صفحه ۱ دکتر زارع

به نام خدا

روشهای آماری توصیفی

- نمادگذاری

تعداد نمونه n ، تعداد جمعیت N

نماد متغیر در نمونه $x, y, z, \dots$

نماد متغیر در واحد i ام $x_i, y_i, z_i, \dots$

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

$$\sum_{i=1}^{n} a = n \times a$$ a یک مقدار ثابت است

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i \quad , \quad \sum_{i=1}^{n} (x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} x_i y_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \quad , \quad \sum_{i=1}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

$$\sum_{i=1}^{n} i = 1 + 2 + \dots + n = \frac{n(n+1)}{2}$$

تمرین :

$x_1 = 5$	$z_1 = -1$	$\sum_{i=1}^{3} x_i^2 = ?$
$x_2 = 4$	$z_2 = 4$	
$x_3 = -3$	$z_3 = 5$	$\sum_{i=3}^{5} (3x_i + 2y_i) = ?$
$x_4 = 1$	$z_4 = 1$	
$x_5 = 7$	$z_5 = 2$	$\left(\sum_{i=1}^{5} z_i\right)^2 = ?$

$$\left(\sum_{i=3}^{5} x_i\right)\left(\sum_{i=1}^{5} z_i\right) = ? \qquad \sum_{i=1}^{4} 2x_i z_i = ?$$

$$\sum_{i=1}^{3} \frac{x_i}{y_i} = ? \qquad \sum_{i=157}^{550} i = ?$$

— شاخصهای مرکزی

— میانگین حسابی   نماد میانگین در جمعیت : μ ,   نماد نمونه : $\bar{x}$, $\bar{y}$, $\bar{z}$, ...

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

— وقتی داده ها گروه بندی شده باشند (بصورت ساده)

x_i	x_1	x_2	x_3	....	x_k
فراوانی f_i	f_1	f_2	f_3	...	f_k

تعداد نمونه: $\sum_{i=1}^{k} f_k = n$

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

به این میانگین، میانگین حسابی موزون می‌گویند

— وقتی داده‌ها بصورت زیر گروه بندی شده‌اند

گروه بندی	فراوانی f_i	نقطه وسط m_i	$f_i m_i$
$L_1 - U_1$	f_1	m_1	$f_1 m_1$
$L_2 - U_2$	f_2	m_2	$f_2 m_2$
$L_3 - U_3$	f_3	m_3	$f_3 m_3$
⋮	⋮	⋮	⋮
$L_k - U_k$	f_k	m_k	$f_k m_k$

نماد :
L_i : کرانه پائینی طبقه i ام
U_i : کرانه بالای طبقه i ام
f_i : فراوانی طبقه i ام
نقطه وسط طبقه i ام : $m_i = \frac{U_i + L_i}{2}$

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i m_i}{\sum_{i=1}^{k} f_i}$$

البته در این حالت $\bar{x}$ برآوردی از میانگین واقعی است. اندازه میانگین واقعی قابل محاسبه نیست

تمرین: میانگین را در دو حالت زیر بدست آورید.

x_i	f_i
۵	۳
۶	۲
۱۰	۵
۱۲	۶
۱۵	۴

طبقه بندی	۵–۹	۱۰–۱۴	۱۵–۱۹	۲۰–۲۴	۲۵–۲۹
فراوانی	۷	۱۰	۱۳	۱۱	۶

خواص میانگین حسابی

- میانگین حسابی یک شاخص ساده و یکتاست و همه مقادیر در آن نقش دارند و نقش آن‌ها یکسان است

- مجموع انحرافات کلیه مقادیر از میانگین خودشان برابر است با صفر یعنی :

$$\sum_{i=1}^{n}(x_i-\bar{x})=0 \quad \text{یا} \quad \sum_{i=1}^{N}(x_i-\mu)=0$$

- اضافه کردن یک مقدار ثابت مثل a یا کم کردن یک مقدار ثابت مثل a از همه مقادیر معادل است با اضافه شدن یا کم شدن همین مقدار ثابت از میانگین

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

- ضرب کردن همه مقادیر در ثابت a یا تقسیم کردن همه مقادیر بر ثابت a معادل است با ضرب همین ثابت در میانگین یا تقسیم کردن میانگین بر همین ثابت a :

$$\frac{\sum_{i=1}^{n}a x_i}{n}=a\bar{x} \quad , \quad \frac{\sum_{i=1}^{n}\frac{x_i}{a}}{n}=\frac{\bar{x}}{a}$$

- اگر متغیر z از مجموع دو متغیر x و y بدست آمده باشد

$$\bar{z}=\bar{x}+\bar{y}$$

- مهمترین خاصیت میانگین حسابی این است که مجموع مربعات انحرافات از میانگین نسبت به هر نقطه دیگر مینیمم است یعنی به ازاء هر مقدار $a\neq\bar{x}$

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

یا $\sum_{i=1}^{n}(x_i-\bar{x})^2<\sum_{i=1}^{n}(x_i-a)^2 \quad , \quad a\neq\bar{x}$

تمرین: خاصیت فوق را بررسی کنید

x_i	۰	۱	۲	۳	۴
f_i	۱۰	۴۰	۱۰	۲۰	۲۰

برای $\begin{cases} a = ۳ \\ a = ۱ \end{cases}$

- عیب میانگین حسابی: مهمترین عیب میانگین حسابی این است که داده‌هایی که از مجموعه داده‌ها خیلی بزرگتر یا خیلی کوچکتر هستند، باعث می‌شوند تاثیر زیادی در مقدار میانگین بگذارند و به طور غیرواقعی میانگین را بزرگتر یا کوچکتر جلوه دهند. چنانچه بخواهیم براین اساس تصمیم‌گیری یا قضاوت کنیم، این تصمیم‌گیری یا قضاوت خالی از اشکال نخواهد بود.

- میانگین پیراسته Trim mean

برای رفع مشکل فوق (رفع مشکل عیب میانگین حسابی) می‌توان پس از مرتب کردن داده‌ها به صورت صعودی یا نزولی درصدی از بزرگترین و درصدی از کوچکترین داده‌ها را حذف کرد و بعد میانگین حسابی را از بقیه داده‌ها حساب کرد. معمولاً بین ۵ تا ۱۰ درصد از داده‌ها یعنی ۲٫۵ تا ۵ درصد از بزرگترین داده‌ها و ۲٫۵ تا ۵ درصد از کوچکترین داده‌ها حذف و از بقیه داده‌ها میانگین حسابی گرفته شده و مبنای قضاوت یا تصمیم‌گیری در مورد جمعیت قرار می‌گیرد. چنانچه میانگین حسابی و میانگین پیراسته تفاوت چندانی نداشته باشند نشان‌دهنده این است که داده‌های پرت یا افراطی (داده‌های خیلی بزرگ یا خیلی کوچک) در مجموعه داده‌ها وجود نداشته است.

تمرین: میانگین حسابی و میانگین پیراسته مقادیر زیر را حساب کنید.

۱۲۶ ، ۱۱۸ ، ۱۲۸ ، ۱۰۶ ، ۹۴ ، ۱۱۶ ، ۴۵۲ ، ۱۲۵ ، ۱۴۲ ، ۱۰۵ ، ۱۴۶

ص ۵

- میانگین هندسی - Geometric Mean $\bar{x}_G$ یا μ_G

میانگین هندسی معمولاً برای محاسبه میانگین اندازه‌های نسبی مثل درصد، نسبت، نرخ رشد، نرخ تورم، درصد خطاهایی که براساس نسبت دو مقدار بدست آمده‌اند، بکار برده می‌شود.

- در حالت ساده:

$$\bar{x}_G = (x_1 \times x_2 \times x_3 \times \cdots \times x_n)^{\frac{1}{n}}$$

- در حالت داده‌های طبقه‌بندی شده

x_i	x_1	x_2	x_3	$\cdots$	x_k
f_i	f_1	f_2	f_3	$\cdots$	f_k

میانگین هندسی موزون:

$$\bar{x}_G = \left(x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \times \cdots \times x_k^{f_k}\right)^{\frac{1}{\sum_{i=1}^{k} f_i}}$$

$$= \left(x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \times \cdots \times x_k^{f_k}\right)^{1/n} \qquad \sum_{i=1}^{k} f_i = n$$

تعداد کل نمونه

$$= \left(\prod_{i=1}^{k} x_i^{f_i}\right)^{\frac{1}{n}}$$

نماد: $\prod_{i=1}^{k} x_i = x_1 \times x_2 \times \cdots \times x_k$

در محاسبه میانگین هندسی بهتر است از خاصیت لگاریتم استفاده شود. خواص مثل:

$$\log x_i^{f_i} = f_i \log x_i \quad \text{یا} \quad \log(x_1 \times x_2 \times \cdots \times x_n)^{\frac{1}{n}} = \frac{1}{n}\log(x_1 \times x_2 \times \cdots \times x_k)$$

$$= \frac{1}{n}\left[\log x_1 + \log x_2 + \cdots + \log x_k\right]$$

تمرین: نسبت سود شرکتی دولتی در سالهای متوالی نسبت به سال ماقبل به شرح زیر است. متوسط سود سالانه این شرکت چقدر است.

نسبت سود سال ۸۱ به ۸۰ برابر با ۳
" " ۸۲ به ۸۱ " " ۴
" " ۸۳ به ۸۲ " " ۱,۵
نسبت سود سال ۸۴ به ۸۳ ۴,۲
نسبت سود سال ۸۵ به ۸۴ ۴,۱

تمرین- میانگین هندسی را برای داده‌های زیر بدست آورید

x_i	۸۸	۸۰	۶۵	۵۵
f_i	۶	۵	۷	۲

- میانگین هارمونیک Harmonic Mean μ_H یا $\bar{x}_H$

از این نوع میانگین برای محاسبه میانگین مقیاسهای ترکیبی مثل کیلومتر در ساعت، دور در ثانیه، نفر ساعت و متغیرهای شبیه به آن استفاده می‌شود

میانگین هارمونیک برابر است با عکس میانگین حسابی عکس داده‌ها

- حالت ساده:

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

- حالت موزون

$$\bar{x}_H = \frac{\sum_{i=1}^{k} f_i}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_k}{x_k}} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

تمرین: مسافت ۴۰۰ کیلومتری بین دو شهر A، B را در رفت با سرعت ۹۰ کیلومتر
① در ساعت طی کرده‌ایم و در برگشت با سرعت ۱۱۰ کیلومتر در ساعت، سرعت متوسط حقیقی چقدر است؟

تمرین ②

x_i	۵	۶	۱۰	۱۵
f_i	۶	۱۰	۸	۶

- میانه Median نماد: Md

میانه عددی است که پس از مرتب کردن داده‌ها (به ترتیب صعودی یا نزولی) دقیقاً ۵۰ درصد داده‌ها از آن بزرگتر و ۵۰ درصد داده‌ها از آن کوچکتر هستند. اگر تعداد داده‌ها فرد باشد، میانه پس از مرتب کردن داده‌ها $\frac{n+1}{2}$ امین عدد است و اگر تعداد داده‌ها زوج باشد، میانه عبارت است از میانگین دو عدد وسطی یعنی $\frac{n}{2}$ امین و $\frac{n+2}{2}$ امین عدد پس از مرتب کردن داده‌ها.

- خواص میانه: - میانه یک شاخص ساده و کم‌تغییر است

- مقادیر خیلی بزرگ یا خیلی کوچک (برعکس میانگین حسابی) در میانه تأثیری ندارند.

- چون در محاسبه میانه همه مقادیر داده‌ها نقش نمی‌کنند، میانه معمولاً کمتر شاخص قوی و مناسبی برای بیان مرکز ثقل داده‌ها است، مگر اینکه میانگین به دلایلی قابل استفاده نباشد.

- مهمترین خاصیت میانه این است که مجموع قدر مطلق انحرافات از میانه مینیمم است. یعنی

برای هر $a \neq Md$

$$\sum_{i=1}^{n} |x_i - M| < \sum |x_i - a|$$

یا

در حالت موزون:

$$\sum_{i=1}^{k} f_i |x_i - Md| < \sum_{i=1}^{k} f_i |x_i - a|$$

از این خاصیت معمولاً برای تعیین محل احداث انبارهای عمومی، ایستگاه‌های قطار یا مترو یا اتوبوس، پمپ بنزین، سیستم‌های فشار آب، سیلوی ذخیره غلات و غیره استفاده می‌شود.

تمرین:
فاصله دو شهر A و B ۱۰۰ کیلومتر است. در این فاصله ۹ مرکز تولید گندم وجود دارد که میزان تولیدات این مراکز متفاوت است. بهترین نقطه برای احداث سیلو یا انبار کجاست.

کیلومتر	A	۵	۱۰	۲۰	۳۰	۵۰	۶۰	۷۵	۸۵	۹۰	B
میزان تولید برحسب ۱۰۰ تن		۱۰	۵	۴۰	۶۰	۳۰	۲۰	۱۵	۱۰	۵	

- محاسبه میانه در حالتی که داده‌ها طبقه‌بندی شده‌اند

L_i : کرانه پایینی مربوط به هر طبقه

U_i : کرانه بالایی مربوط به هر طبقه

$L_i - U_i$	فراوانی f_i	فراوانی تجمعی F_i
$L_1 - U_1$	f_1	f_1
$L_2 - U_2$	f_2	$f_1 + f_2$
$L_3 - U_3$	f_3	$f_1 + f_2 + f_3$
$\vdots$	$\vdots$	
$L_k - U_k$	f_k	$\sum_{i=1}^{k} f_i$

$$Md = L_i + \frac{\frac{n}{2} - F_i}{F_{i+1} - F_i}(U_i - L_i)$$

که در این فرمول :

n : تعداد نمونه

L_i : کرانه پایینی طبقه‌ای که میانه در آن قرار دارد

F_i : فراوانی تجمعی طبقه‌ای که میانه در آن قرار دارد

F_{i+1} : فراوانی تجمعی طبقه بعد از طبقه‌ای که میانه در آن قرار دارد.

$U_i - L_i$: دامنه طبقه‌ای که میانه در آن قرار دارد.

- نما یا مُد Mode

نما یا مُد عبارتست از مقداری که بیشترین فراوانی را دارا می‌باشد

خواص نما : نما یک شاخص ساده است اما برخلاف میانگین و میانه یکتا نیست

به عبارت دیگر یک مجموعه ممکن است بیشتر از یک نما داشته باشد

- معمولاً یا مُد تحت تاثیر مقادیر خیلی بزرگ یا خیلی کوچک قرار نمی‌گیرد

تمرین :- نمای مقادیر زیر را حساب کنید

۷, ۳, ۵, ۶, ۷, ۹, ۱۰, ۱۴, ۱۳, ۹, ۹, ۷, ۶, ۱, ۳

- میانگین حسابی و میانه را برای داده‌های فوق حساب کنید

- این سه شاخص با هم چه تفاوتی دارند؟ هرکدام بیشتر در کجا کاربرد بیشتری دارند؟

- صدک Percentile ، دهک Decile ، چارک Quartile

برای تعیین صدک ، دهک یا چارک یک مجموعه عددی ابتدا باید داده‌ها به صورت صعودی یا نزولی مرتب شوند

صدک P ام عبارت است از مقداری که (پس از مرتب شدن داده‌ها) P درصد مقادیر از آن مقدار کمتر و بقیه از آن مقدار بیشتر هستند. مثلاً صدک بیستم عبارت است از عددی که بیست درصد از داده‌ها از آن کمتر و ۸۰ درصد داده‌ها از آن بیشتر هستند.

- دهک اول برابر است با صدک دهم ، دهک دوم برابر است با صدک بیستم و ...
- چارک اول برابر است با صدک بیست و پنجم . چارک سوم برابر است با صدک هفتاد و پنجم و چارک دوم همان میانه است برابر است با صدک پنجاهم یا دهک پنجم
- برای محاسبه این شاخص‌ها می‌توان از یک روش یکسان یعنی محاسبه صدک‌ها به شرح زیر استفاده کرد.
- برای محاسبه صدک P ام از روش زیر بهره می‌گیریم

$$i = \frac{P(n+1)}{100}$$

که در آن n تعداد داده‌هاست.

- اگر i عدد صحیح بود ، x_i امین عدد (پس از مرتب کردن داده‌ها به صورت صعودی) برابر است با صدک P ام
- اگر i عدد صحیح نبود ، جزء صحیح را با i و جزء اعشار را با w نشان می‌دهیم . در این حالت صدک P ام برابر است با $(1-w)x_i + w\,x_{i+1}$

مثلاً اگر $n=30$ باشد . صدک بیستم برابر است با

$$i = \frac{20\,(30+1)}{100} = 6{,}2$$

$$(1-0/2)\,x_6 + 0/2\,x_7$$