باسمه تعالی

جزوه کابرد آمار در پژوهش :

جلسه اول:

آمار از لحاظ لغوی به معنای شمارش است، اماصرفا علم شمارش نیست. آمار عبارت است از مجموعه فعالیت هایی که از طریق آن اطلاعات را جمع آوری ، توصیف ، طبقه بندی سرانجام داوری می کنیم.

طبقه آماری آمار:

آمار توصیفی و آمار استنباطی :

آمار توصیفی: به آن بخش از علم آمارگفته می شود که هدف آن توصیف واقعیت های موود است، نه قصد رفتن به گذشته را دارد و نه قصد استقبال از آنیده را. در واقع آنجه کههست را مورد تحلیل و ارزیابی قرار می دهد.

مثال: متوسط بهره هوشی دانشجویان در ترم اری واحد علوم تحقیقات به گذشته و آینده توجه نمی شود و فقط حال حاضر بررسی می شود.

در آمار توصیفی کل افراد مدنظر مورد مطالعه قرار می گیرند و نمی توان تعداد خاصی را انتخاب کرد.

به کل افراد جامعه می گویند و منظور از تعداد نمونه است.

آمار توصیفی به دو دلیل همیشه امکان پذیر نیست:

1- صرفه جویی در منابع: در توصیفی هزینه، زمان و نیروی انسانی زیادی مصرف می شود.

2- جلوگیری از مشکلات احتمالی: مثلا برای پایش متوسط عمر لامپ های یک شرکت نمی توان همه لامپ ها را روشن کرد. شرکت با ضرر روبه رو می شود.

آمار استنباطی :

در این آمار از کل جامعه تعدادی را به عنوان نمونه انتخاب می کنیم. اقدامات محاسباتی را انجام می دهیم و نتیجه به دست آمده را به کل جامعه تعمیم می دهیم. از روی نمونه کل جامعه را استنباط می کنیم.

جامعه: گروهی از اشیاء ، افراد، حوادث و ... که حداقل دارای یک صفت یا یک ویژگی مشترک هستند.

نمونه: زیر جامعه ای است که از جامعه انتخاب می شود و مصرف آ« جامعه است.

در این شیوه از آمار امکان خطا بیشتر است.

به عددی که از نمونه بدست می آید (آمار) می گویند و آن را با حروف نشان می دهند و به عددی که از جامعه بدست می آید (پارامتر) می گویند و آن را با حروف یونانی نشان می دهند.

اگر میانگین با روش آمار بدست آید با X و اگر با روش پارامتر باشد با Y نشان می دهیم.

متغیر: در تعریف آمار پس از جمع آوری اطلاعات، صفاتی را اندازه گیری می کنیم تا اطلاعات بدست آید، به صفت مورد نظر متغیر می گویند.

اندازه گیری: عبارتست از تخصیص عدد به یک صفت (متغیر) براساس یک قاعده یا قانون از پیش تعیین شده.

مثال: یکی از صفات یک اتاق طول آن است و برای اندازه گیری از ابزاری به نام متر استفاده می کنیم.

انواع مقیاس:

1- اسمی 2- ترتیبی 3- فاصله ای 4- نسبی

از 1 تا 4 به ترتب کاملتر از قبلی می شود و ویژگی قبلی را هم دارد.

1- مقیاس اسمی : با استفاده از این مقیاس، متغیرها را مقوله بندی و طبقه بندی می کنیم. مثال: مذکر، مونت، رنگ چشم: قهوه‌ای، آبی و ...

برای نشانه گذاری مثلا به مذکر عدد یک و به مونت عدد دو را اختصاص می دهیم.

4 عمل ضرب، تقسیم، جمع، منها در این مقیاس مورد استفاده نیست.

2- مقیاس ترتیبی: علاوه بر مقوله بندی و طبقه بندی متغیر، رتبه بندی هم می کنیم. مثلا از کم به زیاد یا از کوچک به بزرگ مثل میزان تحصیلات دیپلم، لیسانس، فوق لیسانس ، دکتری و ..

4 عمل ضرب، تقسیم، جمع، منها در اینجا هم معنا ندارد.

3- مقیاس فاصله ای: علاوه بر مقوله بندی و رتبه بندی متغیر، فاصله هر متغیر با صفت و مشاهده بعدی مشخص است، این فاصله ها با هم برابر و مساوی است.

مثال: مرتب کردن نمره دانش آموزان یک کلاس: از نمره 2 تا 5، 6-8 ، 9-11، 12-14، 15-17، 18-20

4 عمل ضرب، تقسیم، جمع و منها معنا ندارد.

به صفری که در این مقیاس وجود دارد، صفر قراردادی می گویند و غیر واقعی است.

4- مقیاس نسبی : علاوه برمقوله بندی، رتبه بندی و فاصله مساوی، صفر آن واقعی و مطلق است. در علوم فیزیک بیشتر معنا دارد و در علوم انسانی کمتر.

مثل تعداد دفعات ازدواج: کسی که ازدواج نکرده است.

انواع متغیرها:

1- متغیر کمی و کیفی

متغیرهای کمی : متغیرهایی هستند که به صورت عدد نوشته می شوند.مثل سن

متغیرهای کیفی : متغیرهایی هستند که خود ما به آن عدد می دهیم مثل رنگ چشم

2- متغیرهای پیوسته و گسسته :

در متغیر گسسته بین دو مقدار آن هیچ عدد معنا ندارد و اعشار بین آن قرار نمی گیرد. مثال: تعداد خارجیان مقیم تهران یک میلیون نفر است. یک میلیون و نیم نفر ندارد.

در متغیر پیوسته بین دو مقدار هر عددی دارای معنا و مفهوم است. مثال: سن یک ساله و دو ساله و یک سال و یک ماه و ...

3- متغیر مستقل و وابسته:

متغیر مستقل توسط محقق دخل و تصرف و دستکاری می شود تا تأثیر آن روی متغیر وابسته مشخص شود.

متغیر وابسته : قابل دستکاری و دخل و تصرف نیست. میزان و مقدارآن به متغیر مستقل بستگی دارد. مثل: تأثیر سینمای خشن بر بزهکاری نوجوانان در شهر تهران: بزهکاری وابسته به سینمای خشن است.

منظور از دستکاری از سیمی توان گفت منظور من از سینمای خشن فقط فیلم های پرده سینما است ، یا منظور فقط فیلم ماهواره است.

جدول توزیع فراوانی:

مقیاس براساس متغیر تعیین می شود . به عددهای به دست آمده از طریق یک جدول به نام توزیع فراوانی نظم می دهیم.

این جدول دو ستون اصلی دارد. یک ستون توزیع X و ستون دیگر فراوانی F است.

توزیع X : مجموعه اعدادی که در اثر اندازه گیری یک ویژگی (متغیر) بدست می آید.

فراوانی F : تعداد دفعاتی که یک حادثه یا واقعه اتفاق می افتد یا نمی افتد.

F	X

جلسه دوم:

در جدول توزیع فراوانی اعداد ار از کوچک به بزرگ و در جدول از پایین به بالا می نویسیم. یعنی کوچکترین عدد در پایین قرار می گیرد.

علامت جمع زیگما : $\sum$

تعداد کل فراوانی در نمونه را با n و تعداد کل فراوانی در جامعه را با N نمایش می دهند.

X	F
20	2
9	3
8	2
7	5
6	2
5	3
4	2
3	4
	$\sum f = n = 23$

اگر اختلاف کوچکترین عدد با بزرگترین عدد بیشتر از 20 باشد، در هرطبقه چند عدد می گذاریم و تک به تک نمی نویسیم.

X	F
18-20	5
15-17	4
12-14	2

9-11	5
6-8	3
3-5	2
	$\sum f = n = 21$

ویژگی های جدول توزیع فراوانی :

1-جدول توزیع فراوانی یک جدول پیوست است.

حدود واقعی اعداد که اعشاری است این پیوستگی را نشان می دهد.

برای پیوستگی اعداد سرجدول توزیع فراوانی از عدد کوچکتر نیم کم و به عدد بزرگتر نیم اضافه می کنیم.

حد واقعی اعداد	X	F
8/5 –11/5	*9–11*	5
5/5 – 8/5	6–8	3
3/5 – 5/5	3–5	2

2- حجم طبقه و فاصله

حجم طبقه تعداد اعدادی است که درون هر طبقه قرار دارد. مثلا در طبقه (5–3) سه عدد (3، 4 و 5) وجود دارد.

حجم طبقه این جدول 3 است. فاصله طبقه هم 3 است.

همیشه در جدول توزیع فراوانی حجم و فاصله با یکدیگر برابر هستند.

فراوانی تراکمی هر طبقه نیز برابر است با فراوانی طبقاتی همان طبقه به اضافه فراوانی های طبقات ماقبل آخر.

X	F	Cf	X/	$\frac{f}{n}\%$	$\frac{cf}{n}\%$
18-20	2	19	19	$\frac{2}{19}\times 100$	$\frac{19}{19}\times 100$
15-17	3	17	16	$\frac{3}{19}\times 100$	$\frac{17}{19}\times 100$
12-14	4	14	13	$\frac{4}{19}\times 100$	$\frac{14}{19}\times 100$
9-11	5	10	10	$\frac{5}{19}\times 100$	$\frac{10}{19}\times 100$

6-8	3	5	7	$\frac{3}{19}\times 100$	$\frac{5}{19}\times 100$
3-5	2	2	4	$\frac{2}{19}\times 100$	$\frac{2}{19}\times 100$
	$\sum f=n=19$				

همیشه آخرین طبقه cf با جمع f برابر است.

مرحله به دست آوردن cf :

17 +2=19　　14 +3=17　　4+10=14　　5 + 5=10　　3+2 =5

x' : نقطه میانی طبقات است .

$$x' = \frac{\text{حدبالا} + \text{حدپایین}}{2}$$

$$\frac{6+8}{2} = 7 \qquad \frac{3+5}{2} = 4$$

درصدگیری :

برای درصدگیری هر عدد را در کل تقسیم و در 100 ضرب می کنیم. $\frac{?}{n}\times 100$

مراحل تشکیل جدول توزیع فراوانی :

1-محاسبه دامنه تغییرات :

دامنه تغییرات با (R) نمایش داده می شود. برای تعیین آن، بزرگترین عدد منهای کوچکترین + یک می شود.

(یک اختلاف حدود واقعی اعداد (0/5) است)

$R = X_M - X_L + 1$ مثال $18 - 4 + 1 = 15$

2- تعیین تعداد طبقات :

تعداد طبقات معمولا بین 10 تا 20 است، اما بیشتر یا کمتر نیز ایرادی ندارد.

3- محاسبه فاصله طبقات :

با (i) نمایش داده می شود. $i = \frac{\text{دامنه تغییرات}}{\text{تعداد طبقات}}$

نکته : همیشه حد پایین طبقه اول می بایست مضربی از فاصله طبقات باشد.

انواع نمودار:

1- نمودار ستونی (میله ای یا خطی) 2- نمودار دایره ای 3-نمودارهیستوگرام

4- نمودار چند ضعلی 5-نمودار چند ضلعی تراکمی (اجایو)

جلسه سوم

نمودار ستونی و دایره ای:

از نمودار ستونی و دایره‌ای زمانی استفاده می کنیم که مقیاس اندازه گیری اسمی یا ترتیبی باشد و متغییر مورد مطالعه گسسته باشد.

از نمودار هیستوگرام، چندضلعی و چندضلعی متراکم (تراکمی) هنگامی استفاده می شود که مقیاس اندازه گیری حداقل فاصله‌ای باشد و متغیر مورد مطالعه پیوسته باشد.

نکته: حداقل فاصله‌ای باشد، یعنی قبلی‌ها مثل اسمی و ترتیبی را شامل نمی‌شود.

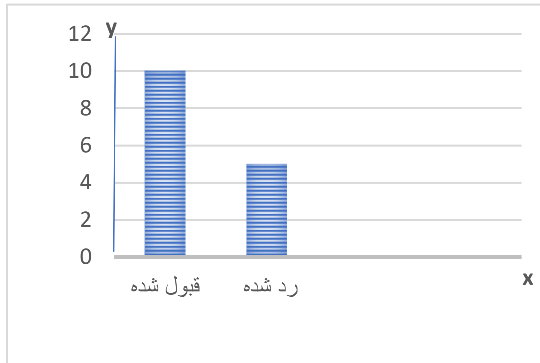
تعداد	نمره درس فیزیک
10	قبول شده

رشد شده	5

مثال:

محور y فراوانی و x ویژگی‌ها است.

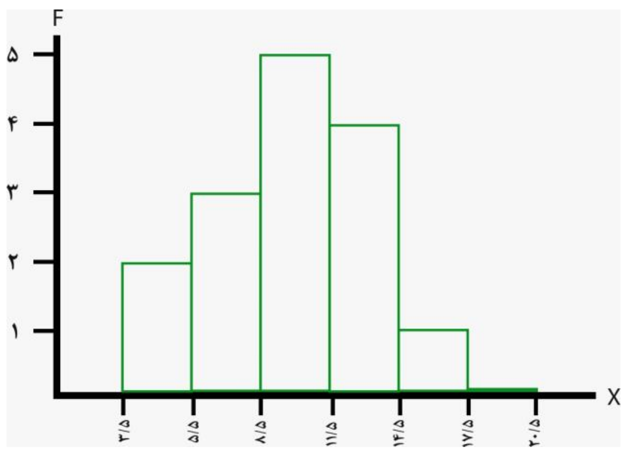
نمودار ستونی



نمودار دایره ای

جدول توزیع فراوانی پیوسته است.

نمودار هیستوگرام

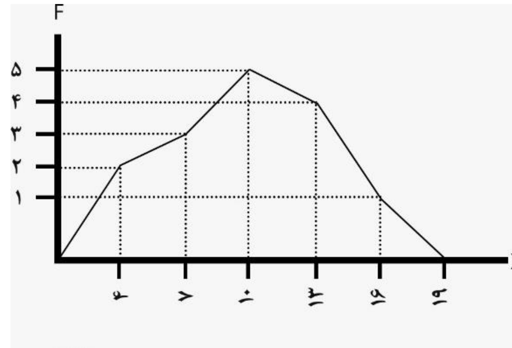


نمرات واقعی	X	f
18.5-20	18-20	0
14.5-17.5	15-17	1
12.5-14.5	12-14	4
9.5-11.5	9-11	5
6.5-8.5	6-8	3
3.5-5.5	3-5	2

نکته 1 : ممکن است فراوانی یک طبقه 0 باشد. وقتی فراوانی 0 باشد یعنی محور y (0) است. در این صورت روی خود محور x آن را پررنگ می کنیم.

نکته 2 : ممکن است کوچکترین عدد 500 باشد، نوشتن از 0 تا 500 مقدور نیست، لذا از 0 تا 499 را فاکتور می‌گیریم و با علامت // روی محور x نشان می دهیم. در واقع قبل از عدد مورد نظر، برای جبران فاصله این علامت را می گذاریم.

نمودار چند ضلعی



X	f	$x^{/}$
15–17	1	16
12–14	4	13
9–11	5	10
6–8	3	7
3–5	2	4

$x^{/}$: ایکس پیریم : نقطه میانی طبقات است.

نحوه بدست آوردن $x^{/}$: اعداد موجود در ستون x ها را با هم جمع کرده و تقسیم بر 2 می‌کنیم $x^{/}$ بدست می آید.

نکته: نقطه ای که بیشترین فراوانی را دارد، قله، نما یا مد نام دارد.

نکته: برای آنکه خطوط ابتدایی و پایانی نمودار بلاتکلیف نمانند، قبل از 4 و بعد از 16 به محور وصـــل می کنیم و فراوانی را 0 در نظر می گیریم .

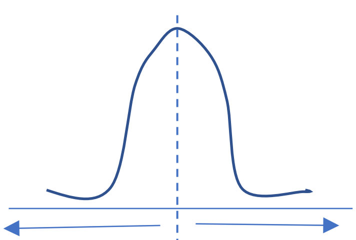
نمودار چند ضلعی متراکمی

X	f	cf
15–17	1	15
12–14	4	14
9–11	5	10
6–8	3	5
3–5	2	2

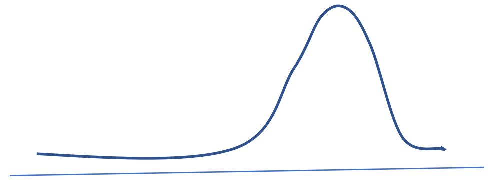
نکته: این نوع نمودار نزول ندارد و سیر صعودی طی می کند.

منحی متقارن :

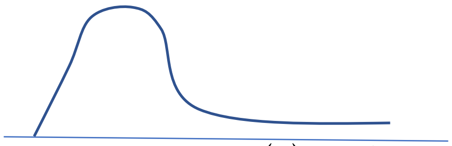
هیچ گاه خط x را قطع نمی کند و موازی ادامه می یابد.



منحنی نامتقارن یا منحنی دارای کجی :



دارای کجی چپ (–)

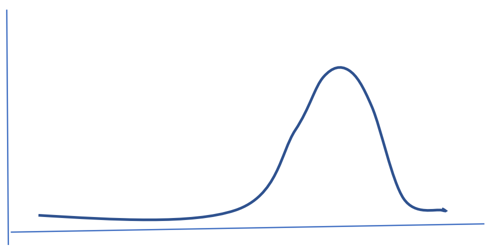


دارای کجی راست (+)

مثال:

سوال: معلم یک کلاس از دانش آموزان آزمون گرفته است، نمرات دانش آموزان دارای نمودار کجی مثبت شده است، آیا امتحان سخت بود؟

پاسخ: امتحان سخت بود، چون تعداد کمی نمره عالی گرفته اند. روی محور X نمرات به ترتیب زیاد می شود.



جلسه چهارم :

شاخص های مرکزی:

برای مقایسه چند مجموعه با یکدیگر به اجبار از این مجموعه‌ها یک نماینده انتخاب می کنیم. حد وسط را به عنوان نماینده بر می‌گزینیم و آن حد وسط شاخص مرکزی است.

شاخص‌های مرکزی آن دسته از شاخص‌هایی هستند که به کمک آن‌ها یک عدد را نماینده یا معرف مجموعه‌ای از اعداد می کنیم و به کمک آن‌ها راجع به کل اعداد قضاوت می کنیم. این اعداد در وسط قرار می‌گیرند.

شاخص‌های مرکزی میزان تمرکز را در وسط توزیع نمرات برای ما مشخص می کند.

انواع شاخص های مرکزی:

1- نما (مد)　　　2- میانه　　　3- میانگین

1- نما:

عددی است که دارای بیشترین فراوانی است و با (μo) نمایش داده می شود.

مثال: در بین اعداد زیر نما چه عددی است؟

2 4 4 4 6 8 9 μo=4

2 2 4 4 5 7 8 12 μo=2 و 4

نکته: اگر چند عدد متوالی پشت سر هم بودند و مد محسوب می‌شدند، باید وسط آن اعداد را به عنوان مد در نظر بگیریم .

مثال : 2 2 4 4 5 5 6 8

μo= 4.5 و 2 $\frac{4+5}{2}=4.5$

2 2 4 4 5 5 6 6 8

μo=5 و 2 $\frac{4+5+6}{3}=5$

زمانی از نما استفاده می کنیم که مقیاس اسمی باشد یا فرصت محاسبه‌ی بقیه شاخص های مرکزی را نداشته باشیم.

2- میانه :

میانه تحت تأثیر تعداده و با حجم کار نداره

اعداد نوعی اعدادی هستند که نوعشان با بقیه همخوانی ندارد اگر در جدول اعداد نوعی داشتیم میریم سراغ میانه چون اگر میانگین بگیریم حق بقیه تضییع میشود و تأثیر حجم را در شاخص مرکزی خنثی می کنیم.

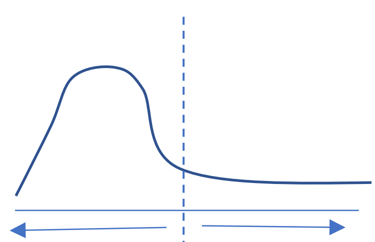
تعداد = جمع فراوانی‌ها

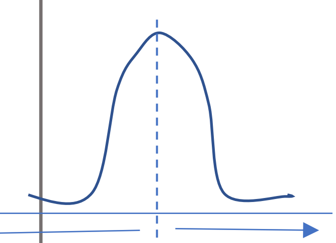
در میانه = اعداد را از کوچک به بزرگ مرتب می کنیم و می‌بینیم وسط کدام قرار میگیرد.

میانه را وقتی استفاده می‌کنیم که مقیاس ترتیبی باشد.

اگر در جدول عدد نوعی داشتیم ولی مقیاس ترتیبی نبود بازم از میانه استفاده می کنیم.

میانه را با (μd) نمایش می‌دهند و زمانی از آن استفاده می‌شود که مقیاس ترتیبی باشد. میانه شاخصی است که توزیع نمرات را به دو قسمت مساوی تقسیم می کند.





میانه :

اعداد طبقه بندی نشده (جدول ندارد)

اعداد طبقه بندی شده (جدول توزیع فراوانی دارد)

الف) اعداد طبقه بندی نشده :

1- زمانی که تعداد اعداد یا نمرات فرد باشد. 6 5 1 8 2

ابتدا اعداد را به ترتیب مرتب می کنیم : 8 6 5 2 1

عددی که در وسط قرار می گیرد میانه است : µd = 5

2- زمانی که تعداد اعداد یا نمرات زوج باشد. 16 12 9 7 6 2

$$\mu d = \frac{7+9}{2} = 8$$

3- زمانی که میانه در محدوده عددی قرار می گیرد که آن عدد تکراری است.

20 19 12 12 12 4

در این صورت ابتدا حدود واقعی اعداد را با استفاده از روش (0/5 کم) پیدا می کنیم و حد وسط 12 را پیدا و تقسیم بر 3 (چون 12 سه بار تکرار شده) می کنیم. مثال:

20 19 12 12 12 4

12.5 حد اختلاف (1) $\frac{1}{3}$

20 19 12.5 **12.16** 11.83 11.5 4

11.5 + 0.33 = 11.83

11.83 + 0.33 = 12.16

12.16 + 0.33 = 12.5

µd = 12.16

مثال:

20 19 12 12 12 12 4

12.5 حد اختلاف (1) $\frac{1}{4} = 0.25$

20 19 12.5 12.25 12 11.75 11.5 4

$$\mu d = \frac{12+12.25}{2} = 12.125$$

µd = 12.125

ب) اعداد میانه طبقه بندی نشده :

با استفاده از فرمول روبرو پیدا می کنیم :

$$\mu d = L + \frac{\frac{n}{2} - cf}{f}\,(i)$$

L : حد پایین طبقه‌ای که در آن میانه قرار دارد.

n : تعداد کل f است.

f : فراوانی همان طبقه‌ای که در آن میانه قرار دارد.

Cf : فراوانی تراکمی ماقبل طبقه‌ای که در آن میانه قرار دارد.

i : فاصله طبقات

نکته: برای پیدا کردن میانه و طبقه آن در جدول، باید $\frac{n}{2}$ از cf کوچکتر یا برابر باشد. $cf \geq \frac{n}{2}$

X	فراوانی f	فراوانی تراکمی Cf
48-50	2	50= $\sum F=n=50$
45-47	3	48
42-44	4	45
39-41	6	41
36-38	8	35
33-35	8	27
30-32	7	19
27-29	6	12
24-26	3	6
21-23	2	3
18-20	1	1
	$\sum F=n=50$	

مثال: $n=50$

$$\frac{50}{2} = 25$$

$$27 \geq 25$$

حد واقعی طبقه میانه را پیدا می کنیم

$$32.5 - 35.5$$

در فرمول جایگزاری می کنیم:

$$\mu\mathrm{d} = L + \frac{\frac{n}{2} - cf}{f}\,(i)$$

$$\mu\mathrm{d} = 32.5 + \frac{50-19}{8}\,(3) = 34.75$$

: فراوانی تراکمی ماقبل طبقه‌ای که در آن میانه قرار دارد.

فراوانی همان طبقه‌ای که در آن میانه قرار دارد.

حد پایین طبقه‌ای که در آن میانه قرار دارد.

ویژگی‌های میانه :

1- نما زمانی استفاده می شود که مقیاس اسمی و میانه زمانی استفاده می شود که مقیاس ترتیبی باشد.

2- **میانه** تحت تأثیر تعداد اعداد است و با حجم کار ندارده و فقط با تعداد کار دارد.

3- اگر در مجموعه اعداد نوعی داشتیم (یعنی نوعشان با بقیه متفاوت است، یا ابتدا است یا انتها) مثلا معلمی برای گرفتن میانگین نمرات دانش آموزان برای آنکه نمره دیگران خراب نشود کسی که 2 گرفته را جمع نمی‌بندد و از راه میانه محاسبه می کنیم. چون اگر از راه میانگین حساب کنیم حق بقیه تضییع می‌شود و تأثیر حجم را در شاخص مرکزی خنثی می‌کنیم.

تعداد = جمع فراوانی‌ها

در میانه : اعداد را از کوچک به بزرگ مرتب می‌کنیم و می‌بینیم کدام عدد در وسط قرار می‌گیرد.

میانه را وقتی استفاده می کنیم که مقیاس ترتیبی باشد.

اگر در جدول اعداد نوعی داشتیم ولی مقیاس ترتیبی نبود بازم از میانه استفاده می‌کنیم.

جلسه پنجم

میانگین :

شاخصی است که میزان تمرکز را در وسط توزیع نمرات برای ما مشخص می کند.

میانگین همان معدل‌گیری است که اعداد را با هم جمع می‌کنیم تقسیم بر تعداد می‌کنیم و بستگی به حجم دارد.

انواع میانگین:

1- هندسی 2- همساز (هارمونیک) 3- حسابی 4- وزین (مرکب)

1- فرمول هندسی : $G = \sqrt[n]{(x1)(x2)(x3)\ldots(xn)}$

مثال: نمرات 2 و 8 $G = \sqrt[2]{2 \times 8} = \sqrt[2]{16} = 4$

2- فرمول همساز : $HM = \frac{1}{\frac{1}{n}(\frac{1}{x1}+\frac{1}{x2}+\cdots\frac{1}{xn})}$

مثال: نمرات 2 و 8 $HM = \frac{1}{\frac{1}{2}(\frac{1}{2}+\frac{1}{8})} =$

3- فرمول حسابی :

هر چه عدد داریم را جمع می‌کنیم و تقسیم بر تعداد می‌کنیم. نماد آن $\bar{x}$ (x بار) است.

در نمونه: $\bar{x} = \frac{\sum x}{n}$ در جامعه : $\mu = \frac{\sum x}{N}$

میانگین حسابی

- اعداد طبقه بندی نشده : $\bar{x} = \frac{\sum x}{n}$
- اعداد طبقه بندی شده:
  - طبقات یکی‌یکی باشد. $\bar{x} = \frac{\sum fx}{n}$
  - i >1 یعنی فاصله طبقات چندتا چندتا باشد. $\bar{x} = \frac{\sum fx^{/}}{n}$ (روش اول)
  - i >1 (میانگین فرضی) $\bar{x} = x^{/} + \frac{\sum fx^{\sim}}{n}(i)$ (روش دوم)

$\bar{x}$= میانگین

i = فاصله طبقات

fx= هر عدد را در فراوانی‌اش ضرب می‌کنیم.

$x^{/}$ = نقطه میانی طبقه‌ای که فرض می‌شود در آن میانگین قرار دارد.

$x\sim$= فاصله هر طبقه با طبقه‌ای که فرض می‌شود در آن میانگین قرار دارد.

مثال برای وقتی که i=1 است :

$$\bar{x} = \frac{\sum fx}{n}$$

$$\bar{x} = \frac{124}{21} = 5.90$$

X	f	fx
8	7	56
7	2	14
6	3	18
5	2	10
4	5	20
3	2	6
	$\sum f = n = 21$	$\sum fx = 124$

مثال1- برای وقتی که i>1 است : (از روش اول)

$$\bar{x} = \frac{\sum fx^{/}}{n}$$

$$\bar{x} = \frac{163}{14} = 11.64$$

$x^{/}$: اعداد موجود در ستون x ها را با هم جمع کرده و تقسیم بر 2 می‌کنیم $x^{/}$ بدست می آید. به $x^{/}$ نماینده هر طبقه هم می گویند.

$fx^{/}$: f را در $x^{/}$ضرب می کنیم.

X	f	$x^{/}$	$fx^{/}$
15-17	4	16	64
12-14	2	13	26
9-11	3	10	30
6-8	5	7	35
3-5	2	4	8
	$\sum f = n = 14$		$\sum fx^{/} = 163$

مثال 2- برای وقتی که i>1 است : (از روش اول)

X	f	$x^{/}$	$fx^{/}$
45-49	1	47	47
40-44	2	42	84
35-39	3	37	111
30-34	6	32	192
25-29	8	27	216
20-24	17	22	374

15–19	26	17	442
10–14	11	12	132
5–9	2	7	14
	$\sum f = n = 76$		$\sum fx^{/} = 1612$

$$\bar{x} = \frac{\sum fx^{/}}{n}$$

$$\bar{x} = \frac{1612}{76} = 21/21$$

$x^{/}$: اعداد موجود در ستون x ها را با هم جمع کرده و تقسیم بر 2 می‌کنیم $x^{/}$ بدست می آید. به $x^{/}$ نماینده هر طبقه هم می گویند.

$fx^{/}$: f را در $x^{/}$ ضرب می کنیم.

میانگین فرضی: (روش دوم)

اعداد طبقه بندی شده را می توان از راه دیگری هم که به آن میانگین فرضی می گویند، بدست آورد.

فرمول میانگین فرضی: $\bar{x} = x' + \frac{\sum f\tilde{x}}{n}(i)$

x′ : نقطه میانی طبقه‌ای که فرض می شود در آن میانگین قرار دارد.

X کلاه دار: $\tilde{x}$: فاصله هر طبقه تا طبقه‌ای که فرض می شود در آن میانگین قرار دارد.

مثال برای روش دوم یا میانگین فرضی:

X	f	$\tilde{x}$	$f\tilde{x}$
42-44	3	+4	12
39-41	11	+3	33
36-38	8	+2	16
33-35	23	+1	23
30-32	35	0	0
27-29	14	-1	-14
24-26	10	-2	-20
21-23	9	-3	-27
18-20	6	-4	-24
15-17	1	-5	-5
	$\sum f = n = 120$		$\sum f\tilde{x} = -6$

$$x' = \frac{30 + 32}{2} = 31$$

$$\bar{x} = x' + \frac{\sum f\tilde{x}}{n}(i)$$

$$\bar{x} = 31 + \frac{(-6)}{120}(3)$$

$$\bar{x} = 30/85$$

فرض می کنیم میانگین در یکی از طبقات است، اغلب طبقه‌ای را در نظر می گیریم که بیشترین فراوانی را دارد.

x′ : اعداد موجود در ستون x ها را با هم جمع کرده و تقسیم بر 2 می کنیم x′ بدست می آید.

به x′ نماینده هر طبقه هم می گویند. لازم نیست برای کل جدول x′رابدست آوریم و فقط برای ردیفی که بیشترین فراوانی را دارد پیدا می کنیم.

fx′ : f را در x′ ضرب می کنیم.

i = فاصله طبقات

میانگین مرکب :

یعنی چند تا میانگین حسابی داریم و می‌خواهیم میانگین، میانگین‌ها را بدست آوریم.

دو حالت دارد:

1- وقتی nها با هم برابر باشند . یعنی n1=n2=n3.....ni

در این حالت فرمولش می‌شود:

$$\overline{x_T} = \frac{\sum \bar{X}i}{n} = \frac{\bar{x}1+\bar{x}2+\bar{x}3}{n}$$

n = تعداد میانگین‌ها

2- n ها با هم متفاوت باشند:

$$\overline{x_T} = \frac{\sum ni \times \overline{xi}}{N} = \frac{(n1 \times \bar{x}1) + (n2 \times \bar{x}2) + (n3 \times \bar{x}3)....}{n1 + n2 + n3 ....}$$

N= حاصل جمع nها است.

مثال برای حالت اول :

فرض می کنیم سه مجموعه داریم که تعداد افراد در هر سه مجموعه 15 نفر است .

$$n1 = 15 \qquad \bar{x}1 = 18$$
$$n2 = 15 \qquad \bar{x}2 = 17$$
$$n3 = 15 \qquad \bar{x}3 = 14$$

$$\overline{x_T} = \frac{\sum \bar{X}\, i}{n} = \frac{\bar{x}1 + \bar{x}2 + \bar{x}3}{n} \qquad \overline{x_T} = \frac{18 + 17 + 14}{3} = 16/33$$

مثال برای حالت دوم:

هر n در $\bar{x}$ ضرب می‌شود. در مخرج هم N منظور جمع n ها است.

$$n1 = 15 \qquad \bar{x}1 = 18$$
$$n2 = 15 \qquad \bar{x}2 = 14$$
$$n3 = 15 \qquad \bar{x}3 = 18$$

$$\overline{x_T} = \frac{\sum ni \times \overline{xi}}{N} = \frac{(n1 \times \bar{x}1) + (n2 \times \bar{x}2) + (n3 \times \bar{x}3)....}{n1 + n2 + n3 ....}$$

$$\overline{x_T} = \frac{(15 \times 18) + (20 \times 14) + (30 \times 18)}{15 + 20 + 30} =$$

نکته:

میانگین تحت تأثیر حجم است. میانگین از بقیه شاخص‌های مرکزی معتبر تر است.

جلسه ششم :

ویژگی‌های میانگین حسابی:

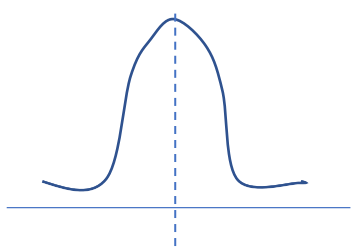
هرگاه عدد ثابتی را به اعداد جدول توزیع اضافه کنیم، میانگین به همان اندازه اضافه می‌شود. (4 عمل اصلی به همین شکل است)

X	x+c
1	3
3	5
4	6
2	4
5	7

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$$

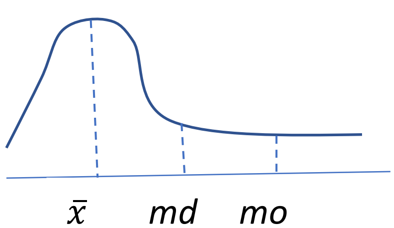
$$\bar{x} = \frac{\sum x + c}{n} = \frac{25}{5} = 5$$

اگر منحنی متقارن باشد، سه شاخص (نما، میان و میانگین) با هم برابر هستند.



mo=md=$\bar{x}$

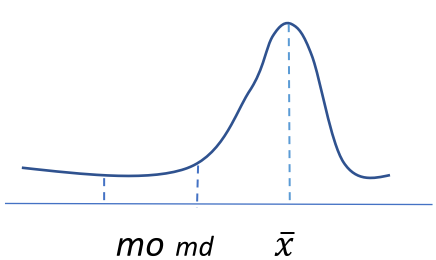
در کجی مثبت جایی که دو نیم می شود با هم برابر نیستند، میانگین بزرگتر می شود و بعد نما و بعد میانه



$\bar{x}$ > md > mo

رابه در کجی منفی برعکس است.

$\bar{x}$ < md < mo



توضیحات کلی در مورد نما، میانه، میانگین:

وقتی از نما استفاده میکنیم که مقیاس اسمی باشد.

وقتی از میانه استفاده می‌کنیم که مقیاس ترتیبی باشد.

وقتی از میانگین استفاده می‌کنیم که مقیاس فاصله‌ای باشد و حجم را در آن بخواهیم دخیل کنیم.

اگر بخواهیم حجم را در شاخص مرکزی دخیل بدهیم به سراغ میانگین که معتبرترین شاخص مرکزی است می‌رویم.

نما بی‌اعتبارترین شاخص مرکزی است.

نمونه سوال:

توزیع فراوانی زیر را در نظر بگیرید، باتوجه به توزیع داده شده نشان دهید که توزیع متقارن است یا دارای کجی مثبت و منفی؟

پاسخ: برای جواب باید نما، میانه و میانگین را بدست آوریم.

cf : فراوانی تراکمی است و هر عدد با قبلی خود جمع می‌شود.

L : حد واقعی طبقه‌ای است که در آن میانه قرار دارد.

n=34 $\frac{n}{2} = 17$

$$\mu d = L + \frac{\frac{n}{2} - cf}{f}(i)$$

میانه : $md = 10/5 + \frac{17 - 13}{5}(1) = 11/3$

میانگین: $\bar{x} = \frac{\sum fx}{n}$

$$\bar{x} = \frac{415}{34} = 12/21$$

در نتیجه $\bar{x} > md > mo$

دارای کجی مثبت است.

X	f	cf	fx
20	3	34	60
18	1	31	18
17	1	30	17
16	2	29	32
14	4	27	56
12	5	23	60
11	5	18	95
10	6	13	60
9	4	7	35
7	3	3	21
	$\sum f = n = 34$		

شاخص های پراکندگی :

برای مقایسه کردن باید علاوه بر شاخص مرکزی، شاخص پراکندگی را هم بدست آوریم.

X	y
2	0
1	3
3	3
4	4
5	5

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{\sum y}{n} = \frac{15}{5} = 3$$

برای شاخص پراکندگی به مبداء نیاز است. ما آن مبداء را شاخص مرکزی (نما، میانه، میانگین) در نظر می‌گیریم که معتبرترین آن میانگین است. در نتیجه ما اغلب پراکندگی را با میانگین مقایسه می‌کنیم.

پراکندگی یک شاخص نسبی است. مثلاً اگر در اولین جلسه کلاس از دانشجویان امتحان گرفته شود، انتظار می‌رود پراکندگی در نمرات زیاد باشد، اما در امتحان آخر انتظار می‌رود پراکندگی کم باشد.

پراکندگی:

پراکندگی شاخص نسبی است که به ما می‌گوید هر یک از اعداد در یک توزیع بطور متوسط چه مقدار از یک شاخص مرکزی اختلاف یا انحراف (تفریق) دارند.

انواع شاخص پراکندگی:

1- دامنه تغییرات   2- انحراف متوسط یا متوسط انحرافات   3- واریانس

4- انحراف استاندارد یا معیار   5- انحراف چارکی

1- دامنه تغییرات :

$$R = X_M - X_L + 1$$

1 + بزرگترین عدد – کوچکترین عدد = دامنه تغییرات

دامنه تغییرات تنها از دو عدد استفاده می‌کند و مابقی را در نظر نمی‌گیرد، همچنین مبداء که میانگین است را ندارد.

2- انحراف متوسط : AD یا DA

مجموع قدر مطلق انحراف نمرات از میانگین تقسیم بر تعداد آن‌ها

$$AD = \frac{\sum|X-\bar{X}|}{n}$$

در قدر مطلق هر عدد منفی + می‌شود.

مثال:

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$$

برای تشکیل ستون دوم، اعداد را از میانگین کم می‌کنیم.

میانگین 3 است.

x	$\bar{x}$	$x - \bar{x}$	$\lvert X - \bar{X}\rvert$
1	3	1-3 = -2	2
3	3	3-3 = 0	0
4	3	4-3 =1	1
2	3	2-3 = -1	1
5	3	5-3 =2	2
∑X=15			

جلسه هفتم :

3- واریانس : $S^2 = \frac{\sum(X-\bar{X})^2}{n}$

واریانس شاخصی است که به ما می‌گوید هر یک از اعداد بطور متوسط چه اندازه از میانگین انحراف دارند.

واریانس

اعداد طبقه بندی نشده

از راه انحراف از میانگین $S^2 = \frac{\sum(X-\bar{X})^2}{n}$

از راه اعداد خام $S^2 = \frac{\sum x^2 - \frac{(\sum X)^2}{n}}{n}$

اعداد طبقه بندی شده

فاصله طبقات ی

1 است.

از راه انحراف از میا $S^2 = \sum f\,(x-\bar{x})^2 = \frac{\sum f x^2}{n}$

از راه اعداد خام $S^2 = \frac{\sum f\,x^2 - \frac{(\sum fX)^2}{n}}{n}$

فاصله طبقات بیشتر از

است.

1- از راه انحراف از میانگین

$$S^2 = \frac{\sum f(X^{/}-\bar{X})^2}{n} = \frac{\sum f x^2}{n}$$

2- از راه اعداد خام

$$S^2 = \frac{\sum f x^2 - \frac{(\sum f\,x^{/})^2}{n}}{n}$$

مثال طبقه بندی شده: 1 = i

X	f
8	2
7	3
6	2
5	3

4	5
3	2

1- راه انحراف از میانگین :

$$S^2 = \sum f\,(x-\bar{x})^2 = \frac{\sum f x^2}{n}$$

2- راه اعداد خام:

$$S^2 = \frac{\sum f\,x^2 - \frac{(\sum f X)^2}{n}}{n}$$

مثال برای طبقه بندی شده : وقتی i > 1

1- از راه انحراف از میانگین :

X	f	$x^/$	$fx^/$	$x^/ - \bar{x} = x$	$(x^/ - \bar{x})^2$	$f\,(x^/ - \overline{x}) = fx^2$
21-23	5	22	110	+9	81	405
18-20	7	19	133	+6	36	252
15-17	8	16	128	+3	9	72
12-14	10	13	130	0	0	0
9-11	8	10	80	-3	9	72
6-8	7	7	49	-6	36	252
3-5	5	4	20	-9	81	405
	$\sum F = n = 50$		$\sum fx^/ = 650$			$\sum f\,(x^/ - \overline{x}) = \sum fx^2 = 1458$

$$\bar{x} = \frac{\sum f x^/}{n} = \frac{650}{50} = 13$$

$$S^2 = \frac{\sum f (X^/ - \bar{X})^2}{n} = \frac{\sum f x^2}{n} = \frac{1458}{50} = 29/16$$

$x^/$ = نقطه میانی بین دو عدد است.

$f\,x^/$ = حاصل ضرب f در $x^/$ است.

$\bar{x}$= میانگین است که از فرمول $\bar{x} = \frac{\sum f x^/}{n}$ بدست می آید.

$(x^/ - \overline{x})$= $x^/$ را منهای میانگین بدست آمده می کنیم.

2- از راه اعداد خام :

$$S^2=\frac{\sum fx^2-\frac{(\sum f\,x^/)^2}{n}}{n}$$

$$S^2=\frac{6887-\frac{(557)^2}{50}}{50}=13/65$$

برای بدست آوردن ستون $fx^{/2}$: ستون $x^/$ را در ستون $fx^/$ ضرب می کنیم.

X	f	$x^/$	$fx^/$	$fx^{/2}$
21–23	5	22	110	484
18–20	7	19	133	722
15–17	8	16	128	1536
12–14	10	13	130	1521
9–11	8	10	80	2200
6–8	7	7	49	392
3–5	5	4	20	32
	∑F=n=50		$\sum fx^/=557$	$\sum fx^{/2}=6887$

جلسه هشتم

4- انحراف استاندارد یا معیار :

اگر بخواهیم واریانس را به توان 2 برسانیم، باید هم اعداد به توان برسد، برای رفع این مشکل جذر می گیریم.

انحراف استاندارد معتبرترین شاخص پراکندگی است.

$$s=\sqrt{\frac{\sum fx^{/2}-\frac{(\sum fx^/)^2}{n}}{n}}$$

$$\sqrt{s^2=s}$$

5- انحراف چارکی

چارک مخفف چهار ک است.

چارک‌ها نقاطی روی مقیاس اندازه گیری هستند که کلیه مشاهدات ما را به چهار قسمت مساوی تقسیم می کنند.

عدد چارک اول Q1 و چارک دوم Q2 ، چارک سوم Q3 و چارک چهارم Q4 است.

٪25 ٪50 ٪75 ٪100

$\frac{n}{4}=Q1$ $\frac{2n}{4}=\frac{n}{2}=Q2$ $\frac{3n}{4}=Q3$ $\frac{4n}{4}=n=Q4$

انحراف چارکی : متوسط پراکندگی را در اطراف میانه نشان می‌دهد و مبداء میانه است.

Q2 میانه است: $Q_2 = md$

اطراف میانه Q1 و Q3 است.

انحراف چارکی اعداد طبقه بندی شده :

مثال: انحراف چارکی این اعداد را بدست آورید؟

9 4 5 10 12 8 2

جواب: اول اعداد را مرتب می‌کنیم از کوچک به بزرگ

12 10 9 8 5 4 2

n=7 یعنی ما 7 تا عدد داریم.

$$\frac{n}{4} = \frac{7}{4} = 1/75 \qquad Q1 = 2 + 0/75 \times (4 - 2) = 3/5$$

1/75 = یک یعنی از اولین عدد 75 صدم می‌رویم جلو

در محاسبه به صورت بالا می نویسم یعنی عدد اول را می نویسیم و 0/75 را ضرب در اختلاف عدد دوم با اول می کنیم.

یعنی از اولین عدد 0/75 فاصله این عدد اول با عدد دوم را باید برویم جلو یعنی Q1 بین اولین عدد و دومین عدد است.

$$\frac{3n}{4} = \frac{21}{4} = 5/25 \qquad Q3 = 9 + 0/25 \times (10 - 9) = 9/25$$

5/25 = 5 یعنی از عدد پنجم 0/25 صدم می رویم جلو

یعنی از عدد پنجم که عدد (9) است ، 0/25 فاصله این عدد ششم با عدد پنجم را باید برویم جلو

$$Q = \frac{Q3 - Q1}{2} = \frac{9.25 - 3.5}{2} = 2/875$$

انحراف چارکی اعداد طبقه بندی شده : $Q = \frac{Q3-Q1}{2}$

$\frac{n}{4}$ = جایگاه چارک اول

$$Q1 = L + \frac{\frac{n}{4} - cf}{f}\,(i)$$

جایگاه چارک دوم = $\frac{n}{2}$

$$Q2 = md = L + \frac{\frac{n}{2} - cf}{f}\ (i)$$

جایگاه چارک سوم = $\frac{3n}{4}$

$$Q3 = L + \frac{\frac{3n}{4} - cf}{f}\ (i)$$

مثال : محاسبه انحراف چارکی :

شرط چارک اول: در این جا چارک اول در اولین طبقه‌ای است که cf آن بشود $\frac{n}{4}$ یا بیشتر از $\frac{n}{4}$ یعنی :

$$CF \geq \frac{n}{4}$$

شرط چارک سوم : چارک سوم در طبقه‌ای است که CF آن بشود $\frac{3n}{4}$ یا بیشتر یعنی :

$$CF \geq \frac{3n}{4}$$

X	f	CF	
85–89	1	40	
80–84	1	39	
75–79	1	38	
70–74	2	37	
65–69	3	35	
60–64	4	32	چارک سو
55–59	6	28	
50–54	7	22	
45–49	5	15	
40–44	3	10	چارک او
35–39	3	7	
30–34	2	4	
25–29	1	2	
20–24	1	1	
	$\sum f = n = 40$		

$$Q1 = L + \frac{\frac{n}{4} - cf}{f}(i)$$

$n= 40$ $\frac{n}{4} = \frac{40}{4} = 10$

در اولین طبقه‌ای که cf آن 10 باشد چارک اول وجود دارد. اعداد را به حد واقعی می نویسیم (نیم کم و نیم زیاد می کنیم)

$$Q1 = \frac{39}{5} + \frac{10-7}{3}(5) = 44.5$$

چارک سوم : $\frac{3n}{4} = \frac{3\times40}{4} = 30$

$$Q3 = L + \frac{\frac{3n}{4} - cf}{f}(i)$$

$$Q3 = 59.5 + \frac{30-28}{4}(5) = 62$$

$$Q = \frac{Q3 - Q1}{2} = \frac{62 - 44.5}{2} = 8/75$$

cf: فراوانی تراکمی ماقبل طبقه‌ای که در آن میانه قرار دارد.

f : فراوانی همان طبقه‌ای که در آن میانه قرار دارد.

L : حد پایین طبقه‌ای که در آن میانه قرار دارد.

نمرات استاندارد:

نمرات خام که اعداد عادی مباحث گذشته بودند با یکدیگر قابل مقایسه نبودند، به همین منظور و برای مقایسه نمرات طی یک انتقال، نمرات خام را به نمرات استاندارد تبدیل می کنیم.

به این نوع انتقال، انتقال خطی می گویند، ویژگی‌های اصلی عدد از بین نمی‌رود و فقط امکان مقایسه را فراهم می کند.

نمرات استاندارد: گاهی اوقات در تحقیق علاقمندیم که نمرات دو یا چند آزمودنی (شخص) را در مورد دو یا چند متغیر مورد مقایسه قرار دهیم و در موردشان قضاوت و داوری کنیم، برای این کار نمرات را استاندارد می کنیم.

در استاندارد سازی نمرات

1- ابتدا واحدها را یکی و یکدست می کنیم.

2- هر نمره را به توزیع خود برمی گردانیم.

نمرات استاندارد نمراتی هستند که در اثر انتقال خطی نمراه خام به دست می‌آید.

انواع نمرات استاندارد:

1- نمرات z 2- نمرات T 3- رتبه درصدی   4- نقطه درصدی

نمرات Z : $Z = \frac{X-\bar{X}}{S}$

$\bar{X}$= میانگین S = انحراف معیار

$$Z = \frac{12-14}{2} = -1$$
$$Z = \frac{14-16}{2} = -2$$
$$Z = \frac{10-8}{2} = 1$$

	X	$\bar{X}$	S	Z
فیزیک	12	14	2	-1
زبان	14	16	1	-2
ریاضی	10	8	2	+1

S: جذر واریانس (معتبرترین شاخص پراکندگی)

نکته: هر چه Z مثبت‌تر باشد فرد در درس قوی‌تر است.

نکته: همیشه در یک توزیع میانگین نمرات Z برابر با 0 و انحراف استاندارد نمرات Z برابر با یک می باشد.

یک توزیع یعنی فقط یک درس باشد.

نقاط قوت نمرات z :

1- نمره استاندار است و امکان مقایسه را فراهم می کند.

نقاطه ضعف :

1- دامنه آن محدود است . از 3/700- تا 3/70 + می تواند باشد. به درد جمعیت زیاد نمی خورد

2- z می تواند منفی باشد. نمره منفی برای یک انسان نیست.

3- با z نمی توان به راحتی با دیگران ارتباط برقرار کرد.

نمرات T : $T = (S)(Z) + \bar{X}$

نمرات استانداری که در اثر انتقال خطی نمره خام بدست می‌آید.

Z وقتی صفر است یعنی نمره با میانگین یکی است که در هم کم می شود صفر می شود.

$$T = 150Z + 500$$

مثال: آزمون تافل است و کسی قبول می شود که T آن حداقل 500 باشد. Z باید حداقل (0) باشد تا فرد بتواند قبل شود.

نقاط قوت T :

1- مثل z امکان مقایسه را فراهم می‌کند.

2- دامنه آن محدود نیست و وسیع است.

3- نمره نهایی منفی نمی شود.

نقاط ضعف :

1- نمی توان با T به راحتی ارتباط برقرار کرد.

رتبه درصدی : $P_R = \frac{CF + \frac{F}{2}}{n} \times 100$

تعیین می کند چند درصد از افراد در یک توزیع از یک نمره معینی کمتر گرفتند.

خود نمره و بالاتر از آن نمره را شامل نمی‌شود و با نماد P_R نمایش می‌دهند.

مثال: نمره چند نفر زیر 17 است؟

پاسخ : یعنی چند درصد از افراد P_{17} است

آن‌هایی که خود 17 و 17 به بالا گرفته‌اند شامل رتبه درصدی نمی‌شوند. $P_{17} = \%77/5$

X	f	CF
19	3	40
18	5	37
17	2	32
16	5	30
15	5	25
14	4	20
13	7	16
12	2	9

11	7	7
	$\sum f = n = 40$	

$$P_R = \frac{CF + \frac{F}{2}}{n} \times 100$$

$$P_{17} = \frac{30 + \frac{2}{2}}{40} \times 100 = 77/5\%$$

$$100 - 77/5 = 22/5\%$$

$$P_{16} = \frac{20 + \frac{5}{2}}{40} \times 100 = 68/75\%$$

$$P_{19} = \frac{37 + \frac{3}{2}}{40} \times 100 = 96/25\%$$

Cf: طبقه ماقبل

$\frac{F}{2}$= نصف f همون طبقه

هر چه فراوانی نمره کوچکتر باشد رتبه درصدی بیشتر می شود.

نقطه درصدی : $P_X = L + \frac{P_n - cf}{f}(i)$

L : حد پایین طبقه‌ای که در آن نقطه قرار دارد.

P_n: جایگاه نقطه را نشان می‌دهد. $p_n = p_R \times n$

Cf : فراوانی تراکمی طبقه ماقبل نقطه

f: فراوانی همان طبقه‌ای که نقطه قرار دارد.

i : فاصله طبقات

نقطه درصدی تعیین می کند که یک درصد معین برابر چه نقطه یا چه نمره‌ای است. (درصد می‌دهد و نمره می‌خواهد)

آن را با نماد P_X نمایش می دهند.

اولین طبقه‌ای که cf آن بشود P_n در نتیجه = $cf \geq P_n$

مثال: معلمی می‌خواهد امتحان بگیرد و می‌خواهد به 5٪ بالای کلاس جایزه بدهد. دانش آموزان باید چه نمره‌ای بگیرند؟

پاسخ: چون 5 درصد بالای نمرات را خواسته پس ما باید 100 درصد را منهای 5 درصد کنیم و 95 درصد را محاسبه کنیم.

٪95 = ٪5 – ٪100 $P_{95} = ?$ در حقیقت P_{95} جایگاه نقطه را نشان می دهد.

X	f	CF
18	7	50

17	4	43
16	4	39
15	6	35
14	4	29
13	5	25
12	3	20
11	7	17
10	5	10
9	5	5
	$\sum f = n = 50$	

p_R= اینجا باید اعشاری نوشته شود. 0/95 = ٪95

$p_n = p_R \times n$ طبقه نقطه

$$p_n = 0/95 \times 50 = 47/5$$

شرط : $cf \geq P_n$

$$P_X = L + \frac{P_n - cf}{f}(i)$$

$$P_{95} = 17/5 + \frac{47/5 - 43}{7}(1) = 18/13$$

جلسه نهم :

همبستگی و پیش بینی :

اگر در امتحان رابطه را خواست، باید از راه همبستگی حل شود.

برای اینکه ببینیم بین دو یا چند متغیر رابطه وجود دارد یا نه از همبستگی استفاده می کنیم.

همبستگی شاخصی است که به کمک آن حدود تغییرات یک یا چند متغیر را نسبت به حدود تغییرات یک یا چند متغیر دیگر تعیین می کند. مثلاً ارتباط همبستگی میان بارندگی و آب رودخانه‌ها

میزان و مقدار این تغییرات را با شاخصی به نام شاخص ضریب همبستگی نشان می دهیم.

مثبت و منفی بودن نشان دهنده قوی و ضعیف نیست بلکه نشان دهنده جهت است.

ضریب همبستگی را با نماد r_{xy} نمایش می‌دهند.

عدد ضریب همبستگی می تواند از 1− تا 1+ باشد. اگر 0 باشد یعنی همبستگی وجود ندارد.

در اینجا عدد منفی (−) به معنای ضعیف بودن نیست. هر چه به سمت صفر (0) برویم همبستگی ضعیف می شود.

به مثبت همبستگی مستقیم و به منفی همبستگی معکوس هم می گویند.

همبستگی مستقیم : $r_{xy} = +1$ در نتیجه $0 < r_{xy} < 1$

تغییرات یک متغیر هم جهت با یک متغیر دیگر است. مثال:

هر چه بهره هوشی بیشتر، میزان یادگیری بیشتر است و هر چه بهره هوشی کمتر، میزان یادگیری کمتر است.

همبستگی معکوس: $r_{xy} = -1$ در نتیجه $-1 < r_{xy} < 0$

با افزایش یک متغیر دیگری کاهش می‌یابد یا برعکس . مثال:

ارتباط میان اضطراب و یادگیری. هر چه اضطراب بیشتر ، میزان یادگیری کمتر است.

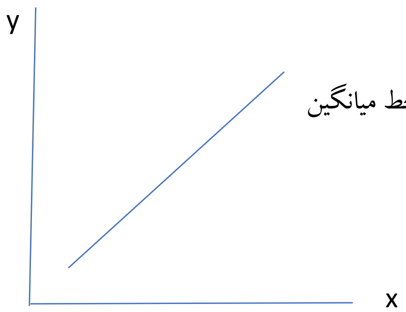
همبستگی از دو مؤلفه تشکیل شده :

1- جهت همبستگی = مثبت و منفی

2- شدت همبستگی = عدد آن است.

شدت به ما می گوید که همبستگی ضعیف است یا قوی.

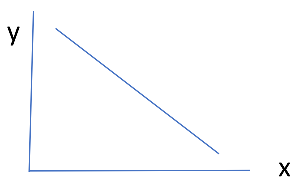
برای تشخیص همبستگی یا برای تعیین رابطه میان متغیرها 2 روش وجود دارد:



روش اول : رسم نمودار

حالت اول: اگر نقاطی که قصد سنجش همبستگی آن را داریم، روی خط میانگین قرار گرفته است، در این حالت می‌گویند همبستگی 1+ ، کامل و مستقیم است.

در حالت دوم: اگر نقاط دور خط میانگین قرار بگیرد، همبستگی بین 1+ و 0 است و هر چه به خط نزدیک شویم به 1+ نزدیک شده‌ایم.



در حالت سوم : خط میانگین برعکس حالت یک است.

همبستگی 1- ، معکوس و منفی است.

در حالت چهارم: همبستگی (خطوط بیضی) بین 1- و 0 قرار دارد و هر چه به خط نزدیک شود به 1- نزدیک‌تر می‌شود و قوی‌تر است.

نکته 1- : برای بحث همبستگی حتماً باید رابطه بین متغیرها خط باشد، خطوط حالت منحنی ندارد.

نکته 2- : در آمار به منفی می‌گوییم انحراف، هر چه از خط میانگین فاصله بگیریم انحراف بیشتر می‌شود و همبستگی کمتر می‌شود.

نکته 3- : وقتی می‌خواهیم + و — را تشخیص دهیم، به شیب نمودار توجه می‌کنیم.

روش دوم : ضرایب همبستگی:

از ضرایب همبستگی استفاده می‌کنیم و عدد همبستگی را بدست می‌آوریم.

زمانی می‌توانیم سراغ همبستگی برویم که رابطه خطی باشد.

خط ، خط میانگین است، هر چه از خط دور می‌شویم همبستگی ضعیف می‌شود چون به صفر نزدیک میشود.

انواع ضرایب همبستگی:

1- ضریب همبستگی گشتاوری پیرسون

2- ضریب همبستگی اسپیرمن براون

در پیرسون مقیاس اندازه گیری باید حداقل فاصله‌ای باشد و متغیر مورد مطالعه پیوسته باشد.

در اسپیرمن مقیاس اندازه‌گیری باید ترتیبی باشد و متغیر مورد مطالعه گسسته باشد.

همبستگی پیرسون، همبستگی بین نمرات است و اسپیرمن رتبه است.

ضریب همبستگی پیرسون:

آن را با نماد *rxy* نمایش می‌دهند و محاسبه آن سه روش دارد.

روش اول: از طریق نمرات استاندارد $rxy = \frac{\sum z_x \times (z_y)}{n}$

Z = همان نمره استاندارد است. $z_x = \frac{x-\bar{x}}{s_x}$ $z_y = \frac{y-y}{s_y}$

روش دوم: از راه انحراف میانگین: $rxy = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$

(مهم) روش سوم: از راه اعداد خام $rxy = \frac{n\sum xy-(\sum x)\,(\sum x)}{\sqrt{[n\,\sum x^2-(\sum x)^2\,] - [n\,\sum y^2-(\sum y)^2\,]}}$

مثال: محققی علاقمند است که همبستگی بین واکنش نسبت به محرک بصری را با واکنش نسبت به محرک‌های شنیداری مورد اندازه گیری قرار دهد، به همین منظور 5 نفر آزمودنی به صورت تصادفی انتخاب و آن‌ها را در معرض محرک‌های مورد پژوهش قرار می‌دهند. در صورتی که اطلاعات جمع آوری شده به شرح زیر موجود باشد، ضریب همبستگی بین این دو متغیر را محاسبه کنید؟

X واکنش نسبت به محرکهای بصری	Y واکنش نسبت به محرکهای شنیداری	x^2	y^2	Xy
5	1	25	1	5
2	2	4	4	4
3	3	9	9	9
4	4	16	16	16
1	5	1	25	5
$\sum x = 15$	$\sum y = 15$	$\sum x^2 = 55$	$\sum y^2 = 55$	$\sum xy = 39$

$$rxy = \frac{n\sum xy - (\sum x)\,(\sum x)}{\sqrt{[n\,\sum x^2 - (\sum x)^2\,] - [n\,\sum y^2 - (\sum y)^2\,]}}$$

$$rxy = \frac{(5)\times(39)\ -\ (15)\ \times\ (15)}{\sqrt{[(5)\times\ (55)\ -\ (15)^2]\ -\ [(5)\times\ (55)\ -\ (15)^2]}}$$

$$rxy = -0/60$$

منفی یعنی معکوس

همبستگی وجود دارد و نسبتاً قوی است چون از وسط بیشتر و واکنش معکوس است.

همبستگی اسپیرمن: $rs = 1 - \frac{n\,(\sum D^2)}{n\,(n^2-1)}$

همبستگی اسپیرمن با نماد rs نمایش داده می‌شود.

D : اختلاف بین رتبه‌ها است. $D = RX - RY$

توضیح:

نمرات هر درس هر نفر روبه‌روی نمره درس خودش است.

برای تبدیل نمره به رتبه دو شرط وجود دارد:

1- همیش به تعداد افراد رتبه داریم.

2- بالاترین نمره بالاترین رتبه است.

مثلاً اگر بالاترین نمره 20 است، رتبه اول است.

مثال:

درس فیزیک:

بالاترین رتبه فیزیک 19 است که می‌شود رتبه 1 ولی چون 2 تا 19 است به یکی رتبه 1 و به دیگری 2 می‌دهم، رتبه‌ها را جمع و تقسیم بر تعدادشان می‌کنیم می‌شود 1/5. $\frac{1+2}{2} = 1/5$ به هر 2 نفر رتبه 1/5 می‌دهیم.

رتبه بعدی 18 است. و باز دو 18 داریم پس به یکی رتبه 3 و به دیگری رتبه 4 می‌دهیم با هم جمع کرده و تقسیم بر تعدادشان می کنیم می‌شود 3/5

به عدد 16 رتبه 5 و عدد 15 رتبه 6 می‌دهیم.

درس ریاضی:

به 19 رتبه 1 ، به 18 چون دو تا است رتبه 2/5 ، به 17 چون دو تا است رتبه 4/5 و به 14 رتبه 6 می‌دهیم.

نمره اختصاصی نمره فیزیک RX		فیزیک x	ریاضی y		رتبه اختصاصی نمره ریاضی RY
$\frac{3+4}{2} = 3/5$	4	18	19	1	1
6	6	15	18	2	$\frac{2+3}{2} = 2/5$

$\frac{1+2}{2}=1/5$	1	19	17	4	$\frac{4+5}{2}=4/5$
$\frac{1+2}{2}=1/5$	2	19	18	3	$\frac{2+3}{2}=2/5$
$\frac{3+4}{2}=3/5$	3	18	17	5	$\frac{4+5}{2}=4/5$
5	5	16	14	6	6

حل مسأله :

فیزیک (X)	ریاضی (Y)	RX	RY	D	D^2
18	19	3/5	1	2/5	6/25
15	18	6	2/5	3/5	12/25
19	17	1/5	4/5	-3	9
19	18	1/5	2/5	-1	1
18	17	3/5	4/5	-1	1
16	14	5	6	-1	1
					$\sum D^2=30/5$

$$rs = 1 - \frac{6(30/5)}{6(36-1)} = 0/13$$

جلسه دهم:

پیش بینی (رگروسیون)

وقتی به یک همبستگی معناداری برسیم (بازه 1- تا 1+) ، در قدم بعدی پیش بینی می کنیم.

در پیش بینی تأثیر را می‌سنجیم.

(متغیر پیش بینی شونده، متغیر پیش بینی کننده (ملاک))

اگر در عنوان سوال رابطه بخواهد همبستگی است، اگر تأثیر بخواهد پیش بینی است.

موفقیت در دانشگاه را به کمک نمره‌ی آزمون داوطلب پیش‌بینی می کنیم.

موفقیت پیش بینی شونده است و نمره آزمون داوطلب پیش بینی کننده

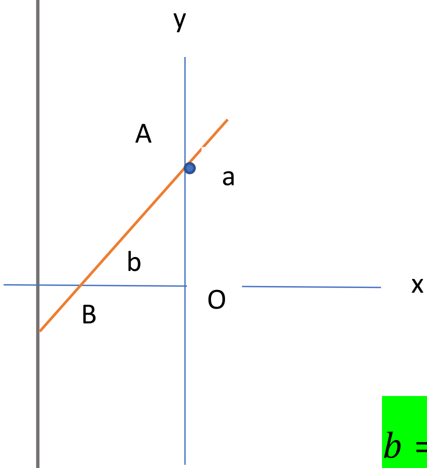
براساس معادله خط پیش بینی را انجام می‌دهیم.

$$y = a + bx$$

X = پیش بینی کننده

y = پیش بینی شونده

a و b = ضرایب معادله رگرسیونی



مثال: $y = 2 + x$

برای پیدا کردن نقاط یک بار به x و یک بار به y عدد صفر می‌دهیم تا نقاط را پیدا کنیم.

A	B	0	-2
		2	0

به x در فرمول صفر می‌دهیم و A بدست می‌اید.

b شیب خط است که در اینجا یک است.

$$b = \frac{OA}{AB} \quad \text{یا} \quad bxy = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$axy = \frac{\sum y - (bxy)(\sum x)}{n}$$

خطای استاندارد برآورد پیش بینی :

شاخصی است که در پیش‌بینی تغییرات یک متغیر بطور متوسط چه مقدار مرتکب خطا شده‌ایم را مشخص می‌کند. با نماد Sxy نمایش می دهند.

$$sxy = sy\sqrt{1 - (rxy)^2}$$

هر چه همبستگی بیشتر باشد، خطا کمتر است.

کمترین خطا 0 است. $sxy=0$ $rxy=\pm 1$

بیشترین خطا sy است. $sxy=sy$ $rxy=0$

ح

معلمی علاقمند است که تغییرات نمره روش تحقیق را به کمک نمرات درس آمار پیش بینی کند. به همین منظور 5 نفر دانش‌جو به صورت کاملاً تصادفی از بین دانشجویان رشته ارتباطات انتخاب می‌کند، سپس آزمون روش تحقیق را برای آن‌ها اجرا و نمره درس آمار آن‌ها را از کارنامه تحصیلیشان استخراج می‌کند. در صورتی که اطلاعات جمع آوری شده به شرح زیر موجود باشد.

1- همبستگی بین 2 درس را محاسبه کنید؟

2- معادله پیش بینی نمرات روش تحقیق به کمک نمرات آمار را تدوین کنید؟

3- درصورتی که دانشجویی در درس آمار نمره 6 گرفته باشد، چه نمره‌ای در درس روش تحقیق برای او پیش‌بینی می‌کنید؟

4- مقدار خطای پیش بینی برای آزمودنی شماره 5 چقدر است؟

5- خطای استاندار برآورد پیش بینی را محاسبه کنید؟

نمرات آمار X	نمرات روش تحقیق Y	X^2	Y^2	Xy
1	2	1	4	2
5	3	25	9	15
4	4	16	16	16
3	5	9	25	15
2	1	4	1	2
$\sum x = 15$	$\sum y = 15$			$\sum xy = 50$

پاسخ سوال اول: فرمول همبستگی

$$rxy = \frac{n\sum xy - (\sum x)\,(\sum x)}{\sqrt{[n\ \sum x^2 - (\sum x)^2\ ]\ -\ [n\ \sum y^2 - (\sum y)^2\ ]}}$$

$$rxy = \frac{5(50) - (15)(15)}{\sqrt{[5(55)\ - (15)^2]\ [5\ (55)\ (15)^2]}} = \ 0/5$$

پاسخ سوال دوم:

$$bxy = \frac{n\sum xy - (\sum x)\ (\sum y)}{n\sum x^2 - (\sum x)^2} \qquad bxy = \frac{(5)(50)-(15)(15)}{(5)(55)-(15)^2} = 0/5$$

$$axy = \frac{\sum y - (bxy)(\sum x)}{n} \qquad axy = \frac{(15)-\ (0/5)(15)}{5} = 1/5$$

$$y^{/} = a + bx \qquad y^{/} = 1/5 + 0/5x$$

پاسخ سوال سوم:

$$y^{/} = 1/5 + 0/5x \qquad y^{/} = 1/5 + 0/5\ (6) = \ 4/5$$

پاسخ سوال چهارم:

$$y^{/}5 = a + bx5\ =\ 1/5\ + (0/5)(2) =\ 2/5$$

$$e5 = y5 - y^{/}5 = 1 - 2/5 = -1/5$$

پاسخ سوال پنجم:

$$sxy =\ sy\sqrt{1 - (rxy)^2} = 1 - (0/5)^2 =$$

$$sy = \sqrt{\frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n}} = \sqrt{\frac{55 - \frac{(15)^2}{5}}{5}} = \sqrt{2} = 1/41$$

نمونه سوالات

۱. نمونه های زیر از امتحان ۴۰ نفر دانشجو در درس آمار بدست آمده است :

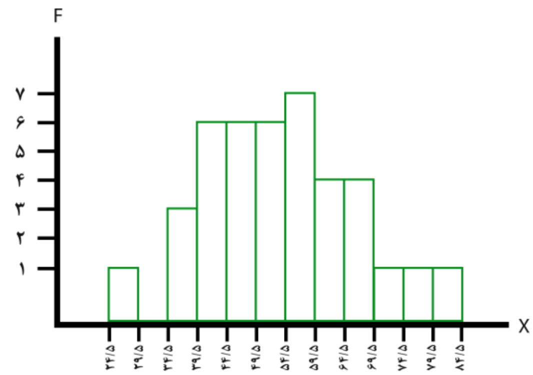
۶۱	۶۳	۶۰	۸۶	۸۴	۹۲	۸۱	۸۲	۸۰	۶۰
۷۶	۷۸	۹۴	۷۵	۷۷	۸۲	۶۷	۷۷	۷۷	۷۷
۷۸	۶۸	۷۰	۷۶	۶۰	۸۱	۸۶	۶۶	۸۷	۸۷
۷۰	۷۷	۷۹	۸۳	۸۷	۷۴	۹۲	۸۱	۸۳	۸۴

الف) جدول توزیع فراوانی همراه با ستون فراوانی تراکمی، نقطه میانی طبقات؛ فراوانی درصدی، فراوانی نسبی، فراوانی تراکمی نسبی، فراوانی تراکمی درصدی، درست کنید. فاصله طبقات را ۳ در نظر بگیرید.

X	F	$\acute{X}$	CF	$\frac{F}{N}$	$\frac{F}{N} \times 100 = F\%$	$\frac{cf}{N}$	$\frac{cf}{N} \times 100 = cf\%$
93-95	1	94	40	$\frac{1}{40}$	$\frac{1}{40} \times 100 = 2/5\%$	$\frac{40}{40}$	$\frac{40}{40} \times 100 = 100\%$
90-92	2	91	39	$\frac{2}{40}$	$\frac{2}{40} \times 100 = 5\%$	$\frac{39}{40}$	$\frac{39}{40} \times 100 = 97/5\%$
87-89	3	88	37	$\frac{3}{40}$	$\frac{3}{40} \times 100 = 7/5\%$	$\frac{37}{40}$	$\frac{37}{40} \times 100 = 92/5\%$
84-86	4	85	34	$\frac{4}{40}$	$\frac{4}{40} \times 100 = 10\%$	$\frac{34}{40}$	$\frac{34}{40} \times 100 = 85\%$
81-83	7	82	30	$\frac{7}{40}$	$\frac{7}{40} \times 100 = 17/5\%$	$\frac{30}{40}$	$\frac{30}{40} \times 100 = 75\%$
78-80	4	79	23	$\frac{4}{40}$	$\frac{4}{40} \times 100 = 10\%$	$\frac{23}{40}$	$\frac{23}{40} \times 100 = 57/5\%$
75-77	8	76	19	$\frac{8}{40}$	$\frac{8}{40} \times 100 = 20\%$	$\frac{19}{40}$	$\frac{19}{40} \times 100 = 47/5\%$
72-74	1	73	11	$\frac{1}{40}$	$\frac{1}{40} \times 100 = 2.5\%$	$\frac{11}{40}$	$\frac{11}{40} \times 100 = 27/5\%$
69-71	2	70	10	$\frac{2}{40}$	$\frac{2}{40} \times 100 = 5\%$	$\frac{10}{40}$	$\frac{10}{40} \times 100 = 25\%$
66-68	3	67	8	$\frac{3}{40}$	$\frac{3}{40} \times 100 = 7/5\%$	$\frac{8}{40}$	$\frac{8}{40} \times 100 = 20\%$
63-65	1	64	5	$\frac{1}{40}$	$\frac{1}{40} \times 100 = 2.5\%$	$\frac{5}{40}$	$\frac{5}{40} \times 100 = 12/5\%$
60-62	4	61	4	$\frac{4}{40}$	$\frac{4}{40} \times 100 = 10\%$	$\frac{4}{40}$	$\frac{4}{40} \times 100 = 10\%$
	$\sum f = 40 = n$						

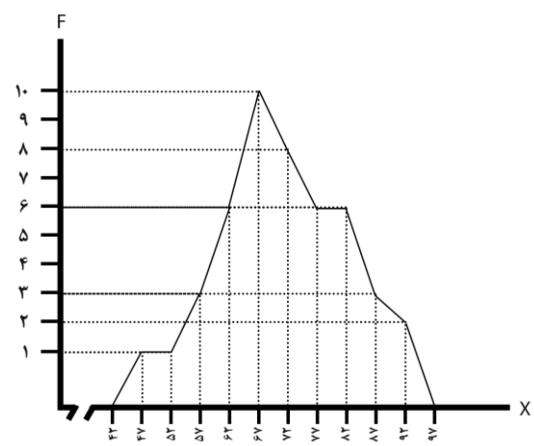
2- برای توزیع فراوانی زیر نمودار هیستوگرام رسم کنید.

X	F
80-84	1
75-79	1
70-74	1
65-69	4
60-64	4
55-59	7
50-54	6
45-49	6
40-44	6
35-39	3
30-34	0
25-29	1



3- برای توزیع فراوانی زیر نمودار چند ضلعی ترسیم کنید.

X	F	$\acute{X}$
90-94	2	92
85-89	3	87
80-84	6	82
75-79	6	77
70-74	8	72
65-69	10	67
60-64	6	62
55-59	3	57
50-54	1	52
45-49	1	47



4- میانه توزیع های زیر را محاسبه کنید.

1-7-5-3-2جواب3 =

8-9-10-11-11-11-11-12جواب10/75 =

3-4-5-6-4جواب4/25 =

1-2-4-3-6-3-7جواب3/25 =

1-2-3-5-7 =

8-9-10-10/5-10/75-11-11/25-11/5-12 =

3-3/5-4-4/5-5-6 =

1-2-2/5-3-3/5-4-6-7 =

۵. در یک توزیع فراوانی بزرگترین عدد ۴۹ و کوچکترین عدد ۵ است :

الف) دامنه تغییرات چقدر است : $45 = 49 - 5 + 1$ $R = X_m - X_l + 1$

ب) در صورتی که بخواهیم این توزیع را به ۱۵ طبقه تقسیم کنیم فاصله طبقات چقدر باید باشد:

$$\text{فاصله طبقات} = \frac{\text{دامنه تغییرات}}{\text{تعداد طبقات}} \qquad 3 = \frac{45}{15}$$

ج) اولین و پایین ترین طبقه را نوشته حدود واقعی و نقطه میانی این طبقه را مشخص کنید

اولین طبقه ۵-۳ حدود واقعی ۵/۵ – ۲/۵ نقطه میانی ۴

6- تعیین کنید نمودار کدام یک از ویژگی های زیر متقارن، دارای کجی منفی یا کجی مثبت است :

الف) توزیع سن اعضای هیات علمی دانشکده ای با تعداد زیادی استاد جوان و چند استاد در حد متوسط ومسن.

کجی مثبت

ب) توزیع بهره هوشی دانش آموزان یک کلاس در صورتی که بهره هوشی نیمی از کلاس در حد متوسط، $\frac{1}{4}$ پایین تر از متوسط و $\frac{1}{4}$ بالاتر از متوسط است.

متقارن

ج) توزیع نمره های داوطلبان ورود به دانشگاه در آزمون هوش و استعداد تحصیلی در شرایطی که اکثر داوطلبان نمره بالا بدست آورده باشند و بقیه نمره ها در حد متوسط و ضعیف باشد.

کجی منفی

7- میانگین توزیع نمره های زیر را با استفاده از روش های مختلف حساب کنید.

X	F	$\acute{X}$	F$\acute{X}$	$\hat{X}$	F$\hat{X}$
55-59	3	57	171	4	12
50-54	4	52	208	3	12
45-49	5	47	235	2	10
40-44	7	42	294	1	7
35-39	5	37	185	0	0
30-34	4	32	128	-1	-4
25-29	3	27	81	-2	-6
20-24	2	22	44	-3	-6

	$\sum f=33=n$		$\sum F\acute{X}=1346$		$\sum F\hat{X}=25$

روش اول : $\bar{X}=\frac{\sum f\acute{x}}{n}$

$$40/78=\frac{1346}{33}$$

روش دوم : $\bar{X}=\acute{x}+\frac{\sum f\hat{x}}{n}(i)$

$$40/78=37+\frac{25}{33}(5)$$

8- میانگین نمره های یک کلاس 200 نفره در یک آزمایش 60 و میانگین یک کلاس 30 نفره در همین آزمایش 45 است میانگین مرکب در دو کلاس چقدر است.

حجم نمونه های نامساوی :

$$\hat{X}T=\frac{(X1*n1)+(X2*n2)+(X3*n3)}{n1+n2+n3+n4+n5\ldots}$$

$$58/04=\frac{(60\times200)+(30\times45)}{200+30}$$

n1=200 X1=60

n2=45 X2=30

9- در توزیع های : 8-5-6-7-8-9-9-10-4-4

الف) میانگین را محاسبه کنید :

$\bar{X}=7$ $7=\frac{4+4+10+9+9+8+7+6+5+8}{10}$ $\bar{X}=\frac{\sum Xi}{n}$

ب) در صورتی که به هر یک از نمره های این توزیع 5 نمره اضافه کنیم میانگین توزیع چقدر می شود.

9-9-15-14-14-13-12-11-10-13

$\bar{X}=12$ $12=\frac{9+9+15+14+14+13+12+11+10+13}{10}$ $\bar{X}=\frac{\sum Xi}{n}$

10- توزیع نمره های هر یک از گروه های زیر در دست است :

گروه 1 = 5-4-3-4

گروه2= 6-4-2-2-1

گروه 3= 5-3-2-1-1-0

الف) میانگین هر یک از گروه ها را محاسبه کنید :

گروه 1 = $\bar{X} = 4$ $\quad 4 = \frac{5+4+3+4}{4}$ $\quad \bar{X} = \frac{\sum Xi}{n}$

گروه 2 = $\bar{X} = 3$ $\quad 3 = \frac{6+4+2+2+1}{5}$ $\quad \bar{X} = \frac{\sum Xi}{n}$

گروه 3= $\bar{X} = 2$ $\quad 4 = \frac{5+3+2+1+1+0}{6}$ $\quad \bar{X} = \frac{\sum Xi}{n}$

الف) میانگین مرکب هر یک از گروه ها را محاسبه کنید :

$$\hat{X}T = \frac{(X1 * n1) + (X2 * n2) + (X3 * n3)}{n1 + n2 + n3 + n4 + n5 \ldots}$$

$$2/86 = \frac{(4 * 4) + (3 * 5) + (6 * 2)}{4 + 5 + 6}$$

11- میانگین 4 گروه عبارتست از : 40-72-73-75

در صورتی که حجم هر یک از گروه ها برابر 25 باشد میانگین مرکب را حساب کنید

$\hat{X}T = \frac{X1+X2+X3+\cdots}{N}$ $\qquad 65 = \frac{75+73+72+40}{4}$