واریانس داده‌ها $S^2/SD^2/V$

دقیق‌ترین، مفیدترین و باثبات‌ترین شاخص پراکندگی است.

در مطالعات آماری و در تحقیقات کاربرد زیادی دارد.

آن را می توان با میانگین که با ثبات‌ترین شاخص مرکزی است، مقایسه کرد.

یکی از معیارهای بسیار مناسب در بیان میزان پراکندگی داده‌ها است. زیرا بر خلاف بعضي از شاخص‌ها، از همه داده‌ها برای محاسبه آن استفاده می‌شود.

متوسط مجذور انحراف‌ها از میانگین را واریانس یا پراش می‌گویند.

جذر واریانس می شود انجراف معیار یا ریشه / توان دوم واریانس را انحراف معیار یا استاندارد داده ها گویند و در مورد جامعه آن را با نماد σ^2 و در مورد نمونه آن را با $S^2/SD^2/V$ نشان می‌دهیم.

واریانس داده‌ها

- **اگر انحراف معیار مجموعه داده‌ها عدد کوچکی باشد، بدین معناست که پراکندگی داده‌ها حول میانگین‌شان کم و در نتیجه داده‌ها به هم نزدیک‌تر است**
و اگر انحراف معیار مجموعه داده‌ها عددی بزرگ باشد، بدین معناست که پراکندگی داده‌ها حول میانگین‌شان زیاد و درنتیجه داده‌ها از هم دورتر است.
- **محاسبه واریانس (پراش) نشان دهنده پراکندگی داده‌ها در اطراف میانگین است. عیب انحراف متوسط را بر طرف می‌کند. قدر مطلق را بر می داریم وبه توان ۲ می رسانیم.**
- **شاخص‌هایی که از مطالعه نمونه به دست می‌آید را آماره و شاخص‌هایی که از مطالعه جامعه به دست می‌آید را پارامتر می‌نامند.**

محاسبه واریانس وانحراف استاندارد (معیار)در داده‌های طبقه بندی نشده

مثال: از یک جامعه آماری اندازه‌های زیر به عنوان نمونه انتخاب شده است، واریانس و انحراف معیار داده‌ها را محاسبه کنید. **2-4-8-12-4**

xi	$X-\bar{X}$	$(X-\bar{X})^2$
۲	2-6=-4	16
۴	-2	4
۸	+2	4
۱۲	+6	36
۴	-2	4
۳۰		64

$$S^2 = \frac{\sum \left(Xi - \overline{X} \right)^2}{n-1}$$

$$S^2 = \frac{64}{5-1} = 16$$

$$S = \sqrt{16} = 4$$

تمرین منزل

برای داده‌های زیر واریانس و وانحراف استاندارد (معیار) را به دست آورید.

۳ ۲ ۴ ۳ ۶ ۳ ۵ ۶

مثال 2: از راه نمره‌های اصلی یا خام

xi	xi^2
12	144
14	196
8	64
6	36
4	16
$\sum xi = 44$	$\sum xi^2 = 456$

$$S^2 = \frac{\sum xi^2 - \frac{(\sum xi)^2}{n}}{n-1} = \frac{456 - \frac{(44)^2}{5}}{5-1} = \frac{68/8}{4} = 17/2$$

$$S = \sqrt{17/2} = 4/147$$

- درعمل محاسبه S^2 برای اینکه واریانس نمونه به ویژه وقتی که تعداد نمونه از 30 مورد کمتر است به واریانس جامعه نزدیک باشد به جای n از n-1 که آنرا درجه آزادی نیز می‌نامند، استفاده می‌کنند واین زمانی است که بخواهیم از روی شاخص آماری نمونه، پارامتر جامعه را استنباط کنیم.

مثال ۳: اگر مجموع مجذورات در یک جامعه آماری ۱۰۰ نفری ۳۵۵۴۰ و مجموع داده‌ها ۱۸۰۰ باشد، واریانس را محاسبه کنید؟

$$S^2 = \frac{\sum xi^2}{N} - (\frac{\sum xi}{N})^2 = \frac{35540}{100} - (\frac{1800}{100})^2 = 31/4$$

مثال ۴: اگر $\sum xi^2 = 220$ و $\mu = 4/5$ و $N = 10$ باشد، واریانس و انحراف معیار چقدر می شود؟

$$S^2 = \frac{\sum xi^2}{n} - \mu^2 = \frac{220}{10} - (4/5)^2 = 1/75$$

$S=\sqrt{1/75} = 1/32$

محاسبه واریانس وانحراف استاندارد(معیار)در داده‌های طبقه بندی شده یا جدول توزیع فراوانی:

۱- روش مستقیم یا از راه نمره‌های اصلی (اعداد خام)

CL	fi	Xi نماینده طبقات	fix`	fix`^{2}
3-7	4	5	20	$4(5)^2=100$
8-12	3	10	30	300
13-17	11	15	165	2475
18-22	7	20	140	2800
23-27	5	25	125	3125
C=5	N=30		$\sum FiXi=480$	$\sum FiXi^2=8800$

$$s^2 = \frac{\sum FiXi^2}{N} - \left(\frac{\sum FiXi}{N}\right)^2$$

$$= \frac{8800}{30} - \left(\frac{480}{30}\right)^2 = 37/33$$

انحراف معیار $S=\sqrt{37/33} = 6/11$

ب)روش غیر مستقیم (انحرافی): زمانی از این روش استفاده می‌کنیم که میانگین داده‌ها عدد صحیح باشد.

CL	fi	x'	Fi x'	Fi x'^2	$(x'-\mu)^2$	Fi$(x'-\mu)^2$
3-7	4	5	20	4(5) ²=100	$(5-16)^2 = 121$	4×121=484
8-12	3	10	30	300	36	108
13-17	11	15	165	2475	1	11
18-22	7	20	140	2800	16	112
23-27	5	25	125	3125	81	405
C=5	N=30		480	8800		$\sum$Fi$(x'-\mu)^2 = 1120$

$$\bar{X} = \frac{\sum Fix'}{N} = \frac{480}{30} = 16$$

$$S^2 = \frac{\sum Fi(x'-\mu)^2}{N}$$

$$= \frac{1120}{30} = 37/33$$

نکات کنکوری

واریانس 5عدد متوالی همیشه مساوی 2 است و از یک نمونه آماری می‌شود

مثال: 5-4-3-2-1 899-898-897-896-895

واریانس 5عدد غیرمتوالی همیشه مساوی 8 است. انحراف متوسط 1.41

اگر در سوال اشاره نشود که داده‌ها از جامعه انتخاب شده است یا نمونه، پیش فرض مسئله جامعه است.

واریانس عبارت است از میانگین نمره‌ها یا مجموع مجذورات انحراف نمره‌ها از میانگین تقسیم بر تعداد نمره‌ها.

تصحیح انحراف استاندارد یا شپرد sheppards Correction

✓خطای طبقه بندی یا گروه بندی Error of grouping در محاسبه انحراف استاندارد به دلیل آنکه انحراف‌ها مجذور می‌شوند، بیشتر می‌شود.

✓این خطا بیشتر تابع فاصله طبقاتی است. هر چه فاصله طبقاتی بزرگتر شود، این خطا بزرگ‌تر است.

✓تصحیح شپرد زمانی به کار برده می‌شود که تعداد طبقات کمتر از ۱۲ باشد، زیرا در غیر اینصورت مقدار خطا ناچیز خواهد بود.

$$\checkmark Sc = \sqrt{S^2 - \frac{C^2}{12}}$$

✓مثال: در یک توزیع فراوانی طبقه بندی شده انحراف استاندارد و فاصله طبقات به ترتیب ۴ و ۳ شده است. انحراف استاندارد اصلاح شده چقدر می‌شود؟

$$Sc = \sqrt{16 - \frac{9}{12}} = 3/9\checkmark$$

فرمول های واریانس و انحراف معیار

$$S^2 = \frac{\sum(Xi-\overline{X})^2}{n-1}$$

$$S^2 = \frac{Xi^2 - \frac{(\sum xi)^2}{n}}{n-1}$$

$$S^2 = \frac{nXi^2-(\sum xi)^2}{n(n-1)}$$

$$S^2 = \frac{\sum xi^2}{N} - \left(\frac{\sum xi}{N}\right)^2$$

$$s^2 = \frac{\sum xi^2}{n} - \mu^2$$

$$s^2 = \frac{\sum Fix'^2}{N} - \left(\frac{\sum Fix'}{N}\right)^2$$

$$S^2 = \frac{\sum Fi(x'-\mu)^2}{N}$$

محاسبه انحراف استاندارد مرکب در فرمول مک نمار Mc Nemoar

$$S_T = \sqrt{\frac{n_A\,(\bar{X}_A^2 + S_A^2) + n_B\,(\bar{X}_B^2 + S_B^2)}{n_A + n_B} - \bar{X}_T^2}$$

- مثال: جدول زیر توزیع نمرات دو گروه از دانشجویان (گروه آزمایشی و گروه کنترل) در یک صفت معین را نشان می دهد. انحراف استاندارد یا معیار مرکب را محاسبه کنید؟

x_A	x_B	X_A^2	X_B^2
2	1	4	1
4	5	16	25
6	7	36	49
8	5	64	25
20	18	120	100
$\bar{X}_A = 5$	$\bar{X}_B = 4/5$		

$$\bar{X}_T = \frac{5 + 4/5}{2} = 4/75$$

$$S_A^2 = \frac{4 \times 120 - (20)^2}{4(4-1)} = \frac{8}{12} = 6/67$$

$$S_B^2 = \frac{4 \times 100 - (18)^2}{4(4-1)} = \frac{76}{12} = 6/33$$

$$S_T = \sqrt{\frac{n_A(\bar{X}_A^2 + S_A^2) + n_B(\bar{X}_B^2 + S_B^2)}{n_A + n_B} - \bar{X}_T^2}$$

$$S_T = \sqrt{\frac{4(25 + 6/67) + 4(20/25 + 6/33)}{4 + 4} - (4/75)^2}$$

$$S_T = \sqrt{6/565} = 2/56$$

ضریب پراکندگی یا ضریب نسبی واریانس: Coefficient of Variation

عاملی است که پراکندگی نسبی یک صفت را نشان می‌دهد معمولاً برای مقایسه بین ویژگی‌های مختلف یک گروه به کار می‌رود.

ضریب تغییرات داده‌ها که آن را با CV نشان می‌دهیم از نسبت انحراف معیار به میانگین به دست می‌آید و به صورت زیر می‌نویسیم.

$$CV = \frac{S_x}{\overline{X}} \times 100$$

$\bar{X}$میانگین مقادیر x ؛ S_xانحراف استاندارد x

مثال :اگر $\bar{X}$ و S_x ما به ترتیب ۹.۶۵ و ۲.۳۷ باشد (V)ضریب تغییرات را محاسبه کنید؟

$$CV = \frac{2/37}{9/65} \times 100 = 24/6$$

ضریب تغییرات واحد ندارد وبه همین دلیل برای مقایسه دو سری از داده‌ها که مقیاس یکسانی ندارند(واحد های اندازه گیری شان متفاوت باشد) ویا میانگین آن متفاوت باشد.

تمرین منزل

- اگر در یک نمونه، میانگین داده‌ها ۵ و واریانس داده‌ها ۴ باشد؛ ضریب تغییرات داده‌ها را به دست آورید؟

ویژگی ها و مفروضات واریانس و انحراف معیار

- اگر کلیه اندازه‌های آماری برابر هم باشند، در این صورت کلیه شاخص‌های گرایش مرکزی برابر و مساوی یکی از اندازه‌هاست و کلیه شاخص‌های پراکندگی برابر صفر است.
- مثال: در مورد اندازه‌های a , a, a, a, a (اگر همه داده‌ها باهم برابر باشند واریانس داده‌ها صفر است و برعکس).

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

- واریانس هرعدد ثابت صفراست.
- واریانس داده‌ها مقداری مثبت است.

$$S^2 \geq 0$$

- مقیاس حداقل فاصله‌ای است.
- اگر داده‌ها کجی یا کشیدگی نداشته باشد، یعنی توزیع بهنجار یا نزدیک به آن باشد، مناسب‌ترین شاخص پراکندگی است.
- انحراف معیار اعداد مساوی هم می‌شود صفر، انحراف معیار منفی نداریم.
- انحراف معیار اندازه‌های یک جامعه آماری همیشه کوچکتر از انحراف معیار اندازه‌های نمونه از همان جامعه آماری است، یعنی $\delta < S$

- **جمع وتفریق اثری در واریانس جدید ندارد**

$$Y_i = a \pm X_i \rightarrow S_Y^{۲} = S_x^{۲}$$

- **به توزیع زیر عدد ثابت ۵ را اضافه می کنیم.**
- **۱،۴،۷،۱۰،۱۳**
- **۶،۹،۱۲،۱۵،۱۸**
- **میانگین نمره‌های اصلی ۷ و میانگین نمره های جدید ۱۲=۵+۷**
- انحراف نمره‌ها از میانگین در هر دو توزیع مساوی است (6+، 3+، 0 ، 2- ، 6-)
- در نتیجه انحراف استاندارد دو توزیع نیز مساوی و برابر 4/74 است.

خواص واریانس

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

ولی اگر تک تک مشاهدات یا داده‌ها را در عدد ثابتی ضرب کنیم، انحراف استاندارد آن توزیع نیز در آن عدد ضرب می شود و واریانس نیز در مجذور آن عدد ضرب می‌شود.

فرض کنید توزیع نمره‌های زیر با میانگین ۷ در دست است. ۱، ۴، ۷، ۱۰، ۱۲

در صورتی که نمره‌های این توزیع را در عدد ثابت ۳ ضرب کنیم، انحراف استاندارد و واریانس نمره‌های جدید به ترتیب در ۳+ و ۹+ ضرب خواهند شد. برای توضیح بیشتر، توزیع نمره‌های ۱۳، ۱۰، ۷، ۴، ۱ را با میانگین ۷ و واریانس ۲۲/۵۰ و انحراف استاندارد ۴/۷۴ در نظر بگیرید، در صورتی که نمره‌های این توزیع را در عدد ۵ ضرب کنیم، توزیع زیر به دست خواهد آمد: ۶۵، ۵۰، ۳۵، ۲۰، ۵

میانگین این توزیع مساوی است با ۳۵= ۷ * ۵ . انحراف‌های نمره‌ها از میانگین برابرند با : ۳۰+، ۱۵+، ۰، ۱۵–، ۳۰– . مجذور این نمره‌ها به ترتیب عبارتند از: ۹۰۰، ۲۲۵، ۰، ۲۲۵، ۹۰۰ . مجموع این نمره‌ها یعنی $\sum x^2$ مساوی ۲۲۵۰ است. واریانس و انحراف استاندارد توزیع جدید برابر خواهد شد با :

$$S^2 = \frac{2250}{5-1} = 562/500$$

$$s = \sqrt{562/500} = 23/72$$

$$S^2C = C^2S^2$$

$$S^2C = (5^2)(22.50) = 562/5$$

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

- **اگر تک تک فراوانی ها را به یک عدد ثابت تقسیم کنیم مقدار واریانس نیز بر مجذور آن عدد تقسیم می‌شود.**

ویژگی ها و مفروضات واریانس و انحراف معیار

- انحراف معیار 5 عدد متوالی از یک جامعه آماری می‌شود $\sqrt{2}$ و از یک نمونه آماری می‌شود $\sqrt{2/5}$
- اگر میانگین چند اندازه صفر باشد انحراف معیار از رابطه $\delta = \sqrt{\frac{\sum Xi^2}{N}}$ به دست می‌آید. مثال: 2+، 1-، 1+، 2-
- اگر اندازه‌های آماری تشکیل تصاعد حسابی بدهند، می‌توان واریانس یا انحراف معیار را از رابطه زیر محاسبه کرد: $\delta^2 = \frac{d^2(n^2-1)}{12}$

مثال: واریانس اندازه‌های 11، 9، 7، 5، 3 را تعیین کنید.

d = قدرنسبت d=2

تفسیر انحراف استاندارد

- انحراف استاندارد شاخصی است **که به منظور تعیین تغییرات یا پراکندگی نمره‌ها به کار برده می‌شود.**
- چنانچه تمام نمره‌ها در اطراف میانگین انباشته یا جمع شده باشند مجموع مجذور انحراف نمره‌ها از میانگین و واریانس کوچک می‌شود و به همین ترتیب انحراف استاندارد نیز کاهش پیدا می‌کند.
- **اندازه های مرکزی هیچگونه اطلاعاتی درباره پراکندگی نمره‌ها به ما نمی‌دهند.**
- بزرگ وکوچک بودن انحراف استاندارد نسبی است. یعنی پژوهشگر انحراف استاندارد را بر اساس بزرگی وکوچکی مورد تفسیر قرار نمی‌دهد، بلکه به کمک انحراف استاندارد به تفسیر موضوع نمره‌های جمع‌آوری شده می‌پردازد.

- **وقتی انحراف استاندارد نمره‌های آزمون ریاضی کلاسی از انحراف استاندارد همین آزمون در کلاس دیگر کمتر باشد، دانش آموز کلاس اول دارای توانش یادگیری کم وبیش یکسانی نسبت به کلاس دیگری هستند. یا دارای تجانس بیشتر در این آزمون هستند.**

- استفاده از انحراف استاندارد مستلزم این است که مقیاس اندازه گیری به کار برده شده حداقل فاصله‌ای یا نسبی باشد و بر آورد معتبر و با ثباتی از پراکندگی جامعه فراهم می‌سازد.
- انحراف استاندارد باثبات ترین و پایدار ترین شاخص پراکندگی است. زیرا در محاسبه آن ارزش کمی کلیه اعداد مورد استفاده قرار می‌گیرد.