۳- انحراف متوسط از میانگین :

قبلاً بیان شد، میانگین مرکز ثقل یا مرکز داده ها

داده های آماری $x_1, x_2, x_3, \dots, x_n \rightarrow \bar{x}_n$

میانگین به تنهایی پارامتر مناسبی برای مقایسه جوامع آماری نیست.

چون تغییرات بسیار بزرگ و بسیار کوچک در آن تاثیرگذار است.

راهکار: میانگین در کنار یک پارامتر پراکندگی

ابتدا میانگین یک پارامتر مرکزی را در نظر می گیریم. $\mu = \bar{x}$

و بعد فاصله آن را با داده های آماری در دسترس می سنجیم $|x_i - \mu|$

فاصله داده آماری x_i ام از میانگین

داده های طبقه بندی نشده: انحراف متوسط از میانگین

$$d \triangleq \frac{\sum_{i=1}^{n} |x_i - \mu|}{n}$$

فراوانی مطلق

داده های طبقه بندی شده (تنظیم شده در جدول): انحراف متوسط از میانگین

$$d \triangleq \frac{\sum_{i=1}^{n} F_i |x_i - \mu|}{n}$$

نکته: $\mu - x_i$ در برخی از داده های آماری مثبت خواهد شد و در برخی از داده های آماری منفی خواهد شد

چون میانگین مرکز ثقل داده هاست

نکته: هرچه داده ها از میانگین μ دورتر باشد، (d) بزرگ تر خواهد بود.

نکته: در برخی از موارد می توانیم انحراف متوسط از میانه را محاسبه نماییم.

$$d' \triangleq \frac{\sum_{i=1}^{n} |x_i - md|}{n}$$ طبقه بندی نشده

$$d' \triangleq \frac{\sum_{i=1}^{n} F_i |x_i - md|}{n}$$ طبقه بندی شده

ویژگی های d:

① اگر تمام داده ها باهم برابر باشند، $d=0$. (اثبات بر عهده دانشجو)

② اگر به متغیرهای عدد ثابت a را اضافه یا کم کنیم d تغییر نمی کند.

③ اگر به متغیرهای آماری عدد ثابت a ضرب یا تقسیم شود انحراف متوسط در قدر مطلق $|a|$ ضرب یا تقسیم خواهد شد.

④ انحراف متوسط اعداد ثابت $d=0$

واریانس (انحراف معیار):

- یکی از شاخص های پراکندگی داده های آماری نسبت به میانگین: واریانس

- دو جامعه یا چند جامعه آماری مقایسه کنیم بهتر است از میانگین به همراه واریانس بهره ببریم.

Subject
Date

جامعه ۱ — مقایسه — جامعه ۲

$\begin{cases} \mu_1 \\ \delta_1^2 \end{cases}$ $\diamond$ $\begin{cases} \mu_2 \\ \delta_2^2 \end{cases}$

در غیر اینصورت، ممکن در تحلیل جوامع آماری به نتایج غیرمنطقی دست یابیم.

در محاسبه انحراف متوسط از میانگین که برای خنثی نمودن عبارت منفی و مثبت $(\mu - x_i)$ از قدر مطلق،

اگر برای این منظور از توان دوم استفاده کنیم به فرمول واریانس می رسیم.

نکته: ویژگی بسیار مهم واریانس

واریانس می تواند تاثیرات انحرافات بزرگ و کوچک را به راحتی نشان دهد یا

واریانس = فاصله اعداد جامعه آماری از هم

= یعنی می خواهیم ببینیم داده ها چه مقدار از هم پراکنده است.

برای داده های طبقه بندی نشده:

δ_x^2 واریانس، $\underbrace{x_1, x_2, x_3, \dots, x_{n-1}, x_n}_{\text{داده های مرتبط با جامعه آماری}}$

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

برای داده‌های طبقه بندی شده (در جدول تنظیم)

x_i	F_i	F_{ci}
x_1	مطلق	تجمعی
⋮		

$$\delta_x^2 \triangleq \frac{\sum_{i=1}^{n} F_i (x_i - \bar{x})^2}{n}$$

کولاک ایران

برای داده‌های طبقه بندی شده، δ_x^2

$$\Rightarrow \delta_x^2 = \frac{\sum_{i=1}^{n} F_i (x_i - \bar{x})^2}{n}$$

$$\Rightarrow \delta_x^2 = \frac{\sum_{i=1}^{n} F_i x_i^2}{n} - 2\left[\underbrace{\frac{\sum_{i=1}^{n} F_i (x_i)\bar{x}}{n}}_{\bar{x}}\right] + \underbrace{\frac{\sum_{i=1}^{n} F_i \bar{x}^2}{n}}_{1}$$

$$\Rightarrow \delta_x^2 = \frac{\sum_{i=1}^{n} F_i x_i^2}{n} - 2(\bar{X})(\bar{X}) + \bar{X}^2$$

$$\Rightarrow \delta_x^2 = \overline{X^2} - 2(\bar{X})^2 + (\bar{X})^2$$

$$\Rightarrow \delta_x^2 = (\overline{X^2}) - (\bar{X})^2$$

$$\Longrightarrow \delta_x^2 = (\overline{X^2}) - (\bar{X})^2$$

(میانگین تمامی داده‌ها به توان ۲
میانگین مربع دوم) − (توان دوم میانگین
داده‌ها)

Subject
Date

بیان دوم :

تعریف : $\delta_x^2 \triangleq \dfrac{\sum_{i=1}^{n} F_i (x_i - \bar{x})^2}{n}$

$$\delta_x^2 = \frac{\sum F_i x_i^2}{n} - \left(\frac{\sum F_i x_i}{n}\right)^2$$

ویژگی های واریانس :

① داده های آماری $x_1, x_2, \ldots, x_n$

اگر داده های آماری مفروض را با عدد ثابت جمع یا تفریق کنیم

$$\delta^2_{x \pm a} = V(x \pm a) = Var(x \pm a) = (\delta^2_x)$$

یعنی جمع یا تفریق یک عدد ثابت در مقدار واریانس تاثیری ندارد.

② اگر داده های آماری در عدد ثابت b ضرب شوند.

$$\delta^2_{bx} = b^2 \delta^2_x$$

در نتیجه گیری از ۱ و ۲ می گوییم.

$$\delta^2_{bx \pm a} = b^2 \delta^2_x$$

③ اگر همه داده های آماری باهم برابر باشند یا عدد ثابت باشند واریانس صفر است.

④ انحراف معیار: جذر مثبت واریانس را انحراف معیار گوییم.

$\delta_x^2 \longrightarrow \delta_x$

واریانس   انحراف معیار

ویژگی انحراف معیار:

$\delta_{x\pm a} = \delta_x$

$\delta_{bx\pm a} = b\,\delta_x$

ارتباط واریانس برای جامعه آماری با واریانس برای نمونه آماری

فرض کنیم نمونه تصادفی n تایی داشته باشیم
طبق اصول آماری داریم:

$$\delta_u^2 \triangleq \frac{\sum_{i=1}^{n} F_i (x_i-\bar{x})^2}{n-1}$$

تعریف واریانس برای یک نمونه تصادفی n تایی

$$\Rightarrow \delta_u^2 = \frac{1}{n-1}\left[\sum F_i x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] \quad (1)$$

طبق تعریف داشتیم

$$\text{جامعه}: \delta_x^2 = \frac{\sum_{i=1}^{n} x_i^2 F_i}{n} - \frac{\left(\sum x_i F_i\right)^2}{n}$$

$$n\delta^2 = \sum x_i^2 F_i - \left(\frac{\sum x_i F_i}{n}\right)^2 \quad (2)$$

از تعریف واریانس یک جامعه آماری می دانیم.

از رابطه (1) (2)

$$\Longrightarrow (n-1)\delta_u^2 = n\delta^2$$

$$\Rightarrow \delta^2 \triangleq \frac{n-1}{n}\delta_u^2$$

واریانس جامعه

واریانس نمونه تصادفی در جامعه متناظر

نتیجه دوم: اگر تعداد نمونه تصادفی اخذ شده در یک جامعه بسیار بزرگ، نزدیک به تعداد جامعه باشد.

$\lim_{n \to \infty} \frac{n-1}{n} \delta^2_u = \delta^2$ یعنی

جامعه $\delta^2_u = \delta^2$ نمونه

وقتی $n \to \infty$

نتیجه اول: پس همواره داریم:

نمونه برود به سمت جامعه نسن $\delta^2_u < \delta^2$

واریانس نمونه $<$ واریانس جامعه

مثال ساده: واریانس و انحراف معیار داده‌های ذیل را به دست آورید. ۱، ۲، ۳، ۴، ۵

حل:

$\delta^2 = \overline{(X^2)} - (\bar{X})^2$

ابتدا محاسبه میانگین $\Rightarrow \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+2+3+4+5}{5} = \boxed{3}$

$\overline{(X^2)} \Rightarrow \frac{\sum_{i=1}^{n} x_i^2}{n} \Rightarrow \frac{1^2+2^2+3^2+4^2+5^2}{5} = \boxed{11}$

$\begin{cases} \delta^2_u = 11 - 9 = \boxed{2} \\ \delta_u = \boxed{\sqrt{2}} \end{cases}$

E★

مثال: حقوق پرداختی به دانشجویان در یک طرح پژوهشی به طور متوسط ۱۵ هزار تومان با انحراف معیار ۳ هزار تومان است. اگر ۲۰٪ میانگین به حقوق دانشجویان اضافه شود میانگین و انحراف معیار حقوق پرداختی به آنها چقدر خواهد شد؟

حل:

$\begin{cases} \mu = 15000 \\ \delta_u = 3000 \end{cases}$ در شرایط اولیه

عدد ثابت $= a = \frac{20}{100}(15000) = \boxed{3000}$ $\quad 20 \times 150$

ویژگی میانگین $\Rightarrow \mu_{new} = \mu + 3000 = 15000 + 3000$

Subject

Date

اما انحراف معیار $\delta_{new} = \delta_{x+۳۰۰۰} = \delta_{x} = ۳۰۰۰$