• پارامترها و آماره‌های مهم:

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

• طبقه‌بندی صفات

از آن‌جا که اطلاعات آماری به صورت اعداد و ارقام بیان می‌شوند، اگر بتوان آن‌ها را به صورت طبقه‌بندی شده بیان کرد، به راحتی می‌توان به خصوصیات مهم آن‌ها پی برد، این داده‌ها به دو دسته پیوسته و ناپیوسته تقسیم می‌شوند.

ـ به اعدادی که طبقات یک جدول توزیع فراوانی را مشخص می‌سازند، حدود طبقات می‌گویند.

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

ـ طول طبقه یا دسته تفاوت بین حدود بالا یا پائین دو طبقه متوالی است.

مثال ۱: به جدول اعداد طبقه‌بندی شده (پیوسته) زیر توجه کنید:

طبقات	$0-5$	$5-10$	$10-15$
فراوانی	3	4	13

به طور مثال در طبقه اول *5* حد *بالا* و *0* 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 و 5 طول طبقات است، چرا که **5**=0-5 ، **5**=5-10 یا **5**=10-15 می‌باشد.

رابطه انواع میانگین‌ها

نکته: همیشه $\overline{X} > \overline{X}_G > \overline{X}_H$ است و فقط زمانی که داده‌ها با یکدیگر برابر باشند $\overline{X} = \overline{X}_G = \overline{X}_H$ خواهد بود.

مثال : کدام‌یک از روابط زیر بین میانگین حسابی $(\bar{x})$ ، میانگین هندسی $(\bar{x}_G)$ و میانگین هارمونیک $(\bar{x}_H)$ برقرار است؟

(اقتصاد ۷۴)

۱) $\overline{X}_H < \overline{X}_G < \overline{X}$ ۲) $\overline{X}_G < \overline{X} < \overline{X}_H$ ۳) $\overline{X}_G < \overline{X}_H < \overline{X}$ ۴) $\overline{X} < \overline{X}_G < \overline{X}_H$

حل : با توجه به تعاریف بالا گزینه (۱) صحیح می‌باشد.

مثال ۲: میانگین داده‌های زیر چیست؟

x_i = داده	2	1	3
F_i فراوانی مطلق	5	4	2

حل : محاسبهٔ میانگین برای داده‌های نیمه طبقه‌بندی شده

$$\overline{X} = \frac{3\times2+1\times4+2\times5}{2+4+5} = \frac{20}{11}$$

● خواص میانگین حسابی

۱) مجموع انحرافات از میانگین همیشه صفر است. $\boxed{\sum(x_i-\mu)=0}$

۲) مجموع مجذور انحرافات از میانگین همیشه می‌نیمم است. a عدد دلخواه است ; $\boxed{\sum\ (x_i-\mu)^2<\sum\ (x_i-a)^2}$

۴) در جامعه آماری فقط یک میانگین داریم.

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

مثال ۱: اگر کمیت‌های $x_1, x_2, ..., x_n$ با حجم n به دست آمده باشند، کدام یک از روابط زیر صادق است؟ (اقتصاد ـ ۷۱)

۱) $\sum X_i = n\overline{X}^2$ ۲) $\sum(X_i-\overline{X})=0$ ۳) $\sum(X_i-\overline{X})^2=0$ ۴) $\sum(X_i-me)=0$

حل : گزینه ۲ صحیح می‌باشد.

مجموع انحرافات از میانگین همواره صفر است. به عبارت دیگر $\sum(x_i-\overline{x})=0$ می‌باشد.

۲ـ میانگین هندسی (μ_G) :

اگر داده‌های بدست آمده نسبت،درصد، شاخص نرخ رشد و ... باشد، برای بدست آوردن مقدار متوسط از میانگین هندسی استفاده می‌کنیم.

۳ـ میانگین هارمونیک (توافقی یا معکوس یا همساز، μ_H)

اگر مقیاس داده‌ها به صورت ترکیبی باشد از این میانگین استفاده می‌کنیم. مانند: متر در ثانیه، کیلومتر بر ساعت و....

مثال ۲: اگر 3 اتومبیل مسیر 60کیلومتری بین دو منطقه را به ترتیب با سرعت 120 و 60 و 90کیلومتر در ساعت طی نمایند. میانگین سرعت این سه اتومبیل برابر با چند کیلومتر در ساعت است؟ (مدیریت ۷۹)

۱) تقریباً 83 ۲) تقریباً 86 ۳) تقریباً 90 ۴) 90

حل : گزینه ۱ صحیح می‌باشد.

$$\overline{X}_H = \frac{n}{\frac{1}{x_1}+....+\frac{1}{x_n}} = \frac{3}{\frac{1}{120}+\frac{1}{60}+\frac{1}{90}} = 83.076$$

مثال۳: در یک کارگاه 5 ماشین با سرعت 4 دور در ثانیه و 3 ماشین با سرعت 6 دور در ثانیه کار می‌کنند. سرعت متوسط این ماشین‌ها چند دور در ثانیه است؟ (حسابداری و مدیریت ۸۵)

۱) 4.85 ۲) 4.75 ۳) 4.63 ۴) 4.57

حل : گزینه ۴ صحیح می‌باشد.

با توجه به واحد ترکیبی (دور در ثانیه) از میانگین هارمونیک استفاده می‌کنیم:

$$\bar{x}_H = \frac{5+3}{\frac{5}{4}+\frac{3}{6}} = \frac{8}{\frac{21}{12}} = \frac{96}{21} = 4.57$$

میانه داده‌های زیر را محاسبه کنید.

حد پایین طبقه میانه‌دار

C − L	10-20	20-30	30-40	40-50	جمع
F_i	10	20	30	40	$N = \sum F_i = 100$
F_{C_i}	10	30	60	100	

فراوانی مطلق طبقه میانه‌دار

فراوانی تجمعی طبقه ماقبل

جواب: ۳۶/۶۶

خواص میانه

50 درصد داده‌ها قبل و 50 درصد داده‌ها بعد از میانه قرار دارند، از این رو بزرگ یا کوچک بودن متغیرها تأثیری بر مقدار میانه نخواهد داشت.

مثال۲: در صورتی که به بزرگ‌ترین عدد یک سری داده مقدار ثابتی اضافه گردد، این افزایش بر کدام معیار تأثیر نمی‌گذارد؟

ا)

۱) ضریب پراکندگی ۲) میانه ۳) میانگین ۴) واریانس

حل : گزینه ۲ صحیح می‌باشد.

مثال ۱: نظر گروهی از سوادآموزان راجع به زمان بخش برنامهٔ نهضت سوادآموزی از سیمای جمهوری اسلامی جمع‌آوری شده است. کدام شاخص مرکزی برای آن داده‌ها مناسب‌تر است؟ (اقتصاد ۷۳)

۱) میانگین　　۲) میانه　　۳) نما　　۴) چارک اول

حل : گزینه ۳ صحیح می‌باشد.

با توجه به آن‌که بیشترین فراوانی سنجیده می‌شود، مد (نما) مناسب‌ترین شاخص مرکزی برای داده‌هاست.

○ تفاوت‌های اساسی بین میانگین، میانه و مد

۱) میانگین برحسب مقیاس داده‌ها است و در محاسبه آن فراوانی و کمیت داده، در نظر گرفته می‌شود، اما میانه و مد تابع ترتیب و فراوانی داده‌ها هستند.

۲) میانگین از ترکیب داده‌ها حاصل نشده و هر افزایش یا کاهش داده‌ها مقدار میانگین را عوض می‌کند، اما اگر افزایش یا کاهش ترتیب داده‌ها را عوض نکند در مقدار مد و میانه تأثیری ندارد.

فرمول‌های لازم برای پیدا کردن محل چارک‌ها، دهک‌ها، صدک‌ها

۱ـ محاسبهٔ چندک‌ها برای داده‌های طبقه‌بندی نشده :

الف) ابتدا داده‌ها را به ترتیب صعودی مرتب می‌کنیم.

ب) سپس با توجه به نوع چندک محل آن را با استفاده از:

Q= چارک: $\frac{aN}{4}+\frac{1}{2}$; $a=1,2,3$

D= دهک: $\frac{aN}{10}+\frac{1}{2}$; $a=1,2,\ldots,9$

P= صدک: $\frac{aN}{100}+\frac{1}{2}$; $a=1,2,\ldots,99$

پیدا می‌کنیم.

۲ـ محاسبه چندک‌ها برای داده‌های طبقه‌بندی شده:

الف) ابتدا از روی جدول، فراوانی تجمعی را محاسبه می‌کنیم.

ب) با استفاده از $\frac{aN}{4}\ (a=1,2,3)$ یا $\frac{aN}{10}\ (1,2,\ldots,9)$ یا $\frac{aN}{100}\ (1,2,\ldots,99)$ اولین طبقه‌ای که فراوانی تجمعی‌اش بیشتر یا مساوی یکی از مقادیر فوق باشد را با توجه به چارک، دهک یا صدک پیدا می‌کنیم.

ج) با استفاده از فرمول زیر آن را محاسبه می‌نماییم:

$$\text{چارک} = \text{مقدار چندک} = \text{حد پائین طبقه چارک‌دار} + \frac{\left(\frac{aN}{4} - \text{فراوانی تجمعی طبقه ماقبل}\right)}{\text{فراوانی مطلق طبقه چارک‌دار}} \times \text{طول طبقه}$$

$$\text{دهک} = \text{مقدار چندک} = \text{حد پائین طبقه دهک‌دار} + \frac{\left(\frac{aN}{10} - \text{فراوانی تجمعی طبقه ماقبل}\right)}{\text{فراوانی مطلق طبقه دهک‌دار}} \times \text{طول طبقه}$$

$$\text{صدک} = \text{مقدار چندک} = \text{حد پائین طبقه صدک‌دار} + \frac{\left(\frac{aN}{100} - \text{فراوانی تجمعی طبقه ماقبل}\right)}{\text{فراوانی مطلق طبقه صدک‌دار}} \times \text{طول طبقه}$$

نکته :دهک پنجم = چارک دوم = صدک پنجاهم = میانه است.

مثال ۱: مطلوبست دهک دوم جدول زیر : (مدیریت ۸۰)

C−L	40-50	50-60	60-70
F_i	5	18	7

۱) 48.2 ۲) 51.5

۳) 50.55 ۴) 62.38

حل : گزینه ۳ صحیح می‌باشد.

ابتدا باید فراوانی تجمعی جدول فوق را محاسبه نمائیم:

حد پائین طبقه دهک‌دار

C − L	40 − 50	50 − 60	60 − 70	
F_i	5	18	7	$N=\Sigma F_i=30$
F_{c_i}	5	23	30	

فراوانی مطلق طبقه دهک‌دار

فراوانی تجمعی طبقه ماقبل

حل :

محل دهک دوم : $\frac{aN}{10}=\frac{2\times 30}{10}=6 \rightarrow$ طبقه $60-50$ ، محل دهک دوم می‌باشد.

مقدار دهک دوم: $50+\frac{(6-5)}{18}\times 10 = 50+\frac{10}{18}=50.55$

فرمول‌های محاسبه واریانس

۲) فرمول محاسبه:

محاسبهٔ واریانس برای داده‌های طبقه‌بندی نشده : $V(x) = \sigma_x^2 = \frac{\sum (x_i - \mu_x)^2}{N} = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N}\right)^2 = \frac{\sum x_i^2}{N} - \mu_x^2$

محاسبه واریانس برای داده‌های طبقه‌بندی شده : $V(x) = \sigma_x^2 = \frac{\sum F_i (x_i - \mu_x)^2}{N} = \frac{\sum F_i x_i^2}{N} - \left(\frac{\sum F_i x_i}{N}\right)^2 = \frac{\sum F_i x_i^2}{N} - \mu_x^2$

$$V(x) = \sigma_x^2 = \sum f_i \ (x_i - \mu_x)^2 = \sum f_i \ x_i^2 - \left(\sum f_i \ x_i\right)^2 = \sum f_i x_i^2 - \mu_x^2$$

مثال ۴: واریانس داده‌ها با جدول فراوانی زیر کدام است؟(حسابداری ۷۷)

x_i	-1	0	1	2
F_i	2	3	4	1

۱) 0.76 ۲) 0.78 ۳) 0.82 ۴) 0.84

حل : گزینه ۴ صحیح می‌باشد.

$$\sigma^2 = \frac{\sum F_i x_i^2}{N} - \left(\frac{\sum F_i x_i}{N}\right)^2 = \frac{2(-1)^2 + 3(0)^2 + 4(1)^2 + 1(2)^2}{2+3+4+1} - \left(\frac{2(-1)+3(0)+4(1)+1(2)}{2+3+4+1}\right)^2 = \frac{10}{10} - \left(\frac{4}{10}\right)^2 = \frac{10}{10} - \frac{16}{100} = 0.84$$

مثال ۷: جدول مقابل توزیع فراوانی فروش یک شرکت را نشان می‌دهد. میانگین و انحراف معیار فروش به ترتیب چقدر است؟

(اقتصاد ۷۰)

تعداد روزها	فروش به هزار تومان
10	20 تا کم‌تر از 30
25	30 تا کم‌تر از 40
15	40 تا کم‌تر از 50

۱) 35 ، 8.6
۲) 36 ، 5.7
۳) 36 ، 7
۴) 35 ، 9

حل : گزینه ۳ صحیح می‌باشد.

C-L	x_i	F_i	$F_i x_i$	$F_i x_i^2$
20-30	25	10	250	6250
30-40	35	25	875	30625
40-50	45	15	675	30375
جمع			1800	67250

(مرکز طبقات) $x_i = \frac{\text{حد بالا + حد پایین}}{2}$

$$\bar{x} = \frac{\sum F_i x_i}{N} = \frac{(10\times25)+(25\times35)+(15\times45)}{10+25+15} = \frac{250+875+675}{50} = \frac{1800}{50} = 36$$

$$var(x) = \sigma_x^2 = \frac{\sum F_i x_i^2}{N} - \left(\frac{\sum F_i x_i}{N}\right)^2 = \frac{67250}{50} - \left(\frac{1800}{5}\right)^2 = 1345 - 1296 = 49$$

$$\sigma_x^2 = 49 \rightarrow \sigma_x = 7$$

مثال : دستگاه A در اندازه‌گیری مکرر از شیء واحدی دارای واریانس $\sigma^2 = 9$ بوده است. دستگاه B در اندازه‌گیری مکرر از همان شیء دارای واریانس $\sigma^2 = 25$ بوده است؟ (مدیریت ۷۴)

۱) دستگاه A دقیق‌تر است.

۲) دستگاه B دقیق‌تر است.

۳) دستگاه A اندازه‌گیری‌های بزرگ‌تری از دستگاه B به دست می‌دهد.

۴) دستگاه B اندازه‌گیری‌های بزرگ‌تری از دستگاه A به دست می‌دهد.

حل : گزینه ۱ صحیح می‌باشد.

واریانس برای مقایسه دوجامعه وقتی به کار می‌رود که:

1) $\mu_1 = \mu_2$

و

2) واحد اندازه‌گیری دوجامعه یکسان باشد.

با حفظ دو شرط بالا وقتی واریانس (انحراف معیار) جامعه‌ای کمتر است، پراکندگی جامعه کمتر، خطا کمتر، دقت بیشتر و کارایی بیشتر است.

در این سوال دو شرط فوق برقرار بوده در نتیجه دستگاه A که دارای واریانس کمتری است دقت بیشتری دارد.

مثال ۲: میانگین 20 داده آماری 15 و واریانس آن‌ها برابر 2.25 است. درصد ضریب تغییرات آن‌ها چقدر است؟(حسابداری ۷۷)

۱) 10 ۲) 12 ۳) 15 ۴) 20

حل : گزینه ۱ صحیح می‌باشد.

$$CV = \frac{\sigma_x}{\mu_x} = \frac{\sqrt{\sigma^2{}_x}}{\mu_x} = \frac{\sqrt{2.25}}{15} = \frac{1.5}{15} = 0.1$$

$$\text{درصد ضریب تغییرات} = CV \times 100 = 0.1 \times 100 = 10$$

مثال۵: میانگین سن یک گروه 12 سال و ضریب تغییرات سن آنان 20 درصد است. انحراف معیار سن آنان چقدر است؟ (مدیریت ۷۹)

۱) 0.6 ۲) 2.4 ۳) 60 ۴) 240

حل : گزینه ۲ صحیح می‌باشد.

$$CV = \frac{\sigma_x}{\mu_x} \to \sigma_x = CV \times \mu_x = \frac{20}{100} \times 12 = 2.4$$

مثال۷: متوسط درآمد ماهانه کارگران کارخانه A ، 17 هزار تومان با واریانس 4 می‌باشد. در کارخانه B متوسط درآمد ماهانه 250 هزار ریال با واریانس 900 می‌باشد. (مدیریت ۷۴)

۱) اختلاف درآمد در کارخانه A بیش از کارخانه B است.

۲) اختلاف درآمد در کارخانه B بیش از کارخانه A است.

۳) درآمدهای اکثر افراد کارخانه A کم‌تر از اکثر افراد کارخانه B است.

۴) کم‌ترین درآمد در کارخانه A بیش از کارخانه B است.

حل : گزینه ۲ صحیح می‌باشد.

واحدهای اندازه‌گیری و میانگین‌ها برابر نیستند بنابراین برای مقایسه بین دو جامعه از ضریب تغییرات استفاده می‌کنیم.

$$CV_A = \frac{\sigma_A}{\mu_A} = \frac{2}{17} \times 100 = 11.76$$

$$CV_B = \frac{\sigma_B}{\mu_B} = \frac{30}{250} \times 100 = 12$$

و چون $CV_A < CV_B$ است یعنی پراکندگی در کارخانه B بیشتر ا ست و اختلاف درآمد در کارخانه B بیش از کارخانه A است.