مدرسان شریف

فصل اول

«آمـــار توصیفی»

درسنامه (۱): مفاهیم اولیه

معرفی علم آمار: علم آمار به سه شاخه‌ی آمار توصیفی، آمار استنباطی (پارامتری) و آمار ناپارامتری تقسیم می‌شود.

آمار توصیفی (Descriptive Statistics)

این شاخه از علم آمار با کمک ابزارهایی مانند جداول فراوانی، نمودارهای آماری و محاسبه‌ی شاخص‌های آماری به توصیف داده‌ها می‌پردازد. در آمار توصیفی همه‌ی داده‌ها به طریق سرشماری جمع‌آوری می‌شوند و هیچ‌گونه استنباطی بر روی آن‌ها انجام نمی‌شود و فقط به توصیف آن‌ها پرداخته می‌شود.

آمار استنباطی (پارامتری) (Inferential Statistics)

در آمار استنباطی به کمک نمونه‌گیری به تحلیل و استنتاج در مورد پارامترهای کلی جامعه‌ی آماری پرداخته می‌شود. توجه کنید که در آمار استنباطی توزیع یا نحوه‌ی پخش داده‌های آماری مشخص است.

آمار ناپارامتری (آزاد از توزیع)(Non – Parametric Statistics)

آمار ناپارامتری دقیقاً مانند آمار استنباطی عمل می‌کند، با این تفاوت که در آمار ناپارامتری اغلب صفات کیفی هستند و توزیع یا نحوه‌ی پخش داده‌های آماری مشخص نیست.

مثال ۱: مرکز آمار ایران هر چند سال یک بار سرشماری انجام می‌دهد. اطلاعات جمع‌آوری شده با استفاده از کدام شاخه علم آمار مورد بررسی قرار می‌گیرند؟

۱) آمار توصیفی　　۲) آمار استنباطی　　۳) آمار ناپارامتریک　　۴) هیچکدام

☑ **پاسخ:** گزینه «۱» در آمار استنباطی و آمار ناپارامتریک نمونه‌گیری انجام می‌گیرد در حالی‌که در سرشماری، مرکز آمار ایران همه‌ی واحدهای جامعه را مورد بررسی قرار می‌دهد.

جامعه‌ی آماری یا جمعیّت (Population)

مجموعه‌ی تمام افراد یا اشیایی که مطالعات آماری در مورد یک یا چند صفت آن‌ها در یک مکان و زمان معین انجام می‌گیرد، جامعه‌ی آماری یا جمعیت گفته می‌شود. تعداد اعضای جامعه، حجم جامعه نامیده می‌شود که آن را با N نشان می‌دهند.

نمونه (Sample): در بررسی‌های آماری به دلیل هزینه‌ی زیاد، کمبود وقت و در بعضی مواقع غیرممکن بودن انجام کار، زیرمجموعه‌ای از جامعه با قاعده و ضابطه‌ی خاصی انتخاب می‌شود که به آن نمونه‌ی آماری گفته می‌شود. تعداد اعضای نمونه، حجم نمونه نامیده می‌شود و آن را با n نشان می‌دهند.

صفت مشخصه (Attribute): صفتی است که بین همه‌ی عناصر جامعه‌ی آماری مشترک بوده و جداکننده‌ی جامعه‌ی آماری از سایر جوامع است. به‌طورکلی صفات، خود به دو دسته‌ی کمی و کیفی تقسیم می‌شوند.

صفات کمی: صفات کمی صفاتی هستند که می‌توانند به صورت عددی بیان شوند. تعداد دانشجویان یک دانشگاه، درآمد یک خانواده و ... دارای صفت کمی هستند.

صفات کیفی: صفات کیفی صفاتی هستند که نمی‌توانند به صورت عددی بیان شوند. گروه خون، رنگ چشم انسان‌ها و . . . دارای صفت کیفی هستند.

مثال ۲: کدام‌یک از گزینه‌های زیر تعریف صفت مشخصه است؟ (مدیریت و حسابداری ـ سراسری ۹۳)

۱) از فردی به فرد دیگر تغییر می‌کند.
۲) صفت مشترک بین کلیه‌ی افراد جامعه است.
۳) متمایزکننده‌ی عناصر جامعه از یکدیگر است.
۴) عنصر مشترک جوامع آماری مختلف است.

☑ **پاسخ:** گزینه «۲» صفت مشترک بین اعضاء جامعه است.

انواع مقیاس‌های اندازه‌گیری صفات (مقیاس‌های استیونز)

الف ـ کیفی (گروهی)

این مقیاس، خود به دو مقیاس اسمی و رتبه‌ای تقسیم می‌شود. ماهیت عددی ندارند، قابل اندازه‌گیری نیستند و واحد ندارند. وضعیت تأهل، رنگ چشم، جنسیت و مهارت، مثال‌هایی از مقیاس کیفی هستند.

۱ـ مقیاس اسمی (Nominal Scale): زمانی که صفتی دارای حالت‌های مختلف باشد، مفهوم بهتر (بدتر) و بزرگتر (کوچکتر) بودن را نداشته باشد، همچنین صفت، ماهیت عددی نداشته باشد و بتوان آن را در طبقات جداگانه‌ای قرار داد، مقیاس این صفت اسمی است. به‌طور مثال جنسیت افراد را در نظر بگیرید. دو حالت مرد و زن دارد و مرد و زن بودن بر یکدیگر ارجحیت ندارند. همچنین ماهیت آن عددی نیست.

۲ـ مقیاس رتبه‌ای (Rank Scale): زمانی که صفتی مفهوم بهتر (بدتر) و بزرگتر (کوچکتر) بودن را داشته باشد و نتوان آن را اندازه‌گیری کرد، از این مقیاس استفاده می‌شود. مثلاً طبقه‌بندی مردم یک کشور به سه طبقه‌ی پر درآمد، با درآمد متوسط و کم درآمد تقسیم شود.

ب ـ مقیاس کمّی (عددی)

وسیله یا روشی خاص برای اندازه‌گیری یک صفت است که ماهیت و حاصل آن بصورت عددی است، مانند تعداد فرزندان، طول قد افراد، درجه‌ی حرارت .... .
این مقیاس خود به سه دسته‌ی شمارشی، فاصله‌ای و نسبی (نسبتی) تقسیم می‌شود.

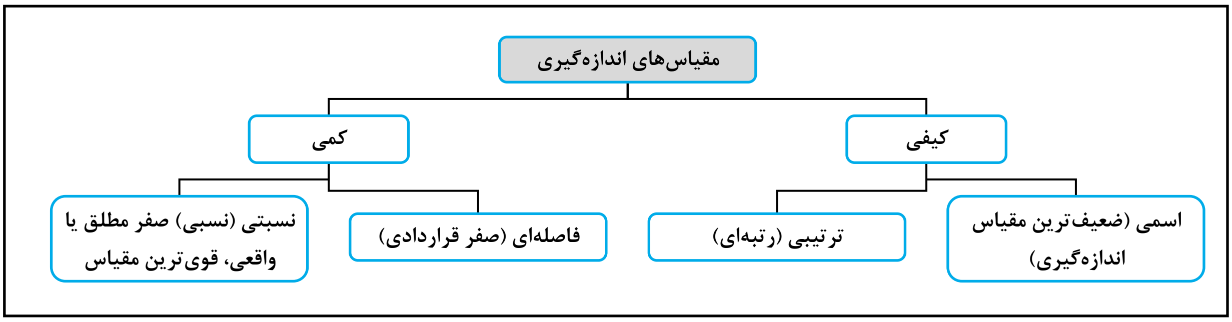
۱ـ مقیاس فاصله‌ای (Interval Scale): زمانی که صفتی دارای ماهیت عددی باشد، اختلاف بین حالتها را بیان کند و مفهوم بهتر (بدتر) و بزرگتر (کوچکتر) بودن را داشته باشد مقیاس آن فاصله‌ای است. اعداد روی دماسنج و یا نمره‌های مربوط به هوش افراد دارای مقیاس فاصله‌ای هستند.

۲ـ مقیاس نسبی (نسبتی) (Ratio Scale): این مقیاس دارای صفر مطلق است و چهار عمل اصلی روی این مقیاس انجام می‌گیرد. طول قد افراد، درآمد و هزینه‌ها همه مثال‌هایی از مقیاس نسبی هستند. در مقیاس نسبتی نسبت حفظ می‌شود، مثلاً اگر برای دو جسم و ویژگی وزن با x_1 و x_2 اندازه‌گیری شود باید $\frac{x_1}{x_2}$ ثابت بماند و به واحد اندازه‌گیری بستگی نداشته باشد.

مثال ۳: در اندازه گیری داده‌های سطح تحصیلات، دمای اجسام و طول قد دانشجویان از چه مقیاس‌هایی استفاده شده است؟

۱) ترتیبی ـ فاصله‌ای ـ نسبتی
۲) فاصله‌ای ـ فاصله‌ای ـ نسبتی
۳) ترتیبی ـ ترتیبی ـ اسمی
۴) اسمی ـ فاصله‌ای ـ فاصله‌ای

☑ **پاسخ:** گزینه «۱» در داده‌های سطح تحصیلات ترتیب رعایت می‌شود ـ دمای اجسام دارای مقیاس فاصله‌ای و طول قد دانشجویان دارای مقیاس نسبی است.



مثال ۴: کدام‌یک از مقیاس‌ها دارای صفر قرار دادی است؟ (مدیریت و حسابداری ـ سراسری ۹۴ ـ طراحی و برنامه‌ریزی شهری ـ سراسری ۸۹)

۱) نسبی
۲) فاصله‌ای
۳) اسمی
۴) رتبه‌ای

☑ **پاسخ:** گزینه «۲» صفر قراردادی به صورت قرارداد صفر در نظر گرفته شده است و مفهوم آن عدد صفر واقعی نیست.

داده‌های آماری (Statistics data)

در بررسی‌های آماری باید صفت مورد بررسی به صورت اعداد و ارقام نمایش داده شود. اگر صفت مورد بررسی کمی باشد، این عمل به سادگی امکان‌پذیر است ولی اگر داده‌ها کیفی باشند، باید طبق ضوابط خاصی با عدد و رقم نمایش داده شوند. به‌طورکلی داده‌های آماری به دو دسته‌ی گسسته (طبقه‌بندی‌نشده) و پیوسته (طبقه‌بندی شده) تقسیم می‌شوند.

داده‌های گسسته (Discrete data): داده‌هایی هستند که بین هر دو مقدار متوالی a مورد نظر از آن‌ها هیچ عددی نمی‌تواند قرار گیرد. تعداد دانشجویان یک دانشکده، تعداد فرزندان یک خانواده و ... به صورت گسسته بیان می‌شوند.

داده‌های پیوسته (Continious data): داده‌هایی هستند که بین هر دو مقدار مورد نظر از آن‌ها بی‌شمار عدد می‌تواند قرار گیرد. طول قد و یا وزن افراد، طول عمر انسان‌ها و ... همه مثالهایی از داده‌های پیوسته هستند. توجه کنید که ماهیت داده‌های پیوسته به صورت اعشاری است، اگرچه در پاره‌ای از مواقع ممکن است آن‌ها را به دلیل رند کردن یا به طور تصادفی به صورت گسسته نمایش دهند، اما این نمایش مهم نیست و ماهیت اعداد برای ما مهم است. به طور مثال اگر وزن فردی ۷۴ کیلوگرم نمایش داده شود این عدد واقعی نیست چرا که وزن، ماهیتی پیوسته دارد و باید اعشاری باشد لذا ما به ماهیت عدد توجه می‌کنیم نه به نوع نمایش آن.

مثال۵: کدام‌یک از موارد زیر داده‌های گسسته می‌باشد؟

۱) قد افراد یک کلاس　　۲) وزن افراد یک شرکت

۳) زمان رسیدن اتوبوس به یک ایستگاه　　۴) تعداد افرادی که به یک فروشگاه وارد می‌شوند.

☑ **پاسخ:** گزینه «۴» بنا به تعریف داده‌های گسسته داریم: داده‌هایی که بین هر دو مقدار متوالی a موردنظر از آن‌ها هیچ عددی نمی‌تواند قرار گیرد بنابراین گزینه‌ی (۴) جواب است.

مراحل یک پژوهش علمی در آمار

در تحقیقات علمی به روش آماری چند مرحله‌ی مهم باید به‌طور کامل انجام شود:

الف) مشخص‌کردن هدف: در این مرحله با کمک روش‌های تحقیق، تلاش برای افزایش اطلاعات از موضوع و بررسی اطلاعات پایدارتر برای رسیدن به هدف انجام می‌گیرد.

ب) جمع‌آوری داده‌ها: در هر پژوهش آماری، تهیه‌ی داده‌های واقعی با توجه به هدفی که از پژوهش داریم، اهمیت اساسی دارد. فرآیند جمع‌آوری داده‌ها ممکن است به روش‌های مختلفی انجام گیرد.

ج) تجزیه و تحلیل داده‌ها: داده‌هایی که به روشهای مناسب گردآوری شده، منبع اساسی برای کسب اطلاعات جدید درباره‌ی پدیده مورد مطالعه هستند در این مرحله داده‌ها را بررسی کرده و معلومات جدیدی به‌دست می‌آوریم که تعیین کننده‌ی نقاط قوّت و ضعف آن‌ها است.

د) بیان یافته‌ها: در این مرحله، تحلیل پایانی بر روی داده‌های آماری که در مرحله‌ی اول تحقیق مشخص شده‌اند، انجام می‌شود و همچنین در این مرحله پاسخ به سؤالات اولیّه امکان پذیر است.

مثال۶: اولین مرحله در تحقیق علمی کدام است؟ (مدیریت بازرگانی ـ آزاد ۸۸)

۱) فرضیه سازی　　۲) تحلیل یافته‌ها　　۳) هدف‌های پژوهش　　۴) جمع‌آوری داده‌ها

☑ **پاسخ:** گزینه «۳» با توجه به ترتیب مراحل گفته شده در بالا هدف‌های پژوهش اولین مرحله است.

مطالعه توصیفی داده‌ها (آمار توصیفی)

مفهوم و کاربرد نمادها $\sum$

مجموعه‌ی داده‌های آماری را معمولاً با نماد $x_1, x_2, \ldots, x_N$ نشان می‌دهند. نماد $\sum$ برای جمع‌کردن داده‌های آماری بکار می‌رود. علامت $\sum$ برای عدم استفاده مکرر از علامت "+" بکار می‌رود. در نماد $\sum$ یک کران پایین، یک کران بالا و یک متغیر (چند متغیر) در جلوی سیگما نوشته می‌شود. جمله‌ای که بعد از نماد $\sum$ نوشته می‌شود، نشان دهنده‌ی مقادیری است که باید با هم جمع شوند و کران‌های پایین و بالای $\sum$ نشان دهنده‌ی ابتدا و انتهای جمع هستند.

به مثال‌های بعد توجه کنید:

۱) $\sum_{i=1}^{5} i = 1+2+3+4+5$

۲) $\sum_{i=1}^{k} x_i = x_1 + x_2 + \cdots + x_k$

۳) $\sum_{i=1}^{4} (x_i+2)^3 = (x_1+2)^3 + (x_2+2)^3 + (x_3+2)^3 + (x_4+2)^3$

۴) $\sum_{i=1}^{3} (x_i-3) = (x_1-3) + (x_2-3) + (x_3-3)$

خواص $\sum$: (c و d اعداد ثابت هستند)

۱) $\sum_{i=1}^{n} c = n.c$

۲) $\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$

۳) $\sum_{i=1}^{n} (x_i+c) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} c = \sum_{i=1}^{n} x_i + nc$

۴) $\sum_{i=1}^{n}(cx_i+dy_i) = \sum_{i=1}^{n} cx_i + \sum_{i=1}^{n} dy_i = c\sum_{i=1}^{n} x_i + d\sum_{i=1}^{n} y_i$

در آمار توصیفی به توصیف اعداد و ارقام می‌پردازیم. توصیف داده‌های آماری بدون توجه به نوع به سه دسته‌ی کلّی تقسیم می‌شود:

۱ـ تنظیم و طبقه‌بندی داده‌ها در یک جدول (جدول فراوانی)۲ـ خلاصه‌کردن داده‌ها به یک یا چند عدد موسوم به شاخص یا آماره

۳ـ رسم نمودارهای گوناگون با استفاده از مقادیر جدول

در اینجا به شرح مراحل بالا می‌پردازیم که برای دو نوع داده‌های طبقه‌بندی نشده (گسسته) و داده‌های طبقه‌بندی شده (پیوسته) تعریف می‌شوند.

جدول فراوانی

این جدول یکی از ساده‌ترین و متداول‌ترین جداول آماری است که شامل پارامترهای مختلف بوده که در زیر به تعریف آن‌ها می‌پردازیم.

فراوانی مطلق و فراوانی نسبی (Frequency-relative Frequency)

اگر n داده از k نوع داشته باشیم و تعداد این داده‌ها در k طبقه به ترتیب $F_1, F_2, \ldots, F_k$ باشند، به تعداد آن‌ها فراوانی‌های مطلق طبقات و به $f_1 = \frac{F_1}{n}$ و $f_2 = \frac{F_2}{n}$ و ... و $f_k = \frac{F_k}{n}$ فراوانی‌های نسبی طبقات گفته می‌شود.

$$1 \le F_i \le n \Rightarrow F_1 + F_2 + \cdots + F_k = \sum_{i=1}^{k} F_i = n \quad ; \quad 0 \le f_i \le 1 \Rightarrow f_1 + f_2 + \cdots + f_k = \sum_{i=1}^{k} f_i = 1$$

فراوانی‌های تجمعی و فراوانی‌های تجمعی نسبی(Cumulative Frequency): اگر در هر طبقه، فراوانی آن طبقه و طبقات ما قبل از آن را با یکدیگر جمع کنیم، فراوانی تجمعی آن طبقه به‌دست می‌آید و اگر در هر طبقه، فراوانی نسبی آن طبقه و طبقات ماقبل را با هم جمع کنیم فراوانی تجمعی نسبی آن طبقه به‌دست می‌آید:

$Fc_j = F_1 + F_2 + \cdots + F_j$ فراوانی تجمعی طبقه j ام

$fc_j = f_1 + f_2 + \cdots + f_j$ فراوانی تجمعی نسبی طبقه j ام

مراحل ساخت یک جدول فراوانی برای داده‌های طبقه‌بندی شده (پیوسته)

۱ـ دریافت داده‌های خام و در صورت لزوم گرد کردن آن‌ها.۲ـ تقسیم داده‌های آماری به تعدادی رده یا طبقه. دو رابطه‌ی تجربی زیر قواعدی مفید هستند:

(دستور استورجس) $K = 1 + 3/322 \operatorname{Log}_{10}^{n}$ تعداد طبقات (دستور تقریبی) $K = \sqrt{n}$ تعداد طبقات

۳ـ میزان تغییرپذیری داده‌ها $= \dfrac{\text{واحد گرد شده‌ی داده‌ها}}{2}$

۴ـ دامنه‌ی داده‌ها را که برابر با تفاضل کوچکترین داده از بزرگترین داده است به‌دست می‌آوریم. (R) $R = \max_i x_i - \min_i x_i$

۵ ـ طول هر رده یا طبقه، از تقسیم دامنه بر تعداد رده (K) به‌دست می‌آید. $I = \dfrac{R}{K}$

مثال ۷: وزن تعدادی دانش‌آموز به‌صورت زیر داده شده است یک جدول فراوانی مناسب برای آن‌ها تشکیل دهید.

۳۷ – ۳۲ – ۳۸ – ۳۴ – ۳۲– ۴۰ – ۳۸ – ۳۲ – ۳۵ – ۳۴ – ۲۹ – ۳۶ – ۳۷ – ۲۴ – ۳۵ – ۳۹ – ۳۳ – ۴۲ – ۴۵– ۳۰ – ۴۳ – ۳۲ – ۲۹ – ۲۷ – ۲۲

پاسخ: مراحل ساخت جدول فراوانی را به‌ترتیب اجرا می‌کنیم:

$$K = 1 + 3/322 \operatorname{Log}_{10} 25 = 1 + 3/322(1/39) = 5/64 \simeq 6$$

توجه کنید که داده‌های داده شده مربوط به وزن دانش‌آموزان است بنابراین پیوسته هستند، داده‌های پیوسته باید به صورت اعشاری باشند بنـابراین متوجـه می‌شویم که آن‌ها را رند کرده‌اند که باعث حذف ممیز آن‌ها شده است در تنظیم جدول فراوانی اگر داده‌ها رند شده باشند باید وضعیت آن‌ها را قبـل از رنـد شدن در جدول نشان دهیم بنابراین: داده‌ها با تقریب ۱، گرد شده‌اند چون ممیز آن‌ها از بین رفته است:

میزان تغییرپذیری $= \frac{۱}{۲} = ۰/۵$

$R = \max_i x_i - \min_i x_i = ۴۵/۵ - ۲۱/۵ = ۲۴$ $\qquad I = \frac{R}{K} = \frac{۲۴}{۶} = ۴$

حدود واقعی طبقات	x_i	F_i	f_i	Fc_i	fc_i
۲۱/۵ ـ ۲۵/۵	۲۳/۵	۲	۰/۰۸	۲	۰/۰۸
۲۵/۵ ـ ۲۹/۵	۲۷/۵	۳	۰/۱۲	۵	۰/۲۰
۲۹/۵ ـ ۳۳/۵	۳۱/۵	۶	۰/۲۴	۱۱	۰/۴۴
۳۳/۵ ـ ۳۷/۵	۳۵/۵	۷	۰/۲۸	۱۸	۰/۷۲
۳۷/۵ ـ ۴۱/۵	۳۹/۵	۴	۰/۱۶	۲۲	۰/۸۸
۴۱/۵ ـ ۴۵/۵	۴۳/۵	۳	۰/۱۲	۲۵	۱

توجه کنید که در جدول فراوانی، x_i نماینده‌ی دسته (مرکز دسته) حد وسط حدود واقعی طبقات است. همچنین توجه کنید که انتهای هر دسته با ابتـدای دسته‌ی بعد مساوی است. در جداول فراوانی این پیوستگی باید رعایت شود و اگر جدول این پیوستگی را نداشت باید نصف فاصله بین دو طبقه‌ی متـوالی از کران پایین هر طبقه کم کرده و به انتهای آن اضافه کنیم که در محاسبه‌ی شاخص در ادامه‌ی این موضوع نشان داده خواهد شد.

تفسیر بعضی از اعداد جدول:

در طبقه‌ی چهارم $F_۴ = ۷$ است این عدد به مفهوم آن است که ۷ نفر از دانش‌آموزان وزنشان در فاصله‌ی $(۳۷/۵, ۳۳/۵]$ قرار دارد.

در طبقه سوم $fc_۳ = ۰/۴۴$ است این عدد به مفهوم آن است که ۴۴ درصد از دانش‌آموزان دارای وزنی کمتر از ۳۳/۵ یا در فاصله $(۳۳/۵ - ۲۱/۵]$ قرار دارند.

نکات مربوط به جدول

همواره جمع فراوانی‌های مطلق برابر با تعداد کل داده‌ها است. $\quad$ ۱) $\sum_{i=۱}^{k} F_i = n$

همواره جمع فراوانی‌های نسبی برابر با ۱ است. $\quad$ ۲) $\sum_{i=۱}^{k} f_i = ۱$

فراوانی نسبی هر طبقه، از تقسیم فراوانی مطلق آن طبقه بر کل داده‌ها حاصل می‌شود. $\quad$ ۳) $f_i = \frac{F_i}{n}$

فراوانی تجمعی طبقه‌ی آخر جدول برابر با تعداد کل داده‌ها است. $\quad$ ۴) $Fc_k = n$

فروانی تجمعی نسبی طبقه‌ی آخر جدول همواره برابر با ۱ است. $\quad$ ۵) $fc_k = ۱$

تفاضل فراوانی تجمعی دو طبقه‌ی متوالی برابر با فراوانی مطلق طبقه‌ی پایین‌تر است. $\quad$ ۶) $F_i = Fc_i - Fc_{i-۱}$

تفاضل فراوانی تجمعی نسبی دو طبقه‌ی متوالی برابر با فراوانی نسبی طبقه‌ی پایین‌تر است. $\quad$ ۷) $f_i = fc_i - fc_{i-۱}$

مثال ۸: کدام‌یک از روابط زیر صحیح است؟ (محیط زیست ـ سراسری ۸۵)

۱) فراوانی طبقه iام = $\frac{\text{فراوانی کل}}{\text{طول فاصله طبقه iام}}$ $\qquad$ ۲) فراوانی طبقه iام = $\frac{\text{فراوانی نسبی طبقه iام}}{\text{فراوانی کل}}$

۳) فراوانی نسبی طبقه iام = $\frac{\text{فراوانی طبقه iام}}{\text{فراوانی کل}}$ $\qquad$ ۴) فراوانی نسبی طبقات × فراوانی طبقه iام = فراوانی کل

☑ **پاسخ:** گزینه «۳» فراوانی نسبی هر طبقه، از تقسیم فراوانی مطلق همان طبقه بر مجموع فراوانی‌ها (کل داده‌ها) بدست می‌آید. $f_i = \frac{F_i}{\sum F_i} = \frac{F_i}{N}$

مثال۹: در یک جدول فراوانی تعداد طبقات $k = 7$، تعداد مشاهدات $n = 120$، فراوانی تجمعی طبقه ۵ برابر $F_5 = 88$ و فراوانی تجمعی طبقه ۶ برابر $F_6 = 108$ هستند. فراوانی نسبی طبقات ۶ و ۷ به‌ترتیب از راست به چپ کدام است؟ (علوم اقتصادی ـ سراسری ۹۷)

۱) ۲۰ و ۱۲ ۲) $\frac{1}{6}$ و $\frac{1}{10}$ ۳) $\frac{1}{6}$ و ۱ ۴) $\frac{5}{6}$ و ۱

پاسخ: گزینه «۲» فراوانی نسبی طبقه ۶ به‌صورت $f_6^* = \frac{F_6 - F_5}{\sum_{i=1}^{7} f_i}$ قابل محاسبه است که خواهیم داشت:

$$f_6^* = \frac{108-88}{120} = \frac{1}{6}$$

برای محاسبه فراوانی نسبی طبقه ۷، با توجه به اینکه فراوانی تجمعی تا طبقه ۶ برابر $F_6 = 108$ می‌باشد و به‌طورکلی ۷ طبقه داریم بنابراین:

$$f_7^* = 1 - \frac{108}{120} = \frac{1}{10}$$

زیرا مجموع فراوانی‌های نسبی برابر یک می‌باشد (مانند احتمال).

مثال۱۰: فرض کنید در یک جدول فراوانی، فراوانی تجمعی تا انتهای یک طبقه برابر ۳۷ و تا قبل از همان طبقه برابر ۲۵ است. اگر تعداد مشاهدات $n = 120$ باشد، فراوانی نسبی این طبقه کدام است؟ (علوم اقتصادی ـ سراسری ۹۶)

۱) ۰/۱ ۲) $\frac{5}{24}$ ۳) $\frac{37}{120}$ ۴) $\frac{11}{120}$

پاسخ: گزینه «۱» تفاضل دو فراوانی تجمعی پشت سر هم در جدول فراوانی، فراوانی مطلق را نتیجه می‌دهد، بنابراین:

$$F_i = Fc_i - Fc_{i-1} = 37 - 25 = 12$$

طبق رابطه فراوانی نسبی و مطلق داریم:

$$f_i = \frac{F_i}{n} = \frac{12}{120} = 0/1$$

مثال۱۱: در ۴۰ داده‌ی آماری در یک جدول فراوانی که در ۸ دسته طبقه‌بندی شده است اگر مجموع فراوانی‌های نسبی تا طبقه‌ی هفتم برابر با ۰/۷۵ باشد فراوانی مطلق طبقه‌ی هشتم کدام است؟

۱) ۹ ۲) ۱۱ ۳) ۱۰ ۴) ۱۲

پاسخ: گزینه «۳» می‌دانیم مجموع فراوانی‌های نسبی در <u>۸</u> طبقه برابر با <u>۱</u> است.

$$f_1 + f_2 + \cdots + f_8 = 1$$

بنابراین فراوانی نسبی طبقه‌ی آخر برابر با ۰/۲۵ است.

$$f_8 = 1 - 0/75 = 0/25$$

اکنون فراوانی نسبی طبقه‌ی آخر و تعداد داده‌ها را داریم، پس فراوانی مطلق را به‌دست می‌آوریم:

$$f_8 = \frac{F_8}{n} \Rightarrow \frac{25}{100} = \frac{F_8}{40} \Rightarrow F_8 = 10$$

مثال۱۲: در ۱۲۰ داده‌ی آماری کوچک‌ترین و بزرگترین مقادیر به ترتیب ۳۵ و ۵۷ هستند. این داده‌ها در ۹ طبقه دسته‌بندی شده‌اند. اگر ۳۲ درصد داده‌ها کمتر از ۴۵ و همچنین ۴۷ درصد داده‌ها کمتر از ۴۷/۵ باشند، فراوانی مطلق دسته‌ی وسط کدام است؟

۱) ۱۲ ۲) ۱۵ ۳) ۱۶ ۴) ۱۸

پاسخ: گزینه «۴» ابتدا دسته‌ی وسط را مشخص می‌کنیم:

$$R = \max x_i - \min x_i = 57 - 35 = 22 \Rightarrow (\text{فاصله طبقات})\ I = \frac{R}{K} = \frac{22}{9} \simeq 2/5$$

در ۹ دسته، طبقه‌ی وسط طبقه‌ی پنجم است. اکنون نقطه‌ی ابتدایی (min) و فاصله‌ی طبقات را داریم، طبقات به صورت زیر هستند:

$$35 - 37/5\ ,\ 37/5 - 40\ ,\ 40 - 42/5\ ,\ 42/5 - 45\ ,\ \underbrace{45 - 47/5}_{\text{دسته وسط}}$$

اکنون با توجه به این‌که ۳۲% از داده‌ها کمتر از ۴۵ و ۴۷% از داده‌ها کمتر از ۴۷/۵ است، می‌توانیم به صورت زیر فراوانی نسبی طبقه‌ی وسط (پنجم) را به‌دست آوریم و سپس با داشتن فراوانی نسبی طبقه‌ی پنجم و تعداد داده‌ها، فراوانی مطلق این طبقه را مشخص کنیم:

$$fc_5 - fc_4 = f_5 \Rightarrow 0/47 - 0/32 = 0/15 \Rightarrow f_5 = \frac{F_5}{n} \Rightarrow \frac{15}{100} = \frac{F_5}{120} \Rightarrow F_5 = 18$$

درسنامه (۲): خلاصه کردن داده‌ها به یک یا چند عدد به نام شاخص یا آماره

با استفاده از شاخص‌ها یا آماره‌ها می‌توانیم نتایج کلی را به صورت ساده‌تر ارائه دهیم شاخص‌ها خود به سه بخش شاخص‌های تمرکز یا متمرکز(مرکزی)، شاخص‌های پراکندگی و شاخص‌های نسبی پراکندگی تقسیم می‌شوند.

۱. شاخص‌های مرکزی (Measure of centeral tendency)

این شاخص‌ها میزان تمرکز داده‌ها را نشان می‌دهند شاخص‌های مرکزی عبارتند از:

میانگین ــ میانه ــ مد

میانگین: اگر داده‌ها بر روی یک محور به صورت منظم ردیف شوند، مقدار میانگین، نقطه تعادل یا مرکز ثقل داده‌ها است میانگین با توجه به نوع صفت اندازه‌گیری شده دارای انواع گوناگونی است که در زیر به معرفی آن‌ها می‌پردازیم. شاخص مرکزی میانگین برای متغیرهای کمی استفاده می‌شود.

میانگین حسابی: معدل داده‌ها را میانگین حسابی می‌نامند. میانگین حسابی برابر است با مجموع داده‌ها تقسیم بر تعداد آن‌ها:

$$\mu_x = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

در پاره‌ای از مواقع ممکن است داده‌ها دارای تکرار F_i باشند، یعنی داده‌ها دارای ارزش و اهمیت یکسانی نباشند. در این صورت میانگین حسابی را میانگین وزنی یا موزون می‌نامند. (Weighted Mathematical Mean)

f_i : فراوانی نسبی F_i : فراوانی مطلق

$$\mu = \frac{F_1x_1 + F_2x_2 + \dots + F_Nx_N}{F_1 + F_2 + \dots + F_N} = \frac{\sum_{i=1}^{N} F_i x_i}{\sum_{i=1}^{N} F_i} = \sum_{i=1}^{N} f_i x_i$$

میانگین جامعه را با حرف یونانی "μ" نشان می‌دهند و آن را "مو" تلفظ می‌کنند.

ولی میانگین نمونه‌ای از جامعه را با $\overline{X}$ نمایش می‌دهند تفاوت علائم در آمار بسیار پر اهمیت است. در فصل‌های بعد خواهیم دید که این تفاوت علائم در روابط و مفاهیم آماری باعث به وجود آمدن مفاهیم متفاوت خواهد شد، لذا به دانشجویان گرامی توصیه می‌شود به این تفاوت‌ها کاملاً اهمیت دهند.

نکات مربوط به میانگین

۱ـ میانگین داده‌های مساوی $(a,a,a,a,a,\dots)$ برابر با خود آن‌ها است که در اینجا منظور از a یک عدد کمی می‌باشد. $(\mu = a)$

۲ـ مجموع جبری اختلاف داده‌ها از میانگین برابر با صفر است.

$$\sum_{i=1}^{N} F_i (x_i - \mu) = 0$$

۳ـ اگر به تک‌تک داده‌ها مقدار ثابتی مانند a را اضافه یا از آن‌ها مقدار ثابت a را کم کنیم، به میانگین نیز مقدار ثابت a اضافه یا کم خواهد شد.

$$y_i = x_i \pm a \Rightarrow \overline{y} = \overline{x} \pm a$$

۴ـ اگر تک‌تک داده‌ها را در عدد ثابتی مانند b ضرب کنیم میانگین نیز در b ضرب خواهد شد. $y_i = bx_i \Rightarrow \overline{y} = b\overline{x}$

۵ـ مجموع توان دوم انحرافات نقاط از میانگین همیشه حداقل است. این خاصیت به مفهوم آن است که مجموع توان دوم انحرافات از میانگین از مجموع توان دوم انحرافات نسبت به هر عدد دلخواه دیگری مانند $a \neq \mu$ کوچک‌تر است.

$$\sum_{i=1}^{n} (x_i - \mu)^2 < \sum_{i=1}^{n} (x_i - a)^2$$

مثال ۱۳: در داده‌های آماری دسته‌بندی شده جدول زیر میانگین حسابی کدام است؟ (برنامه‌ریزی محیط زیست ـ سراسری ۸۹)

حدود دسته	۲۱۵_۲۲۵	۲۲۵_۲۳۵	۲۳۵_۲۴۵	۲۴۵_۲۵۵	۲۵۵_۲۶۵
فروانی تجمعی	۱۱	۲۸	۵۱	۷۲	۹۰

۱) ۲۴۱　۲) ۲۴۲
۳) ۲۴۳　۴) ۲۴۴

☑ **پاسخ:** گزینه «۲» با توجه به اینکه فراوانی‌های تجمعی داده شده است، ابتدا فروانی‌های مطلق را به‌دست می‌آوریم:

$$F_i = FC_i - FC_{i-۱} \qquad F_i = ۱۱,۱۷,۲۳,۲۱,۱۸$$

اکنون مقادیر x_i ها (مرکز دسته‌ها) را محاسبه می‌کنیم:

$$x_i : \frac{۲۱۵+۲۲۵}{۲} = ۲۲۰ \ ,\ ۲۲۰+۱۰=۲۳۰ \ ,\ ۲۳۰+۱۰=۲۴۰ \ ,\ ۲۴۰+۱۰=۲۵۰ \ ,\ ۲۵۰+۱۰=۲۶۰$$

در رابطه‌ی میانگین حسابی قرار می‌دهیم:

$$\mu = \frac{\sum F_i x_i}{\sum F_i} = \frac{۱۱\times۲۲۰+۱۷\times۲۳۰+۲۳\times۲۴۰+۲۱\times۲۵۰+۱۸\times۲۶۰}{۱۱+۱۷+۲۳+۲۱+۱۸} = ۲۴۲$$

مثال ۱۴: اطلاعات جدول زیر پس از کسر عدد ۹۵ از مقادیر اصلی و تقسیم هریک از ارقام به‌دست آمده بر ۱۰۰ نتیجه شده است. میـانگین مقـادیر اصلی کدام است؟

x_i	-۳	-۲	-۱	۰	۱	۲
F_i	۲	۴	۵	۶	۲	۱

۱) ۱۰　۲) ۲۰
۳) ۳۰　۴) ۴۰

☑ **پاسخ:** گزینه «۲» اگر مقادیر اصلی را y بنامیم رابطه y با x (تغییرات بر روی داده‌ها) به صورت زیر است:

$$x = \frac{y-۹۵}{۱۰۰} \Rightarrow ۱۰۰x = y-۹۵ \Rightarrow y = ۱۰۰x+۹۵$$

با توجه به خاصیت (۳ و ۴) میانگین خواهیم داشت:

$$\mu_y = ۱۰۰\mu_x + ۹۵$$

بنابراین مقدار μ_x را به‌دست آورده و در رابطه‌ی بالا قرار می‌دهیم:

$$\mu_x = \frac{\sum F_i x_i}{\sum F_i} = \frac{(۲\times-۳)+(۴\times-۲)+(۵\times-۱)+(۶\times۰)+(۲\times۱)+(۱\times۲)}{۲+۴+۵+۶+۲+۱} = -۰/۷۵$$

$$\Rightarrow \mu_y = ۱۰۰(-۰/۷۵)+۹۵ = -۷۵+۹۵ = ۲۰$$

مثال ۱۵: میانگین ۱۰ داده آماری برابر عدد ۵ محاسبه شده است. پس از محاسبه معلوم گردید که دو مقدار ۱۰ و ۱۲ نیز باید به داده‌ها اضافه شـود، میانگین جدید کدام است؟ (حسابداری ـ آزاد ۸۸)

۱) ۵　۲) ۶　۳) ۷　۴) ۸

☑ **پاسخ:** گزینه «۲» میانگین ۱۰ داده آماری برابر با ۵ است، بنابراین جمع این داده‌ها برابر با ۵۰ است.

$$\mu = \frac{\sum x_i}{N} \Rightarrow \sum x_i = N\times\mu = ۱۰\times۵ = ۵۰$$

اکنون به جمع داده‌ها اعداد ۱۰ و ۱۲ را نیز اضافه می‌کنیم، بنابراین به تعداد داده‌ها نیز ۲ داده اضافه می‌شود.

$$\mu = \frac{(\text{جدید})\sum x_i}{(\text{جدید})N} = \frac{۵۰+۱۲+۱۰}{۱۰+۲} = \frac{۷۲}{۱۲} = ۶$$

میانگین پیراسته (Trimmed Mean)

در برخی از داده‌ها، تعدادی از مشاهدات با بقیه‌ی داده‌ها همخوانی و تجانس ندارند. توجه داشته باشید که تعدادی از داده‌ها که به‌طور غیرعـادی بـزرگ یـا کوچک باشند اثر زیادی بر روی میانگین جامعه دارند. برای این دسته از داده‌ها ابتدا داده‌هـا، را بـه صـورت صـعودی مرتـب مـی‌کنـیم، سـپس درصـدی از مشاهدات برابر با [N.P] از پایین و از بالا حذف می‌شوند (توجه کنید که در اینجا تعداد داده‌ها و درصدی است که در صورت مسأله مشخص می‌شود) و در پایان، میانگین حسابی مشاهدات باقیمانده را محاسبه می‌کنیم.

مثال ۱۶: در جدول زیر کدام میانگین قابل محاسبه است؟

x_i	کمتر از ۲۰	۲۱	۲۳	۲۴	بزرگتر از ۲۵
F_i	۵	۱۰	۷	۳	۲

۱) هیچ میانگینی قابل محاسبه نیست　　۲) میانگین هندسی　　۳) میانگین پیراسته　　۴) میانگین هارمونیک

☑ **پاسخ:** گزینه «۳» گاهی ممکن است اطلاعات به گونه‌ای باشند که نتوان برای آن‌ها جداول منظمی به وجود آورد، در عمل جداول مربوط به مرگ و میر انسان‌ها می‌توانند این‌گونه باشند در این‌صورت ابتدا یا انتها یا هر دو طرف جدول است. در این‌گونه جداول ما نمی‌توانیم در طبقه اول و طبقه آخر مرکز دسته $\left(X_i = \frac{\text{حد بالای هر طبقه + حد پایین هر طبقه}}{۲}\right)$ را بدست آوریم چرا که حد آن‌ها مشخص نیست و بنابراین نمی‌توان $\mu = \frac{\sum_{i=۱}^{k} F_i X_i}{\sum_{i=۱}^{k} F_i}$ را محاسبه کرد، لذا از میانگین پیراسته استفاده می‌کنیم همان‌طور که مشاهده شد میانگین پیراسته درصدی از پایین و بالای اطلاعات را حذف می‌کند در ادامه نیز خواهیم دید که همه‌ی شاخص‌های آماری که به نوعی وابسته به X_i (مرکز دسته‌ها) هستند، در جداول باز قابل محاسبه نیستند.

مثال ۱۷: هزینه‌ی ماهانه یک خانواده برحسب صدهزار ریال به این صورت است: ۱۰,۸,۱۵,۲۰,۹,۱۶,۱۷,۱۸,۳۰,۲۵,۱۴/۵,۱۵ میانگین پیراسته کدام است؟

۱) ۱۵/۹۱　　۲) ۱۴　　۳) ۱۳/۷　　۴) ۱۲/۵

☑ **پاسخ:** گزینه «۱» تعداد داده‌هایی که باید از ابتدا و انتها حذف شوند برابر است با:

$$N.P = ۱۲ \times ۰/۲۵ = ۳$$

$$\underbrace{۸,۹,۱۰}_{\text{پایین } \%۲۵},۱۴/۵,۱۵,۱۵,۱۶,۱۷,۱۸,\underbrace{۲۰,۲۵,۳۰}_{\%۲۵}$$

اکنون میانگین باقیمانده‌ی داده‌ها را به‌دست می‌آوریم:

$$\mu'_x = \frac{\sum_{i=۱}^{N} X'_i}{N} = \frac{۱۴/۵+۱۵+۱۵+۱۶+۱۷+۱۸}{۶} = ۱۵/۹۱$$

میانگین وینزوری (winsorized mean)

نوعی دیگر از میانگین پیراسته است که در آن به جای حذف تعداد $[N.P]$ داده از پایین و از بالا، نماینده‌ای از داده‌ها را در پایین و بالای داده‌ها قرار می‌دهیم و سپس میانگین کل مشاهدات را به‌دست می‌آوریم. به مثال زیر توجه کنید:

مثال ۱۸: درآمد چند بازاریاب بصورت زیر است میانگین وینزوری کدام است؟

۹۳/۹,۱۰۵/۸,۱۰۶/۵,۱۱۶/۶,۱۲۵,۱۲۸/۳,۱۳۲/۱,۱۳۶/۷,۳۵۲/۴

۱) ۱۲۱/۷　　۲) ۱۲۰/۶　　۳) ۱۱۹/۸　　۴) ۱۲۱

☑ **پاسخ:** گزینه «۲» به جای ۲۵% پایین مقدار ۱۰۶/۵ و به جای ۲۵% بالا عدد ۱۳۲/۱ قرار می‌گیرد. سومین مشاهده از ابتدا و انتها را قرار می‌دهیم. دو عدد از پایین و دو عدد از بالا حذف کرده و به جای آن‌ها اولین عدد از پایین و آخرین عدد از بالا را قرار می‌دهیم.

توجه کنید که اگر مقدار $N.P$ عدد صحیح نشد، جزء صحیح آن را در نظر می‌گیریم.

$$۹ \times ۰/۲۵ = ۲/۲۵ \Rightarrow [۲/۲۵] = ۲$$

$$\mu'_x = \frac{۱۰۶/۵+۱۰۶/۵+۱۰۶/۵+۱۱۶/۶+۱۲۵+۱۲۸/۳+۱۳۲/۱+۱۳۲/۱+۱۳۲/۱}{۹} = ۱۲۰/۶$$

میانگین هندسی (Geometric Mean)

در داده‌هایی که اندازه‌های آن‌ها نسبی یا بصورت درصد یا نرخ رشد و شاخص باشند، از میانگین هندسی استفاده می‌شود که رابطه‌ی آن به صورت زیر است.
در این رابطه X_i مشاهده iام و f_i تعداد تکرار یا فراوانی داده‌ی i ام است. توجه کنید که در رابطه‌ی میانگین هندسی x_i ها حتماً باید بر حسب درصد یا نسبت باشند. اگر این طور نبود، ابتدا باید آن‌ها را به درصد یا به نسبت تبدیل کنیم.

$$\boxed{\mu_G = \sqrt[N]{x_1^{F_1} \dots x_k^{F_k}}} \text{ که در آن } N = \sum_{i=1}^{k} F_i$$

مثال ۱۹: کدام شاخص برای محاسبه متوسط نرخ رشد اقتصادی مناسب است؟ (علوم اقتصادی ـ سراسری ۹۶)

۱) میانگین هندسی ۲) میانگین حسابی ۳) میانه ۴) نما

☑ **پاسخ:** گزینه «۱» برای محاسبه متوسط داده‌هایی که به‌صورت نسبتی یا درصد (مانند نرخ‌ها) می‌باشند از میانگین هندسی استفاده می‌شود.

مثال ۲۰: میانگین هندسی داده‌های آماری $۲, \frac{۲}{۲۷}, \frac{۸}{۳}, \frac{۲۷}{۱۶}, \frac{۳}{۲}$ کدام است؟ (مدیریت و حسابداری ـ سراسری ۹۴)

۱) ۱/۲ ۲) ۱ ۳) ۱/۵ ۴) ۱/۶

☑ **پاسخ:** گزینه «۲» مقادیر را در فرمول میانگین هندسی قرار می‌دهیم:

$$\mu_G = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \dots x_k^{f_k}} = \sqrt[5]{\frac{۳}{۲} \times \frac{۲۷}{۱۶} \times \frac{۸}{۳} \times \frac{۲}{۲۷} \times ۲} = \sqrt[5]{۱} = ۱$$

مثال ۲۱: واردات یک قلم کالا طی سال‌های ۱۹۷۳ تا ۱۹۷۸ به صورت زیر به دست آمده است. سرعت رشد واردات این کالا کدام است؟
(مدیریت بازرگانی ـ آزاد ۸۲)

واردات بر حسب میلیون تن	۱۳۲	۱۳۶	۱۴۱	۱۴۵	۱۴۷	۱۵۱

۱) ۱/۰۱۷ ۲) ۱/۰۲۲ ۳) ۱/۰۱۱ ۴) ۱/۰۲۷

☑ **پاسخ:** گزینه «۴» سرعت رشد را باید از میانگین هندسی بدست آورد لذا ابتدا رشدها (x_i) را بدست می‌آوریم یعنی نسبت واردات هر سال بـه سـال قبل را محاسبه می‌کنیم:

$$\mu_G = \sqrt[N]{x_1^{F_1} . x_2^{F_2} \cdots x_k^{F_k}} = \sqrt[5]{\frac{۱۳۶}{۱۳۲} \times \frac{۱۴۱}{۱۳۶} \times \frac{۱۴۵}{۱۴۱} \times \frac{۱۴۷}{۱۴۵} \times \frac{۱۵۱}{۱۴۷}} = \sqrt[5]{\frac{۱۵۱}{۱۳۲}} = ۱/۰۲۷$$

مثال ۲۲: رشد سود سالیانه یک فروشگاه در ۴ سال متوالی به صورت روبرو بدست آمده است: ۸ و ۴ و ۴ و ۲ متوسط رشد سود سالیانه چقدر است؟
(حسابداری ـ آزاد ۸۹)

۱) ۳ ۲) ۶ ۳) ۴/۵ ۴) ۴

☑ **پاسخ:** گزینه «۱» با توجه به این که متوسط رشد سالانه خواسته شده است از میانگین هندسی استفاده می‌کنیم:

$$\mu_G = \sqrt[N]{x_1^{f_1} . x_2^{f_2} \dots x_k^{f_k}} = \sqrt[4]{۲ \times ۴ \times ۴ \times ۸} = \sqrt[4]{۲ \times ۲^۲ \times ۲^۲ \times ۲^۳} = \sqrt[4]{۲^۸} = ۲^{\frac{۸}{۴}} = ۲^۲ = ۴$$

اکنون چون رشد سالیانه خواسته شده میانگین هندسی را از رشد سال پایه ۱ کم می‌کنیم. $G - ۱ = ۴ - ۱ = ۳$ متوسط رشد سالیانه

مثال ۲۳: فروش کارخانه آلفا در سال گذشته ۹۱ درصد کاهش و امسال ۴۴ درصد افزایش داشته است. در این صورت متوسط نرخ رشد فروش در این دو سال، چند درصد است؟ (علوم اقتصادی ـ سراسری ۹۵)

۱) -۶۴ ۲) -۸ ۳) ۶ ۴) ۳۶

☑ **پاسخ:** گزینه «۱» اگر سال پایه ۱۰۰% در نظر گرفته شود، ۹۱ درصد کاهش برابر است با: ۹% و ۴۴ درصد افزایش برابر با ۱۴۴% می‌باشد. اکنون متوسط نرخ رشد در دو سال برابر است با:

$$\text{میانگین هندسی } G = \sqrt{x_1 \times x_2} = \sqrt{۱۴۴ \times ۹} = \%۳۶ \quad , \quad \text{نرخ رشد سالیانه} = G - ۱ = \%۳۶ - \%۱۰۰ = -\%۶۴$$