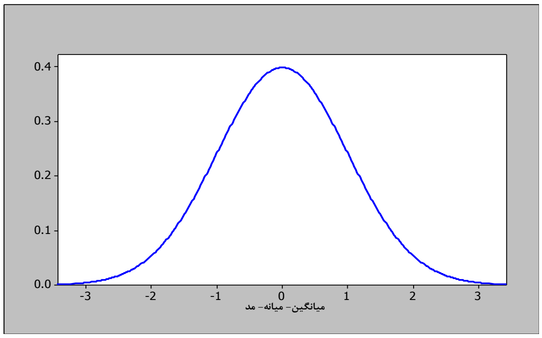
پارامترهای تعیین انحراف از قرینگی

در هنگام مقایسه دو یا چند جامعه، در صورت مساوی بودن پارامترهای مرکزی و پراکندگی، این پارامترها با بهره‌گیری از ضریب چولگی کارساز خواهند بود.

انواع حالات توزیع‌ها

۱- متقارن (نرمال):	مد = میانه = میانگین
۲- چوله به راست:	مد < میانه < میانگین
۳- چوله به چپ:	مد > میانه > میانگین

قله نمودار همیشه نشان دهنده مد است. **میانه نیز همیشه بین مد و میانگین است.**



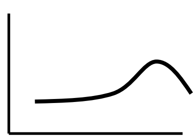
در این حالت (چولگی برابر صفر)، میانگین، میانه و مد برابر با یکدیگر خواهند شد.

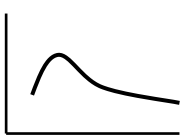
مقادیر مختلف ضریب چولگی (SK)

۱- صفر: در صورت متقارن بودن توزیع جامعه
۲- مثبت: در صورت چوله به راست بودن توزیع جامعه
۳- منفی: در صورت چوله به چپ بودن توزیع جامعه

مفهوم چولگی

اگر دم توزیع جامعه به سمت راست باشد، توزیع را چوله به راست و در صورت عکس، آن را چوله به چپ می‌نامند.


چوله به چپ


چوله به راست

تفسیر مقادیر SK

۱- $|SK| \leq 0/1$ جامعه تقریباً نرمال

۲- $0/1 < |SK| \leq 0/5$ تفاوت اندک با توزیع نرمال

۳- $|SK| > 0/5$ تفاوت فاحش با توزیع نرمال

فرمول‌های محاسبه ضریب چولگی (SK)

۱- ضریب چولگی گشتاوری

۲- ضریب های چولگی پیرسون

۳- ضریب های چولگی چندکی

$$SK = \frac{r_3}{\delta_x^3}$$

- **ضریب چولگی گشتاوری**

$r_3 = \frac{\sum F_i (X_i - \mu_x)^3}{N}$ گشتاور مرتبه سوم به مبدأ میانگین

δ_x انحراف معیار

- **ضریب چولگی پیرسون**

فرمول شماره ۱

$$S \cdot K_1 = \frac{(\mu_X - Mo)}{\delta_x}$$

فرمول شماره ۲

$$S \cdot K_2 = \frac{3(\mu_X - Md)}{\delta_x}$$

رابطه تجربی و معروف پیرسون

$$\mu - Mod = 3(\mu - Med)$$

پارامترهای تعیین انحراف از کشیدگی

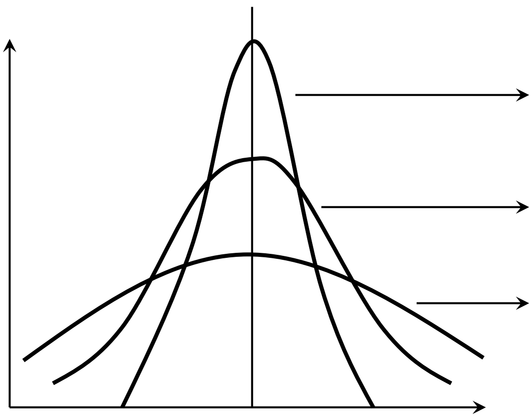
این پارامترها برای مقایسه توزیع جوامع موردنظر با توزیع جامعه نرمال به لحاظ کشیدگی (کوتاهی و بلندی توزیع) مورد استفاده قرار می‌گیرد

انواع توزیع به لحاظ کشیدگی و مقدار ضریب آن(E)

۱- مساوی توزیع نرمال ($E=0$)
۲- بلندتر از توزیع نرمال ($E>0$)
۳- کوتاه تر از توزیع نرمال ($E<0$)

Kurtosis

مقایسه انواع کشیدگی



توزیع کشیده تر $\mathcal{E}>0$

توزیع نرمال $\mathcal{E}=0$

توزیع پراکنده تر $\mathcal{E}<0$

$$\mu_x = Md = Mo$$

تفسیر ضریب کشیدگی (E)

۱- توزیع نرمال $|E| \le 0/1$

۲- توزیع نسبتاً بلندتر از نرمال $0/1 < |E| \le 0/5$

۳- توزیع کاملاً کشیده تر از نرمال $|E| > 0/5$

انواع ضرایب کشیدگی

۱- ضریب کشیدگی گشتاوری؛ با استفاده از گشتاور مرتبه چهارم به مبدأ میانگین

۲- ضریب کشیدگی چندکی؛ با استفاده از انحراف چارکی و صدک‌های دهم و نودم

فرمول ضریب کشیدگی گشتاوری

$$E = \frac{r_4}{\delta_x^4} - 3 \longrightarrow r_4 = \frac{\sum F_i (X_i - \mu_x)^4}{N}$$

کشیدگی توزیع نرمال= 3

آزمون(۳)

بخش چهارم: ارزش زمانی پول

ارزش زمانی پول

- مثال معروفی که بیانگر ارزش زمانی پول می باشد خرید جزیره منهتن از سرخپوستان آمریکایی به مبلغ ۲۴ دلار در سال ۱۶۲۶ است. فرض کنید سرخپوستان می توانستند مبلغ ۲۴ دلار را با نرخ ۶٪ در سال پس انداز نمایند. جدول زیر ارزش ۲۴ دلار سرمایه اولیه را در زمان های متفاوت نشان می دهد:

سال	ارزش ۲۴ دلار سرمایه اولیه
۱۶۲۶	۲۴
۱۶۷۶	۴۴۲.۰۸
۱۷۲۶	۸۱۴۳.۲۴
۱۷۷۶	۱۴۹۹۹۹.۹۲
۱۸۲۶	۲۷۶۳۰۲۱.۶۹
۱۸۷۶	۵۰۸۹۵۲۸۵
۱۹۸۶	۱۷۲۶۸۸۷۶۴۸۴.۳۸

بهره

"بهره عبارت است از قیمتی که می‌بایست در ازای استفاده از پول در یک دوره پرداخت کرد"

$$\text{نرخ بهره برحسب درصد} = \frac{\text{مقدار سرمایه اولیه} - \text{مقدار اصل و فرع}}{\text{مقدار سرمایه اولیه}} \times 100$$

مقدار سرمایه اولیه – مقدار اصل و فرع = مقدار بهره

به عنوان مثال، ۱۰۰ واحد پول امروز با نرخ بهره ۶٪ در سال برابر است با ۱۰۶ واحد پولی سال آینده در همین روز.

- سرمایه‌گذاران باید از بالا بودن نرخ بهره نسبت به نرخ تورم اطمینان حاصل کنند. اگر یک فرد پول خود را با نرخ ۴ درصد سرمایه‌گذاری کند و نرخ تورم نیز ۴ درصد باشد، ثروت فرد افزایش نخواهد یافت.
- نرخ بهره واقعی، نرخی است که یک سرمایه‌گذار انتظار دارد بعد از کسر تورم دریافت کند. نرخ بهره واقعی به دلیل متفاوت بودن انتظارات سرمایه‌گذاران از تورم آینده، عدد واحدی نیست.

نرخ تورم – نرخ بهره اسمی = نرخ بهره واقعی

رابطه دقیق و چند نکته

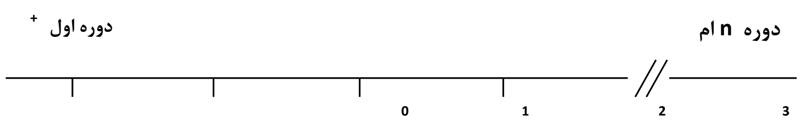
- مرکب کردن طی دوره زمانی می‌تواند موجب اختلاف زیاد نرخ بهره اعلام شده و موثر شود.
- رابطه دقیق بین نرخ بهره اسمی، تورم و نرخ بهره واقعی:

$$\text{نرخ بهره واقعی} = \frac{1+\text{نرخ بهره اسمی}}{1+\text{نرخ تورم}} - 1$$

نرخ بازده مورد انتظار اسمی را زمانی که نرخ هزینه سرمایه ۹٪ و نرخ تورم ۶٪ باشد، تعیین کنید؟

شکل فرآیند مالی

شکل کلی یک فرآیند مالی به صورت زیر است:



در پایان هر دوره درآمدها یا هزینه ها به صورت خط مستقیم عمودی ظاهر می شوند، با این شرط که درآمدها در قسمت **بالا** و هزینه ها در قسمت **پایین** قرار خواهند گرفت.

تمرین فرآیند مالی

شرکت X دستگاه پرسی را ۱۰ سال پیش ۳۵۰۰۰ واحد پولی خرید. درآمد سالیانه دستگاه ۱۲۰۰۰ واحد پولی بوده است. هزینه نگهداری و تعمیرات در سال اول ۱۵۰۰ واحد پولی، در سال دوم ۱۷۵۰ واحد پولی و در سال سوم ۲۰۰۰ واحد پولی بوده و به همین ترتیب هر سال ۲۵۰ واحد پولی افزایش داشته است. شرکت قصد دارد دستگاه پرس را به مبلغ ۲۵۰۰ واحد پولی (ارزش اسقاطی) در سال آینده بفروشد. شکل فرآیند مالی را رسم کنید.

نرخ بهره ساده و مرکب

- تفاوت بهره ساده با بهره مرکب: هنگامی که پولی با نرخ بهره مرکب سرمایه‌گذاری می‌شود **به بهره پول هم بهره تعلق** می‌گیرد. اما در بهره ساده فقط روی اصل پول، بهره حساب می‌شود

- فرض کنید امروز مبلغ ۱.۰۰۰ دلار سرمایه‌گذاری می‌کنید. اگر به این سرمایه‌گذاری بهره‌ای با نرخ ۱۰٪ در سال تعلق گیرد، ارزش آتی (**ارزش مرکب**) دو سال بعد این مبلغ برابر خواهد بود با:

$$FV = PV(1+i)^n = 1.000\ (1+\%10)^2 = 1.210$$

نرخ بهره موثر، یا نرخ تنزیل، نرخ هزینه سرمایه، نرخ بازده مورد انتظار: i

مقایسه

نحوه محاسبه بهره ساده و بهره مرکب را نشان می‌دهد.

بهره مرکب	
P_1	۱,۰۰۰ + (۱,۰۰۰ × ٪۱۰) = ۱,۱۰۰
P_2	۱,۱۰۰ + (۱,۱۰۰ × ٪۱۰) =۱,۲۱۰

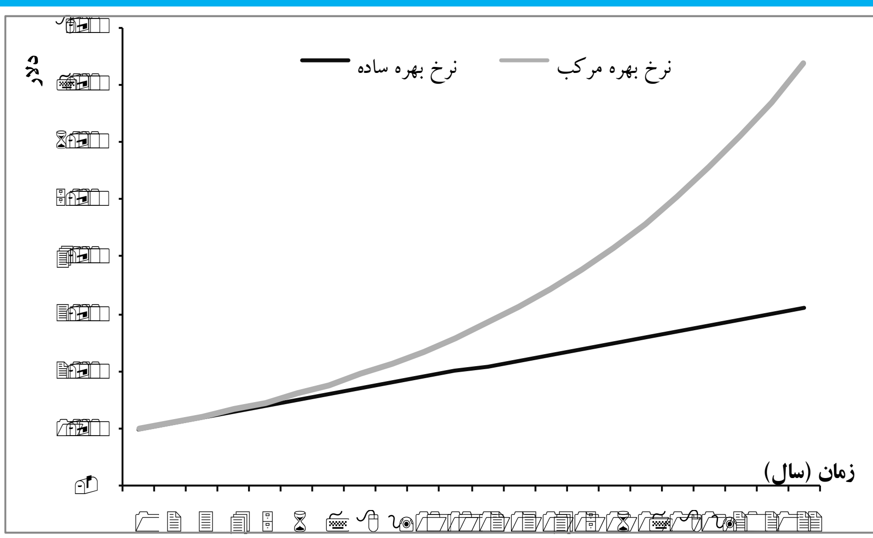
بهره ساده	
P_1	۱,۰۰۰ + (۱,۰۰۰ × ٪۱۰) = ۱,۱۰۰
P_2	۱,۱۰۰ + (۱,۰۰۰ × ٪۱۰) = ۱,۲۰۰

اصل و فرع سرمایه‌گذاری در پایان سال دوم با بهره ساده برابر ۱.۲۰۰ دلار و با بهره مرکب برابر ۱.۲۱۰ دلار است. بنابراین، بهره ساده ارزشی کمتر از بهره مرکب ایجاد می‌کند.

در طول زمان، ارزش ایجاد شده از بهره مرکب و بهره ساده از هم فاصله می‌گیرند.

مقایسه

۹۴



مثال

اگر نرخ بهره ماهانه در سه ماه اول۲٪ و سه ماه دوم ۳٪ باشد، مبلغ ۳۶۰۰۰ ریال امروز بعد از ۶ ماه چند ریال است؟

۱) ۳۵۵۶۲ ۲) ۳۸۰۰۰ ۳) ۴۱۷۴۶ ۴) ۵۱۵۱۵

ارزش آتی با .... و مدت ...... دارای رابطه .... است:

۱- نرخ بهره – پس انداز – غیر مستقیم
۲- نرخ بهره – پس انداز –مستقیم
۳- تورم – پس انداز – غیرمستقیم
۴-بهره بانکی – پس انداز - مستقیم

جداول

حل مساله‌هایی که با نرخ بهرهٔ مرکب محاسبه می‌شوند، چندان مشکل نیست. می‌توان به سادگی $(1+i)^n$ را حل کرد و سپس این مقدار را در سرمایه‌گذاری اولیه(PV) ضرب کرد. مقدار $(1+i)^n$ را عامل ارزش مرکب می‌نامند.

مقدار عامل بهره ارزش آتی (مرکب) از نتایج محاسبه شده در جداول مربوطه استخراج می‌شود. این جداول برای محاسبه آن دسته از تصمیمات مالی در خصوص سرمایه‌گذاری که بلندمدت هستند، بسیار مفید خواهد بود؛ زیرا صرفه‌جویی در وقت و افزایش دقت را به همراه می‌آورد و امکان اشتباه محاسباتی را کاهش می‌دهد.

جداول

	۱٪	۲٪	۳٪	۴٪	۵٪	۶٪	۷٪	۸٪	۹٪	۱۰٪
۱	۱.۰۱۰	۱.۰۲۰	۱.۰۳۰	۱.۰۴۰	۱.۰۵۰	۱.۰۶۰	۱.۰۷۰	۱.۰۸۰	۱.۰۹۰	۱.۱۰۰
۲	۱.۰۲۰	۱.۰۴۰	۱.۰۶۱	۱.۰۸۲	۱.۱۰۳	۱.۱۲۴	۱.۱۴۵	۱.۱۶۶	۱.۱۸۸	۱.۲۱۰
۳	۱.۰۳۰	۱.۰۶۱	۱.۰۹۳	۱.۱۲۵	۱.۱۵۸	۱.۱۹۱	۱.۲۲۵	۱.۲۶۰	۱.۲۹۵	۱.۳۳۱
۴	۱.۰۴۱	۱.۰۸۲	۱.۱۲۶	۱.۱۷۰	۱.۲۱۶	۱.۲۶۲	۱.۳۱۱	۱.۳۶۰	۱.۴۱۲	۱.۴۶۴
۵	۱.۰۵۱	۱.۱۰۴	۱.۱۵۹	۱.۲۱۷	۱.۲۷۶	۱.۳۳۸	۱.۴۰۳	۱.۴۶۹	۱.۵۳۹	۱.۶۱۱
۶	۱.۰۶۲	۱.۱۲۶	۱.۱۹۴	۱.۲۶۵	۱.۳۴۰	۱.۴۱۹	۱.۵۰۱	۱.۵۸۷	۱.۶۷۷	۱.۷۷۲
۷	۱.۰۷۲	۱.۱۴۹	۱.۲۳۰	۱.۳۱۶	۱.۴۰۷	۱.۵۰۴	۱.۶۰۶	۱.۷۱۴	۱.۸۲۸	۱.۹۴۹
۸	۱.۰۸۳	۱.۱۷۲	۱.۲۶۷	۱.۳۶۹	۱.۴۷۷	۱.۵۹۴	۱.۷۱۸	۱.۸۵۱	۱.۹۹۳	۲.۱۴۴
۹	۱.۰۹۴	۱.۱۹۵	۱.۳۰۵	۱.۴۲۳	۱.۵۵۱	۱.۶۸۹	۱.۸۳۸	۱.۹۹۹	۲.۱۷۲	۲.۳۵۸
۱۰	۱.۱۰۵	۱.۲۱۹	۱.۳۴۴	۱.۴۸۰	۱.۶۲۹	۱.۷۹۱	۱.۹۶۷	۲.۱۵۹	۲.۳۶۷	۲.۵۹۴

اگر بانکی و یا صندوقی تبلیغ کند که برای حساب‌های پس‌انداز، بهره i درصد پرداخت می‌کند و سود را هر چند ماه یک بار می‌پردازد؛ در چنین حالتی برای محاسبه ارزش آتی می‌توان از رابطه زیر استفاده نمود:

$$FV = PV(1 + \frac{i}{m})^{mn}$$

FV = ارزش آتی (مرکب)
PV = ارزش فعلی (حال)
i = نرخ بهره سالیانه
m = تعداد دفعات پرداخت بهره در سال
n = تعداد سال‌ها

محاسبه بهره مرکب (چند بار در یک سال)

اگر نرخ بهره اعلام شده بانکی ۱۲٪ و به صورت فصلی و مرکب باشد، بدان معنی است که ۳٪ بهره هر سه ماه یک بار می‌پردازد. از این رو، اگر کسی ۱.۰۰۰ دلار در آغاز سال سرمایه‌گذاری کند در پایان سال یا پس از گذشت چهار فصل، حساب پس انداز او چقدر خواهد شد.

محاسبه بهره مرکب (چند بار در یک سال)

اگر نرخ بهره اعلام شده بانکی ۱۲٪ و به صورت فصلی و مرکب باشد، بدان معنی است که ۳٪ بهره هر سه ماه یک بار می‌پردازد. از این رو، اگر کسی ۱.۰۰۰ دلار در آغاز سال سرمایه‌گذاری کند در پایان سال یا پس از گذشت چهار فصل، حساب پس انداز او چقدر خواهد شد.

$$FV = PV(۱+\frac{i}{m})^{mn} = ۱.۰۰۰(۱+\frac{12\%}{4})^{۴\times۱} = ۱/۱۲۵۵ \times ۱.۰۰۰ = ۱۱۲۵/۵$$

نکته مهم با توجه به مثال قبل

اگر نرخ بهره اعلام شده بانکی ۱۲٪ و به صورت فصلی و مرکب باشد، بدان معنی است که ۳٪ بهره هر سه ماه یک بار می‌پردازد. از این رو، اگر کسی ۱.۰۰۰ دلار در آغاز سال سرمایه‌گذاری کند در پایان سال یا پس از گذشت چهار فصل، حساب پس انداز او چقدر خواهد شد.

$$FV = PV(۱ + \frac{i}{m})^{mn} = ۱.۰۰۰(۱ + \frac{12\%}{4})^{۴\times۱} = ۱/۱۲۵۵ \times ۱.۰۰۰ = ۱۱۲۵/۵$$

بنابراین۱۲٪ نرخ بهره مرکب که به صورت سه ماهه پرداخت می‌شود، معادل ۱۲/۵۵٪ نرخ بهره موثر سالانه است. برای محاسبه بهره موثر سالانه (r_e) که در سال، m مرتبه بهره به آن تعلق می‌گیرد و به صورت مرکب پرداخت می‌شود را می‌توان از معادله زیر به دست آورد:

$$R_e = (۱ + \frac{i}{m})^m - ۱$$

مثال

یک بانک صنعتی اعلام کرده است که نرخ بهره این بانک برای تاسیس واحدهای صنعتی ۱٪ در ماه (بحث کنید) است. نرخ موثر سالیانه را حساب کنید.

$$R_e = (۱ + \frac{12\%}{12})^{۱۲} - ۱$$

$R_e =$ نرخ موثر سالیانه 12/7٪

زمانی که شرکت از بانک وام می‌گیرد، برای محاسبه نرخ موثر وام می‌توان از رابطه زیر استفاده نمود:

$$\text{نرخ موثر وام} = \frac{\text{خالص دریافتی} - \text{خالص پرداختی}}{\text{خالص دریافتی}}$$

مثال

شخصی قصد دارد مبلغ ۱۰۰۰۰۰۰ ریال وام یکساله با کارمزد ۱۰٪ که در ابتدا کسر می‌شود دریافت نماید. اگر اخذ وام مستلزم افتتاح یک حساب سپرده غیرقابل برداشت تا سر رسید به مبلغ ۵۰۰۰۰ ریال باشد، نرخ کارمزد موثر این وام چند درصد است؟

خالص دریافتی: 1,000,000-0/1(1,000,000)-50,000=850,000
خالص پرداختی: 1,000,000-50,000=950,000

نرخ کارمزد موثر این وام = ٪۱۱.۷۶

ارزش آتی جریانات نقدی نامساوی

برای محاسبه ارزش آتی در مواقعی که پرداخت‌ها در طول سال نامنظم باشند، رابطه زیر می‌تواند مورد استفاده قرار گیرد:

$$FV = \sum_{n=1}^{N} A_n\,(1+i)^{N-n}$$

پرداخت (اقساط) در پایان سال n ام می‌باشد. اولین پرداخت بهره برای سال N-۱، دومین پرداخت بهره در سال N-۲ ،... و آخرین پرداخت در پایان سال N صورت می‌گیرد.

(مثال)ارزش آتی جریانات نقدی نامساوی

ارزش آتی سرمایه‌گذاری ۳ ساله که قسط اول آن ۱.۰۰۰ دلار در پایان سال اول، قسط دوم ۱.۲۰۰ دلار در سال دوم و قسط سوم ۱.۵۰۰ دلار در پایان سال سوم است را با نرخ بازده ۵٪ محاسبه کنید.

$$FV = \sum_{n=1}^{N} A_n (1+i)^{N-n}$$

$$FV = 1.000(1+0.05)^{3-1} + 1.200(1+0.05)^{3-2} + 1.500(1+0.05)^{3-3}$$

FV =۱۱۰۲/۵+ ۱.۲۶۰+ ۱.۵۰۰ = ۳۸۶۲/۵

ارزش آتی جریانات نقدی مساوی

قسط عبارت است از یک رشته جریانات نقدی مساوی که در فواصل معین زمانی دریافت یا پرداخت می‌شود. ارزش آتی اقساط مساوی که در پایان هر سال پرداخت می‌شود را می‌توان از معادله زیر به دست آورد:

$$A \left[\frac{(1+i)^n - 1}{i}\right] = FV$$

FV = ارزش آتی اقساط مساوی
A = اقساط
i = نرخ بهره سالانه
n = مدت (تعداد اقساط)

در معادله فوق، فرض بر این است که اقساط در پایان هر سال پرداخت می‌شود و اصل و فرع پول بلا فاصله پس از پرداخت آخرین قسط محاسبه می‌شود و این بدان معنی است که به آخرین قسط، بهره تعلق نمی‌گیرد.

(مثال)ارزش آتی جریانات نقدی مساوی

چه مقدار پول در حساب بانکی شما خواهد بود، اگر شما از سال آینده به مدت ۸ سال، هر سال ۱.۰۰۰ واحد پولی در حساب بانکی خود پس‌انداز نمائید. نرخ بهره برای پس‌انداز در بانک ۴٪ فرض می‌شود.

$$FV = 1.000\left[\frac{(1+\%4)^{8}-1}{\%4}\right] = 9.214$$

چه مقدار پول باید هر سال (شروع یک سال بعد) در بانک پس‌انداز کنید تا پس از ۷ سال با نرخ بهره ۵٪، مبلغ ۱۰۰.۰۰۰ واحد پولی در حساب شما باشد؟

$$100.000 = A\left[\frac{(1+\%5)^{7}-1}{\%5}\right] \Rightarrow A = ۱۲.۲۸۲$$

ارزش فعلی (تنزیل)

ارزش فعلی عبارت است از ارزش جریانات نقدی آتی در زمان صفر که بر اساس ارزش زمانی پول محاسبه شده باشد. چگونگی محاسبه ارزش فعلی جریانات نقدی آتی را «تنزیل» می‌گویند. برای تنزیل یک جریان نقدی و محاسبه ارزش فعلی آن می‌توان بدین صورت عمل کرد:

$$PV = \frac{FV}{(1+i)^{n}}$$

از این رو، ارزش فعلی درست برعکس ارزش مرکب است. اگر پولی به تنزیل گذاشته شود باید جریان نقدی آینده را بر $(1+\mathbf{i})^{n}$ تقسیم کرد.

مثال ساده از ارزش فعلی (تنزیل)

به آقای رضوی پیشنهاد شده است که یک برگ اوراق بهادار بخرد. سند مزبور به او چنین حقی می‌دهد که ۵ سال بعد مبلغ ۱۲۰۰۰ دلار دریافت کند. آقای رضوی پس از بررسی ریسک مربوطه به این نتیجه رسید که نرخ بهره ۱۲٪، نرخ تنزیل مناسبی است. حالا آقای رضوی چه مبلغی باید بپردازد که ۵ سال بعد ۱۲۰۰۰ دلار دریافت کند.

$$\boxed{PV = \frac{FV}{(1+i)^n}}$$

$$PV = \frac{12000}{(1+\%12)^5} = 6809$$

هر چقدر نرخ تنزیل (**i**) بیشتر و زمان سررسید (**n**) طولانی تر باشد، ارزش فعلی کمتر خواهد بود.

مثال

ارزش فعلی جریان‌های نقدی زیر با نرخ ۱۰٪ مبلغ ۵.۵۶۷ ریال است. مبلغ جریان نقدی سال دوم چند ریال می‌باشد.

ریال	۴۰۰۰	؟	۱۰۰۰	۰
سال	سوم	دوم	اول	۰

ارزش فعلی جریانات نقدی میان دوره‌ای (تنزیل چند بار در یک سال)

برای محاسبه ارزش فعلی جریان‌های نقدی که به صورت میان دوره‌ای (چند بار در یک سال) تنزیل می‌گردد، می‌توان از معادله زیر استفاده کرد:

$$PV = \frac{FV}{(1+\frac{i}{m})^{mn}}$$

ارزش فعلی با نرخ بهره و مدت پس‌انداز دارای رابطه معکوس است.

مثال

ارزش فعلی یک سرمایه‌گذاری آتی ۲۰۰۰ دلاری که بعد از ۵ سال به دست می‌آید با نرخ بهره ۱۲٪ که به صورت فصلی پرداخت می‌شود را محاسبه کنید.

$$PV = \frac{FV}{(1+\frac{i}{m})^{mn}} = \frac{2.000}{(1+\frac{\%12}{4})^{4\times5}} = 1107$$

در صورتی که بهره به دفعات (چندین بار در سال) تنزیل شود، ارزش آن کاهش می‌یابد. در مثال فوق اگر پرداخت‌ها را به صورت سالانه تنزیل کنیم، ارزش فعلی آن حدود ۱۱۳۵ دلار به دست می‌آید که بزرگتر از ۱.۱۰۷ دلار است.

ارزش فعلی جریانات نقدی نامساوی

برای محاسبه ارزش فعلی در مواقعی که جریانات نقدی نامنظم (اقساط نامساوی) باشند، رابطه زیر می‌تواند مورد استفاده قرار گیرد:

$$PV = \sum_{n=1}^{N} \frac{A_n}{(1+i)^n}$$

PV = ارزش فعلی (حال)
i = نرخ بهره
A_n = مبلغ جریان نقدی (پرداخت) در پایان سال nام
n = تعداد سال

ارزش فعلی جریانات نقدی نامساوی

یک سازمان سرمایه‌گذاری تعهد می‌کند که در پایان سال اول ۱.۰۰۰ دلار، پایان سال دوم ۲.۵۰۰ دلار و پایان سال سوم ۴.۰۰۰ دلار پرداخت کند. اگر نرخ بهره مرکب ۱۰٪ در سال باشد، چه مبلغ باید در این سازمان سرمایه‌گذاری شود؟

$$PV = \frac{1.000}{(1+\%10)^1} + \frac{2.500}{(1+\%10)^2} + \frac{4.000}{(1+\%10)^3} = 5.980$$

ارزش فعلی جریانات نقدی مساوی

ارزش فعلی اقساط مساوی که در پایان هر سال پرداخت می‌شود را می‌توان از معادله زیر به دست آورد:

$$PV = A\left[\frac{1-(1+i)^{-n}}{i}\right]$$

PV = ارزش حال اقساط مساوی

A = اقساط

i = نرخ بهره سالانه

n = مدت (تعداد اقساط)

ارزش فعلی جریانات نقدی مساوی

اگر بانکی متعهد شود از سال آینده در چنین روزی تا مدت ۹ سال، همه ساله مبلغ ۶.۰۰۰ واحد پولی به شما پرداخت کند و نرخ بانک ۷٪ در سال فرض شود، چه مقدار پول در این طرح سرمایه‌گذاری می‌کنید؟

$$PV = A\left[\frac{1-(1+i)^{-n}}{i}\right]$$

$$PV = A\left[\frac{1-(1+i)^{-n}}{i}\right] = 6.000\left[\frac{1-(1+\%7)^{-9}}{\%7}\right] = 39.091$$

مطلوب است ارزش فعلی ۳ قسط مساوی ۱۵۰ هزار تومانی سالانه با فرض این که نرخ تنزیل برابر با ۲۰٪ باشد؟

$$PV_A = ۱۵۰۰۰۰\left(\frac{۱-\frac{۱}{(۱+۰/۲)^{۳}}}{۰/۲}\right) = ۳۱۵۹۰۰$$

نکات مهم

در صورتی که در مسائل مربوط به ارزش فعلی و آتی اقساط، صریحا به زمان وقوع جریانات نقدی اشاره نشده باشد، **فرض بر این است که جریانات در پایان دوره اتفاق افتاده است** و برای ایجاد ارتباط بین این گونه جریانات نقدی و جریانات نقدی که در ابتدای دوره پرداخته شده خواهیم داشت:

ارزش فعلی(آتی)اقساط ابتدای دوره=ارزش فعلی (آتی)اقساط پایان دوره× (1+ i)

به عبارتی، زمانی که در مسئله صراحتا بیان می‌شود که اقساط در ابتدای دوره پرداخت یا دریافت می‌شود. جهت محاسبه ارزش فعلی یا آتی آن‌ها می‌بایست، پاسخی که از روابط قبل به دست آمده را در عبارت (1+i) ضرب نمود تا زمان محاسبه ارزش فعلی یا آتی درست لحاظ شده باشد.

مثال

در صورتی که در مسائل مربوط به ارزش فعلی و آتی اقساط، صریحا به زمان وقوع جریانات نقدی اشاره نشده باشد، **فرض بر این است که جریانات در پایان دوره اتفاق افتاده است** و برای ایجاد ارتباط بین این گونه جریانات نقدی و جریانات نقدی که در ابتدای دوره پرداخته شده خواهیم داشت:

ارزش فعلی(آتی)اقساط ابتدای دوره=ارزش فعلی (آتی)اقساط پایان دوره× (1+ i)

به عبارتی، زمانی که در مسئله صراحتا بیان می‌شود که اقساط در ابتدای دوره پرداخت یا دریافت می‌شود. جهت محاسبه ارزش فعلی یا آتی آن‌ها می‌بایست، پاسخی که از روابط قبل به دست آمده را در عبارت (1+i) ضرب نمود تا زمان محاسبه ارزش فعلی یا آتی درست لحاظ شده باشد.

مثال

فردی برای باز پرداخت بدهی یک میلیون ریالی خود در ۳ سال دیگر اقدام به افتتاح حساب پس‌انداز نزد بانک نموده است. در صورتی که وی در ابتدای هر سال اقدام به سپرده‌گذاری نماید و ضمنا نرخ سود بانکی سالیانه ۱۰٪ باشد، مبلغ سپرده سالیانه وی چه میزان خواهد بود؟

274,725(1 333,333(2 302,115(3 432,900(4

$$FV_A=A\left(\frac{(1+i)^t-1}{i}\right)(1+i)$$

$$1000000=A\left(\frac{(1+i)^3-1}{0/1}\right)(1+0/1) \qquad A=274725$$

مثال

با توجه به شکل، ارزش فعلی مجهول چند میلیون ریال است؟

سال					ارزش فعلی با نرخ ۲۰٪
۴	۳	۲	۱		
۸	۸	۸	۸		۲۰/۷۱میلیون ریال
۵	۵	۵	۵		۱۲/۹۴میلیون ریال
۳	۳	۳	۳		؟

FV(A±B)=FV(A)±FV(B)
PV(A±B)=PV(A)±PV(B)

اقساط مادام العمر

$$PV = \frac{A}{i}$$

سرمایه‌گذاری در پروژه‌ای درآمد سالیانه دایمی معادل یک میلیون ریال ایجاد می‌نماید. ارزش بازار این سرمایه‌گذاری در حالی که نرخ سود مورد انتظار ۲۰٪ باشد، چقدر است؟

$$PV = \frac{1{,}000{,}000}{0/2} = 5{,}000{,}000$$

اقساط مادام‌العمری که هرسال با نرخ g رشد نمایند.

$$PV = \frac{A}{i - g}$$

اقساط مادام العمر

زمانی که تعداد دفعات پرداخت بهره زیاد شود، در آن صورت بهره به صورت پیوسته محاسبه شده و بهره مرکب در بیشترین مقدار خود خواهد بود که از رابطه زیر محاسبه می‌شود: ($n \to \infty$ نهایت میل کند)

$$FV = Pe^{i \times n}$$

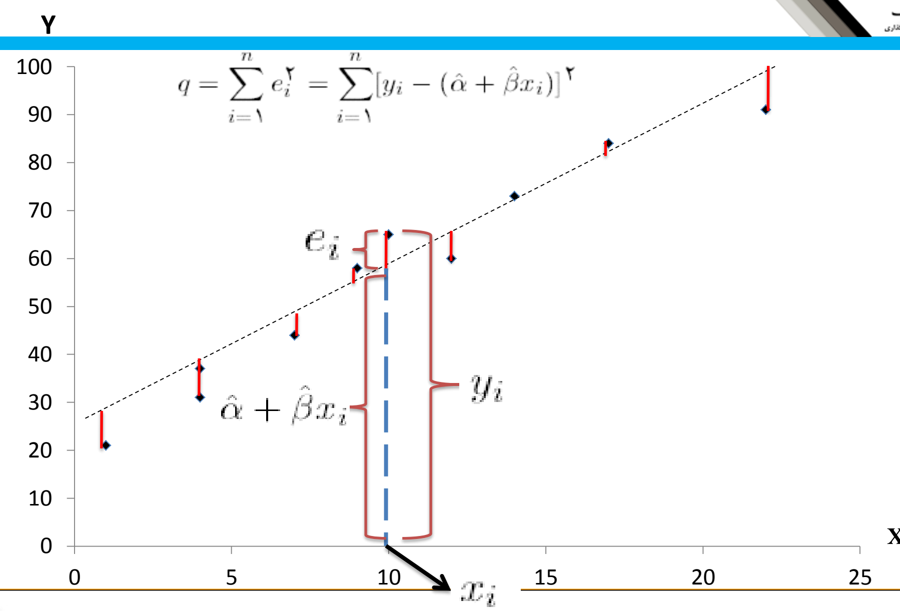
روش کمترین مربعات مجذورات (حداقل مربعات)

این روش رابطه ای ریاضی را بین دو متغیر وابسته و مستقل تعریف می کند. ساده ترین ارتباط یک ارتباط خطی است. هدف، ترسیم بهترین خطی است که مقدار خطا را حداقل کند.

خطای پیش بینی = $e = Y_i - \widehat{Y}_i$

که در آن :

- Y_i = اطلاعات حقیقی سال i
- $\widehat{Y}_i$ = اطلاعات تخمینی سال i



$$e^2 = (Y_i - \hat{Y}_i)^۲$$

اکنون هدف دست یافتن به کمترین مجموع مجذورات خطای تخمینی خواهد بود: = **خطای پیش بینی**

$$\boxed{Min : \sum_{i=۱}^{n} e_i^۲ = \sum_{i=۱}^{n} (Y_i - \hat{Y}_i)^۲}$$

در صورتی که خط تخمینی دارای معادله $\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i$ **باشد.**

شرط لازم برای مینیمم شدن این رابطه آن است که مشتقش نسبت به متغیرهای مستقل مختلف برابر صفر گردد.

$$\sum_{i=۱}^{n} e_i^۲ = \sum_{i=۱}^{n} (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

مشتق نسبت به $\hat{\alpha}$:

$$\frac{\partial \sum e_i^۲}{\partial \hat{\alpha}} = -۲ \left[\sum (Y_i - \hat{\alpha} - \hat{\beta} X_i)\right] = ۰$$

$$\rightarrow \sum (Y_i - \hat{\alpha} - \hat{\beta} X_i) = ۰$$

$$\rightarrow \sum Y_i = \sum \hat{\alpha} + \hat{\beta} \sum X_i \rightsquigarrow$$ معادله اول نرمال

با تقسیم معادله اول نرمال به n خواهیم داشت:

$$\frac{\sum Y_i}{n} = \frac{\sum \hat{\alpha}}{n} + \frac{\hat{\beta} \sum X_i}{n} \rightarrow \bar{Y} = \hat{\alpha} + \hat{\beta} \bar{X}$$

$$\rightarrow \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

مشتق نسبت به $\hat{\beta}$:

$$\frac{\partial \sum e_i^{۲}}{\partial \hat{\beta}} = -۲ \sum X_i (Y_i - \hat{\alpha} - \hat{\beta} X_i) = ۰$$

$$\sum X_i Y_i = \hat{\alpha} \sum X_i + \hat{\beta} \sum X_i^{۲}$$ معادله دوم نرمال

مقدار $\hat{\alpha}$ را از رابطه $\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$ در معادله دوم نرمال قرار می‌دهیم :

$$\sum X_i Y_i - \bar{Y} \sum X_i = \hat{\beta} (\sum X_i^{۲} - \bar{X} \sum X_i)$$

$$\rightarrow \hat{\beta} = \frac{\sum X_i Y_i - \bar{Y} \sum X_i}{\sum X_i^{۲} - \bar{X} \sum X_i}$$

دو فرمول دیگر؛

اگر صورت و مخرج رابطه فوق را در n ضرب کنیم :

$$\hat{\beta} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^{۲} - (\sum X_i)^{۲}}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

اگر صورت و مخرج رابطه فوق را در n تقسیم کنیم :

$$\hat{\beta} = \frac{E(XY) - E(X)E(Y)}{E(X^{۲}) - E(X)^{۲}} = \frac{COV(X, Y)}{\sigma_x^{۲}}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

($\overline{Y}$ و $\overline{X}$) نقطه میانگین بوده که خط رگرسیون لزوما از آن عبور می‌کند.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \,, \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

مثال : فرض کنید در یک مسئله توسعه سیستم حمل و نقل، مقادیر زیر براساس یک نمونه تصادفی n=۷ به دست آمده‌اند، برآورد حداقل مربعات مقادیر a ، b در معادله رگرسیونی $Y = a + bx + \varepsilon$ را به دست آورید؟

$$\sum X_i = ۰ \qquad \sum X_i^{۲} = ۲۸ \qquad \sum Y_i = ۲۱ \qquad \sum X_i Y_i = ۵۶$$

در معادله رگرسیون $Y = a + bx + \varepsilon$ مقادیر b (شیب خط) و a (مقدار ثابت) از روابط زیر به دست می‌آید.

$$b = \frac{COV(x,y)}{\delta_X^{۲}} = \frac{\frac{\sum XY}{n} - \frac{\sum X}{n} \times \frac{\sum Y}{n}}{\frac{\sum X^{۲}}{n} - \left(\frac{\sum X}{n}\right)^{۲}} = \frac{\frac{۵۷}{۷} - \frac{۰}{۷} \times \frac{۲۱}{۷}}{\frac{۲۷}{۸} - \left(\frac{۰}{۷}\right)^{۲}} = ۲$$

با توجه به آن که خط رگرسیون همیشه از نقطه $(\overline{X}, \overline{Y})$ عبور می‌کند، مقدار a از رابطه زیر به دست می‌آید.

$$\begin{cases} \overline{Y} = b\overline{X} + a \implies ۳ = ۲ \times ۰ + a \implies a = ۳ \\ \overline{Y} = \frac{\sum Y}{n} = ۳ \\ \overline{X} = \frac{\sum X}{n} = ۰ \end{cases}$$

$$\hat{\beta} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^{۲} - (\sum X_i)^{۲}}$$

ضریب همبستگی بین دو متغیر وابسته X و Y را محاسبه نمایید:

X_i	۷	۱۰	۴	۱۱
Y_i	۱۴	۲۰	۸	۲۲

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
۷	۱۴	۴۹	۱۹۶	۹۸
۱۰	۲۰	۱۰۰	۴۰۰	۲۰۰
۴	۸	۱۶	۶۴	۳۲
۱۱	۲۲	۱۲۱	۴۸۴	۲۴۲
$\sum X_i = 32$	$\sum Y_i = 64$	$\sum X_i^2 = 286$	$\sum Y_i^2 = 1144$	$\sum X_iY_i = 572$

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

$$= \frac{\frac{۵۷۲}{۴} - \frac{۳۲}{۴} \times \frac{۶۴}{۴}}{\sqrt{\left(\frac{۲۸۶}{۴} - \left(\frac{۳۲}{۴}\right)^۲\right)\left(\frac{۱۱۴۴}{۴} - \left(\frac{۶۴}{۴}\right)^۲\right)}}$$

$$= \frac{۱۴۳ - ۸ \times ۱۶}{\sqrt{(۷۱.۵ - ۶۴)(۲۸۶ - ۲۵۶)}} = \frac{۱۵}{\sqrt{۲۲۵}} = ۱$$

اگر شیب یک معادله رگرسیون خطی ۱۰- باشد و $\sum X = 100$ و $\bar{X} = 20$ و $\sum Y = 20$ باشند، عرض از مبدا معادله کدام است ؟

$$\begin{cases} \bar{X} = \dfrac{\sum X}{n} \Rightarrow ۲۰ = \dfrac{۱۰۰}{n} \Rightarrow n = ۵ & (۱) \\ \bar{Y} = \dfrac{\sum y}{n} = \dfrac{۲۰}{۵} = ۴ & (۲) \\ a = \text{شیب خط} = -۱۰ & (۳) \end{cases}$$

$$\xrightarrow{(۱),(۲),(۳)} \begin{cases} \bar{Y} = a\bar{X} + b \\ ۴ = -۱۰ \times ۲۰ + b \rightarrow b = ۲۰۴ \end{cases}$$

راه ساده‌تر:

فرض می کنیم : $x_i = X_t - \bar{X} \quad , \quad y_i = Y_t - \bar{Y}$

$$SS_x = S_{xx} = \sum x_i^۲ = \sum (X_i - \bar{X})^۲ \rightarrow \sum (X_i^۲ - ۲\bar{X}X_i + \bar{X}^۲)$$

$$\rightarrow \sum X_i^۲ - ۲\bar{X}\sum X_i + \sum \bar{X}^۲ \rightarrow \sum X_i^۲ - ۲\bar{X}(n\bar{X}) + n\bar{X}^۲$$

$$\rightarrow \sum X_i^۲ - n\bar{X}^۲ \rightarrow \sum X_i^۲ - \frac{(\sum X_i)^۲}{n}$$

این عبارت معادل مخرج عبارتی است که قبلا به دست آورده بودیم؛

$$\hat{\beta} = \frac{\sum X_i Y_i - \bar{Y}\sum X_i}{\sum X_i^۲ - \bar{X}\sum X_i}$$

راه ساده‌تر:

$$x_i = X_t - \bar{X} \quad , \quad y_i = Y_t - \bar{Y}$$

$$SS_{xy} = S_{xy} = \sum x_i\ y_i = \sum (X_i - \bar{X}).(Y_i - \bar{Y}) = \sum (X_i Y_i - \bar{X}Y_i - \bar{Y}X_i + \overline{XY})$$

$$\rightarrow \sum X_i Y_i - \bar{X} \sum Y_i - \bar{Y} \sum X_i + n\,\overline{XY} \rightarrow \sum X_i Y_i - n\,\overline{XY} - n\,\overline{XY} + n\,\overline{XY}$$

$$\rightarrow \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}$$

این عبارت معادل صورت عبارتی است که قبلا به دست آورده بودیم؛

$$\hat{\beta} = \frac{\sum X_i\ Y_i - \bar{Y} \sum X_i}{\sum X_i^2 - \bar{X} \sum X_i}$$

$$\hat{\beta} = \frac{SS_{xy}}{SS_x}$$

یا

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X}$$

- گام اول : محاسبه : $\bar{X}, \bar{Y}, x_i = (X_i - \bar{X}),\ y_i = (Y_i - \bar{Y})$
- گام دوم : محاسبه : x_i^2 و $\sum x_i^2$
- گام سوم : محاسبه : $\sum x_i y_i$

ضریب همبستگی

همبستگی یک پارامتر مرتبط با کواریانس است که شدت رابطه(همبستگی) بین متغیرهای مطرح در مدل رگرسیون را از نظر خطی بررسی میکند

- همبستگی مثبت (مستقیم) : افزایش یا کاهش یکی باعث افزایش یا کاهش دیگری می شود.
- همبستگی منفی (غیر مستقیم) : افزایش یکی باعث کاهش دیگری و بلعکس.

$-1 \leq \rho \leq +1 \rightarrow \rho = 0 \rightarrow$ X و Y مستقل بوده و رابطه خطی ندارند.

$-1 < \rho < 0 \rightarrow$ X و Y رابطه غیرمستقیم دارند

$0 < \rho < 1 \rightarrow$ X و Y رابطه مستقیم دارند

ضریب همبستگی

$$\rho = \frac{COV(x,y)}{\sigma_x \sigma_y} = \frac{\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\frac{\sum (X_i - \bar{X})^2}{n} \times \frac{\sum (Y_i - \bar{Y})^2}{n}}} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \cdot \sigma_Y}$$

$$= \frac{\frac{\sum X_i Y_i}{n} - \left(\frac{\sum X_i}{n}\right) \cdot \left(\frac{\sum Y_i}{n}\right)}{\sqrt{\left[\frac{\sum X_i^2}{n} - \left[\frac{\sum X_i}{n}\right]^2\right]\left[\frac{\sum Y_i^2}{n} - \left[\frac{\sum Y_i}{n}\right]^2\right]}} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}}$$