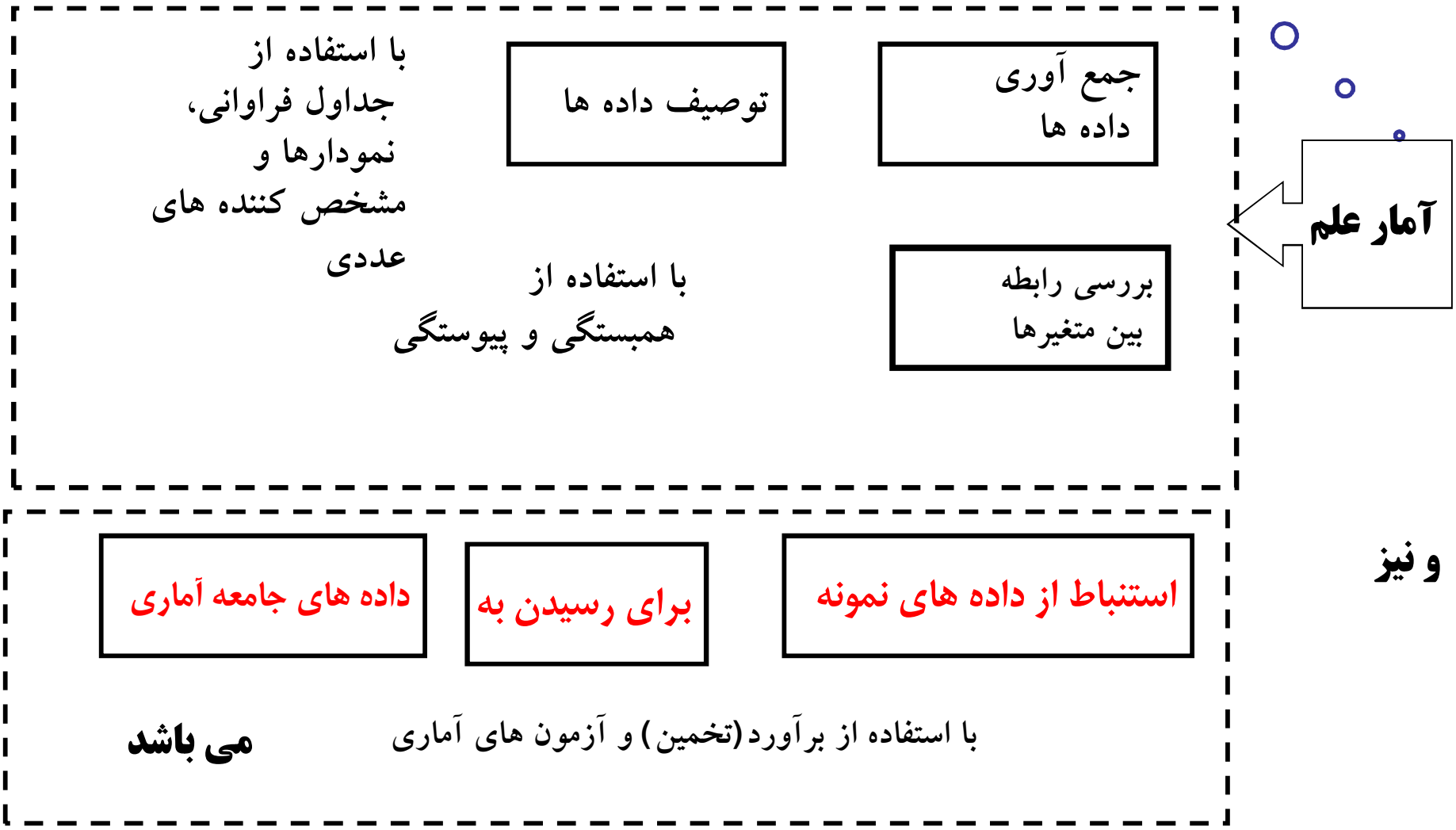
تعریف علم آمار



آمار توصیفی : روش هایی است که برای سازمان دادن ، خلاصه کردن و توصیف مشاهده ها به کار می رود و موضوع آن بیان دقیق ، کامل و نظام دار داده های تجربی و نتایج عینی پژوهش است .

آمار استنباطی : برای استنتاج کلی بر پایه ی احتمالات از طریق اطلاعات یک گروه نمونه ، از آمار استنباطی استفاده می شود و موضوع آن ، تبیین نتایج توصیفی ، تفسیر و میزان اهمیت و اعتبار آن هاست .

به عبارت دیگر روش هایی است برای استنباط خصوصیات گروه بزرگ (جامعه) از طریق مطالعه ی ویژگی های گروه های کوچکتر (نمونه) .

مقیاسهای اندازه گیری

اسمی :
ویژگی‌ها صرفاً در مقوله‌ها رده‌بندی می‌شوند بی‌آنکه هیچ رابطه ریاضی بین مقوله‌ها ضرورت داشته باشند.

رتبه ای :
در مقیاس رتبه ای نه تنها تفاوت کیفی متغیر ها مشخص می شود.
افراد مورد مطالعه از نظر صفت مورد نظر، از بیشترین تا کمترین مقدار درجه بندی و مرتبه هر فرد نسبت به دیگران مشخص می شود.

فاصله ای :
نه تنها افراد از نظر صفت مورد مطالعه طبقه بندی می شوند و رتبه هر فرد تعیین می شود بلکه تفاوت هر فرد از فرد دیگر را نیز می توان تعیین کرد.

نسبتی :
خصوصیت مقیاس نسبی داشتن نقطه ای دقیق برای شروع است که آن را صفر مطلق می نامیم

فراوانی - اندازه

به ارقامی که از طریق شمارش داده ها به دست می آید ، **فراوانی** گفته می شود .

مثال : مشخص کردن تعداد مراجعه کنندگان به یک مرکز درمانی به تفکیک ماه های سال .

اندازه ، ارقامی است که بیشی یا کمی یک چیز یا حالت را بر حسب یک مقیاس اندازه گیری نشان می دهد .

مثال : سن ، قد ، وزن ، هوشبهر و ...

علائم مورد استفاده

X ⟵ متغیر اول ، مرکز طبقه یا دسته

f ⟵ فراوانی: به تعداد دفعات وقوع یک متغیر گفته می شود .

$\sum$ ⟵ مجموع

N ⟵ تعداد داده ها ، تعداد نفرات ، تعداد نمره ها

توزیع فراوانی : فهرستی از تعدادی داده که صعودی یا نزولی مرتب شده و فراوانی آنها نیز مشخص شده باشد

فراوانی - اندازه

به ارقامی که از طریق شمارش داده ها به دست می آید ، فراوانی گفته می شود .

مثال : مشخص کردن تعداد مراجعه کنندگان به یک مرکز درمانی به تفکیک ماه های سال .

اندازه ، ارقامی است که بیشی یا کمی یک چیز یا حالت را بر حسب یک مقیاس اندازه گیری نشان می دهد .

مثال : سن ، قد ، وزن ، هوشبهر و ...

فراوانی : به تعداد دفعات وقوع یک متغیر گفته می شود .

توزیع فراوانی : فهرستی از تعدادی داده که صعودی یا نزولی مرتب شده و فراوانی آنها نیز مشخص شده باشد .

توزیع فراوانی طبقه بندی نشده

14 ، 15 ، 13 ، 9 ، 15 ، 12 ،
13 ، 10 ، 14 ، 12 ، 17 ، 11 ،
12 ، 14 ، 11 ، 13 ، 16 ، 15 ،
13 ، 11 ، 16 ، 12 ، 14 ، 17 ،
14 ، 10 ، 13 ، 15 ، 16 ، 14

$N = 30 = \sum f$

X	f
17	2
16	3
15	4
14	6
13	5
12	4
11	3
10	2
9	1
$\sum$	30

بازگشت

دامنه ی تغییر نمرات

عبارت است از تفاضل بزرگترین و کوچکترین عدد به علاوه ی یک .

دامنه ی تغییر = بزرگترین داده - کوچکترین داده + 1

$$R = X_{\max} - X_{\min} + 1$$

$$R = Scor_{\max} - Scor_{\min} + 1$$

بازگشت

فاصله ی طبقاتی (اندازه هر طبقه)

به پهنای دسته های ایجاد شده یا به عبارت دیگر به اندازه ی هر دسته یا طبقه گفته می شود .

$$\text{فاصله ی طبقاتی} = \frac{\text{دامنه ی تغییر}}{\text{تعداد طبقات}} \longleftarrow i$$

تعداد طبقات مناسب	فاصله ی طبقاتی مناسب
بین 10 تا 25 طبقه	1 ، 2 ، 3 ، 5 ، 10 ضرایب

جدول توزیع فراوانی طبقه بندی شده

طبقات	خط نشان T	f
37 - 41	////	4
32 - 36	⧸⧸⧸⧸ /	6
27 - 31	⧸⧸⧸⧸ ////	9
22 - 26	⧸⧸⧸⧸ ⧸⧸⧸⧸ /	11
17 - 21	⧸⧸⧸⧸ //	7
12 - 16	///	3
		40

بازگشت

25 ، 34 ، 15 ، 40 ، 29 ، 36 ،
24 ، 12 ، 30 ، 38 ، 31 ، 28 ،
18 ، 29 ، 23 ، 17 ، 39 ، 24 ،
20 ، 38 ، 27 ، 16 ، 18 ، 32 ،
28 ، 26 ، 22 ، 27 ، 33 ، 22 ،
23 ، 32 ، 21 ، 26 ، 35 ، 30 ،
23 ، 19 ، 25 ، 20

$N = 40 =$ f

$R = 40 - 12 + 1 = 29$

$i = 29 \div 6 = 4,83$

$i = 5$

حدود واقعی طبقات - مرکز طبقه (دسته)

L_l	طبقات	L_u	X
36,5	37 - 41	41,5	39
31,5	32 - 36	36,5	34
26,5	27 - 31	31,5	29
21,5	22 - 26	26,5	24
16,5	17 - 21	21,5	19
11,5	12 - 16	16,5	14

L_l = حد واقعی پایین L_u = حد واقعی بالا

X = مرکز طبقه

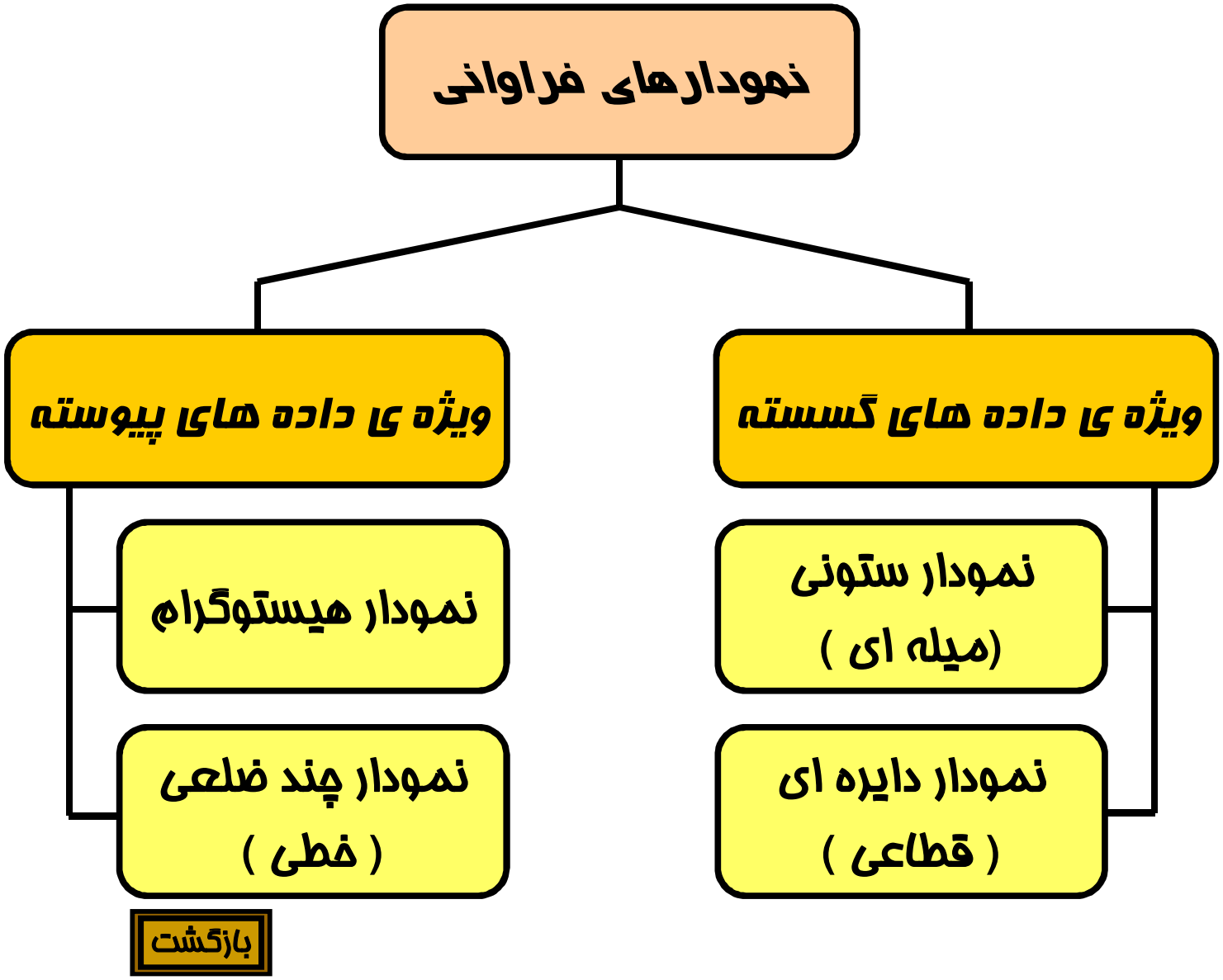
بازگشت

درصد فراوانی – فراوانی تراکمی

طبقات	نقطه میانی X	فراوانی مطلق f	درصد فراوانی P	فراوانی تراکمی cf	درصد فراوانی تراکمی P.cf
37- 41	39	4	10	40	100
32- 36	34	6	15	36	90
27- 31	29	9	22/5	30	75
22- 26	24	11	27/5	21	52/5
17- 21	19	7	17/5	10	25
12- 16	14	3	7/5	3	7/5
		N =40			

$$p = \frac{f}{N} \times 100$$

$$p.cf = \frac{Cf}{N} \times 100$$


نمودارهای فراوانی
ویژه ی داده های گسسته
ویژه ی داده های پیوسته
نمودار ستونی
(میله ای)
نمودار دایره ای
(قطاعی)
نمودار هیستوگرام
نمودار چند ضلعی
(خطی)
بازگشت

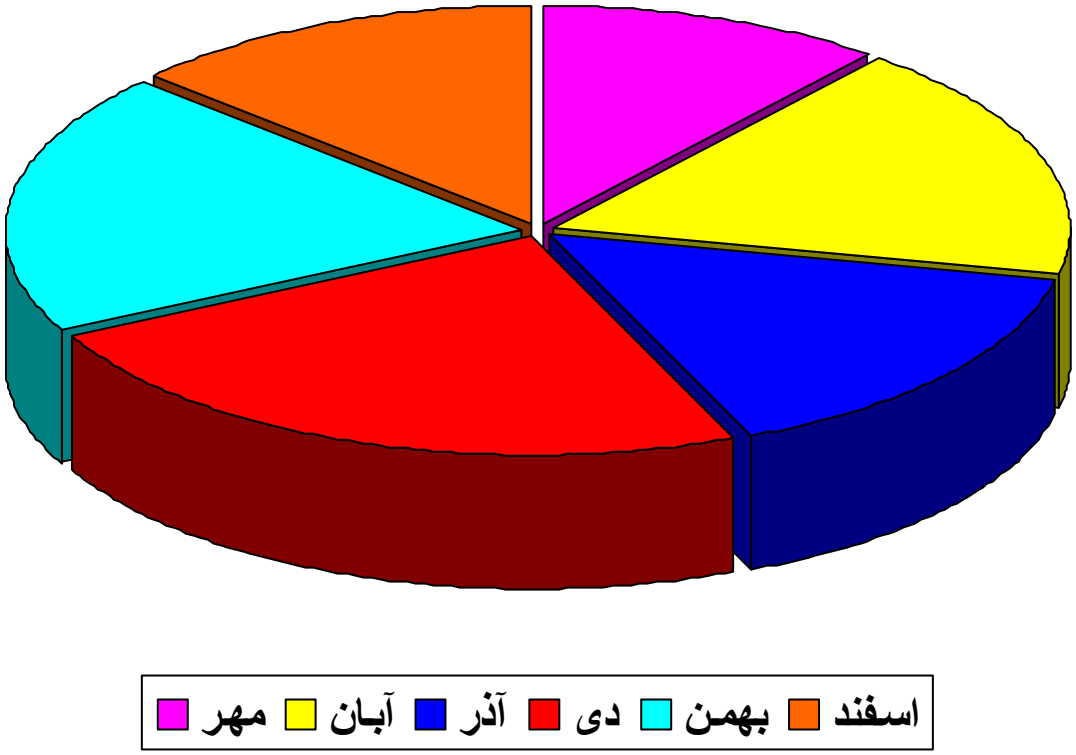
نمودار میله ای

ماه ها	f
مهر	25
آبان	40
آذر	35
دی	55
بهمن	45
اسفند	30
$\sum$	230

بازگشت

نمودار دایره

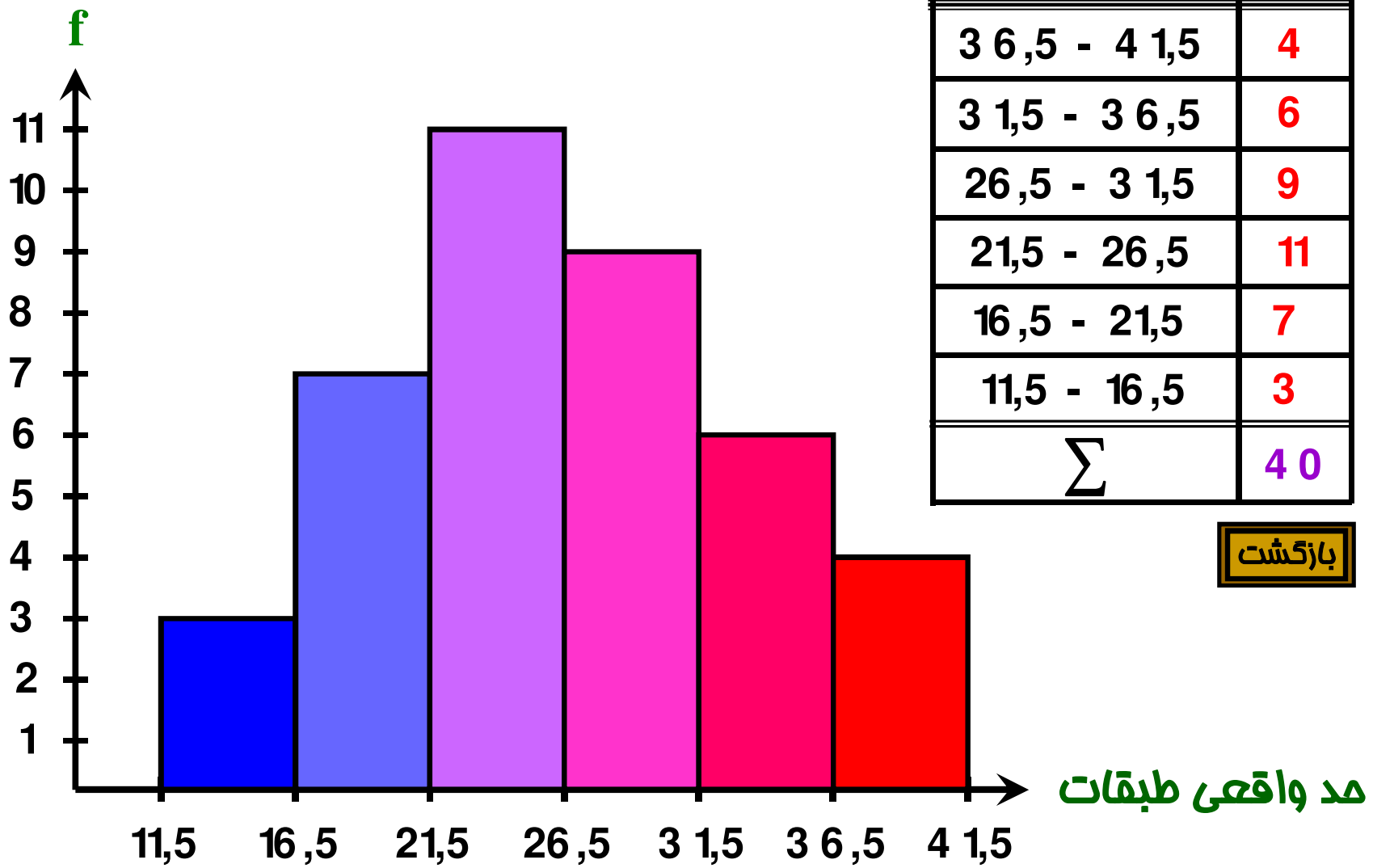
ماه ها	f
مهر	25
آبان	40
آذر	35
دی	55
بهمن	45
اسفند	30
$\sum$	230



$$\text{درجه} = \frac{f}{N} \times 365$$

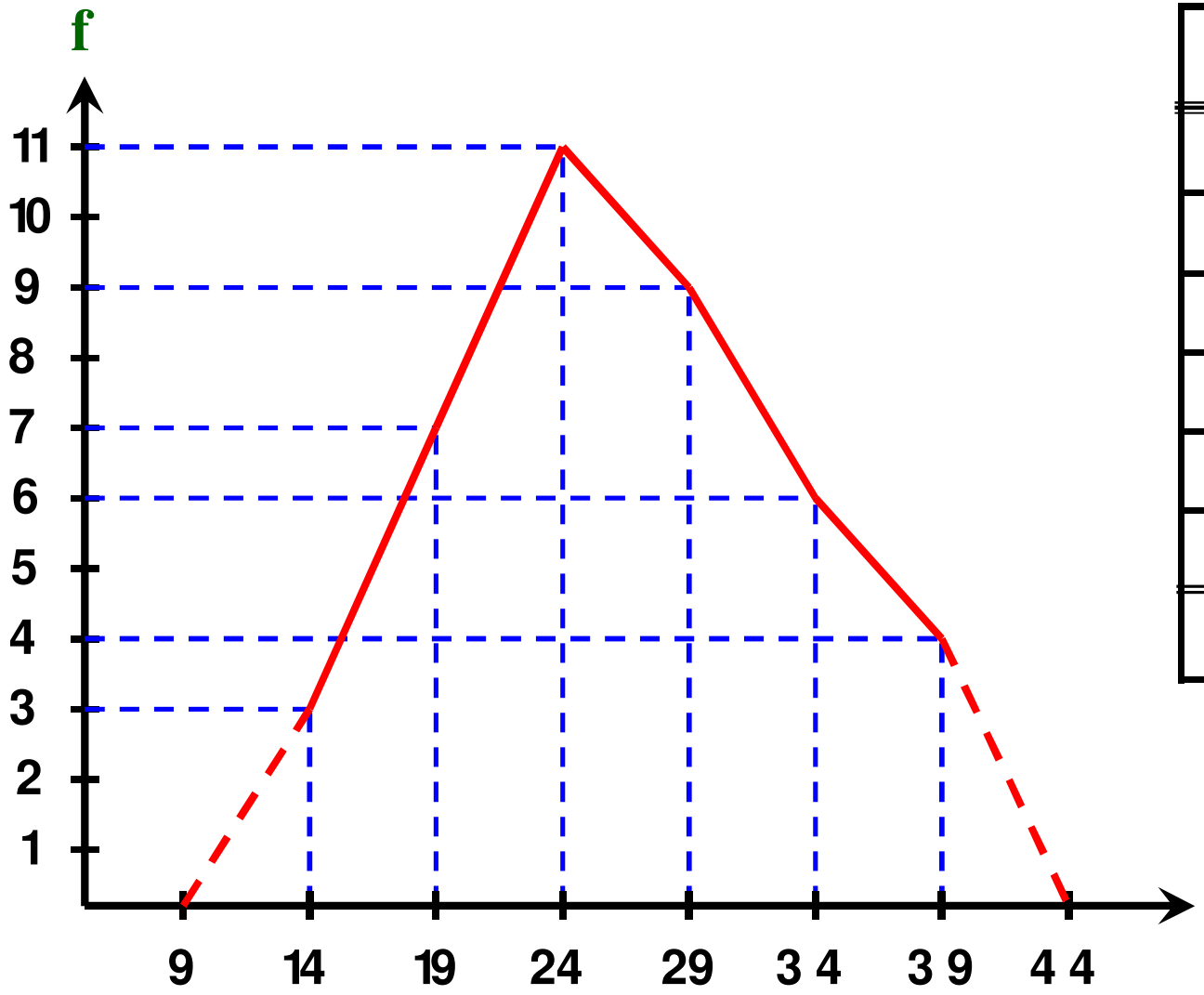
نمودار هیستوگرام

حد واقعی طبقات	f
36,5 - 41,5	4
31,5 - 36,5	6
26,5 - 31,5	9
21,5 - 26,5	11
16,5 - 21,5	7
11,5 - 16,5	3
$\sum$	40



بازگشت

نمودار چند ضلعی



X	f
3 9	4
3 4	6
29	9
24	11
19	7
14	3
∑	4 0

بازگشت

مقادیر متوسط (اندازه های گرایش به مرکز)
میانگین - میانه - نما (مد)

مقدار متوسط یا شاخص گرایش به مرکز:

نمره ای است که اگر آن را به جای نمره های اصلی به همه ی آزمودنی ها بدهیم ، مقدار خطای محاسبه کمینه گردد .

مقاصدی که این شاخص می تواند به آن کمک کند :

1 - توصیفی سریع از یک جامعه یا تعداد زیادی داده ی کمی .

2 - توصیف غیرمستقیم جامعه ای که گروه از آن بیرون آمده .

3 - مقایسه ی گروه های مختلف با هم .

نما (مد)

داده ای است که در توزیع داده ها بیشترین فراوانی را دارد .

نما (مد) ⟵ Mo

13 ، 7 ، 12 ، 10 ، 13 ، 8 ، 12 ، 13 ، 9 ، 14 ، 13 ، 10

در صورت وجود یک نما در بین توزیع :

نما ⟵ 13

12 ، 7 ، 13 ، 10 ، 13 ، 8 ، 12 ، 13 ، 9 ، 10 ، 12 ، 14

در صورت وجود دو نما متوالی در بین توزیع :

نما ⟵ 12,5 = 2 ÷ 25 　　12 + 13 = 25

14 ، 10 ، 12 ، 7 ، 14 ، 8 ، 12 ، 13 ، 9 ، 14 ، 10 ، 12

در صورت وجود دو نما در بین توزیع (عدم توالی) :

دو نمایی ⟵ نما ⟵ 14 و 12

نما را وقتی محاسبه می کنیم که :

1 - برآورد آسان و سریعی از اندازه های مرکزی توزیع لازم باشد .

2 - برآورد تقریبی از اندازه های مرکزی توزیع کافی باشد .

3 - بخواهیم بدانیم کدام نمره بیشتر تکرار شده است .

4 – مقیاس داده ها اسمی باشد .

5 - برای داده های کم مناسب نیست و به سرعت تغییر می کند

6 - در صورتی که نما در ابتدا یا انتها توزیع عددی باشد شاخص مرکزی مناسبی نیست

تعریف میانه

داده ای است که داده های مرتب شده را (صعودی یا نزولی) از نظر تعداد به دو قسمت مساوی تقسیم کند .

به عبارت دیگر ارزش عددی نقطه ای که 50 درصد بالای نمره ها را از 50 درصد پایین آن جدا می کند .

$$\text{محل قرار گرفتن میانه} = \frac{N+1}{2}$$

میانه $\longrightarrow$ Md

بازگشت

محاسبه ی میانه ی اعداد گروه بندی نشده

9 ، 12 ، 10 ، 9 ، 8 ، 13 ، 11 ، 9 ، 12

میانه

8 ، 9 ، 9 ، 9 ، 10 ، 11 ، 12 ، 12 ، 13

1 2 3 4 5 4 3 2 1

$\frac{9+1}{2} = 5$

18 ، 15 ، 13 ، 18 ، 17 ، 14 ، 18 ، 19 ، 16 ، 15 ، 18 ، 15

16,5 ← میانه

13 ، 14 ، 15 ، 15 ، 15 ، 16 ، 17 ، 18 ، 18 ، 18 ، 18 ، 19

1 2 3 4 5 6 6 5 4 3 2 1

$\frac{12+1}{2} = 6,5$

16 + 17 = 33

33 ÷ 2 = 16,5

بازگشت

محاسبه ی میانه ی اعداد گروه بندی شده

طبقات	f
37 - 41	4
32 - 36	6
27 - 31	9
22 - 26	11
17 - 21	7
12 - 16	3
∑	40

طبقه ی میانه

$$Md = L + \frac{\frac{N}{2} - cf}{f} \times i$$

$$Md = 21{,}5 + \frac{\frac{40}{2} - 10}{11} \times 5$$

$$Md = 21{,}5 + 5 \times 0{,}91$$

$$Md = 26{,}05$$

میانه را وقتی محاسبه می کنیم که :

1 - وقت کافی برای محاسبه ی میانگین نداشته باشیم .

2 - توزیع دارای چولگی قابل ملاحظه باشد .

3 - توزیع نمره ها باز یا ناقص باشد .

4 - بخواهیم بدانیم که نتایج اندازه گیری در نیمه ی بالای توزیع یا در نیمه ی پایین آن قرار می گیرد .

5 - اندازه گیری مقادیر انتهایی توزیع دقیق نباشد .

6 - مقیاس اندازه گیری رتبه ای باشد

بازگشت

تعریف میانگین

به متوسط عددی همه ی نمرات توزیع **میانگین** گفته می شود .

میانگین ، با ثبات ترین شاخص مرکزی است که با داده های فاصله ای و نسبی بکار می رود و مرکز ثقل داده هاست .

μ یا $\bar{X}$ ⟵ میانگین (میانگین حسابی)

محاسبه ی میانگین داده های طبقه بندی نشده

7 ، 11 ، 13 ، 16 ، 17 ، 20

$$\overline{X} = \frac{\sum X}{N}$$

$$\overline{X} = \frac{7 + 11 + 13 + 16 + 17 + 20}{6} = \frac{84}{6} = 14$$

8 ، 8 ، 8 ، 8 ، 11 ، 11 ، 15 ، 15 ، 15

$$\overline{X} = \frac{\sum X}{N} = \frac{4 \times 8 + 2 \times 11 + 3 \times 15}{4 + 2 + 3} = \frac{99}{9} = 11$$

$$\overline{X} = \frac{\sum f X}{N}$$

بازگشت

محاسبه ی میانگین در داده های طبه بندی شده

طبقات	f	X	f X
37 - 41	4	39	156
32 - 36	6	34	204
27 - 31	9	29	261
22 - 26	11	24	264
17 - 21	7	19	133
12 - 16	3	14	42
$\sum$	40	-	1060

بازگشت

$$\overline{X} = \frac{\sum f X}{N}$$

$$N = \sum f = 40$$

$$\overline{X} = \frac{1060}{40}$$

$$\overline{X} = 26,5$$

محاسبه ی میانگین مرکب یا میانگین میانگینها (وزنی)

10 ، 10 ، 12 ، 12 ، 12 ، 12 ، 12 ، 16 ، 16 ، 16

$X = 10$ $X = 12$ $X = 16$

$W = 2$ $W = 5$ $W = 3$

$$\overline{X} = \frac{\sum X}{N} = \frac{2 \times 10 + 5 \times 12 + 3 \times 16}{2 + 5 + 3} = \frac{128}{10} = 12{,}8$$

$$\overline{X} = \frac{\sum WX}{\sum W}$$

بازگشت

ویژگی های میانگین

1 - میانگین تنها مقدار متوسطی است که از ضرب آن در تعداد دفعات (N) ، حاصل جمع مقادیر به دست می آید .

2 - اگر همه ی داده ها را با عدد ثابتی جمع یا از همه ی آن ها ، مقدار ثابتی را کم کنیم ، میانگین نیز همان مقدار اضافه یا کم می شود .

3 - اگر همه ی داده ها را در عدد ثابتی ضرب یا بر آن عدد تقسیم کنیم ، میانگین نیز در همان عدد ضرب یا بر همان عدد تقسیم می شود .

4 - همیشه جمع جبری نمرات انحراف از میانگین برابر صفر است .

5 - همیشه مجموع مجذور نمرات انحراف از میانگین کمترین مقدار است .

6 - میانگین نسبت به تک تک نمرات توزیع حساسیت دارد و تغیر هر نمره باعث تغیر میانگین می کردد .

شاخص های پراکندگی

شاخص های پراکندگی میزان پراکنده بودن نمرات حول و حوش مرکز داده ها را نشان می دهد

دامنه تغییرات R

$$R = X_{\max} - X_{\min} + 1$$

انحراف چارکی

$$Q = \frac{Q_3 - Q_1}{2}$$

واریانس و انحراف استاندارد

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

انحراف چارکی

به نصف دامنه ی تغییر بین دو چارک گفته می شود . $Q = \frac{Q_3 - Q_1}{2}$

از این شاخص زمانی استفاده می شود که میانه را بر میانگین به عنوان شاخص مقدار مرکزی ترجیح دهیم .

محاسبه انحراف چارکی اعداد طبقه بندی نشده :

1 - اعداد را از کوچک به بزرگ مرتب کنید.

2 - میانه را حساب کنید

3 - میانه اعداد سمت راست و چپ میانه را محاسبه کنید .

6 - 8 - 9- 11- 12- 13 - 14 - 17

Q1=8/5 Q2=11/5 Q3=13/5

واریانس و انحراف استاندارد

واریانس :

نوعی مقدار متوسط و متعلق به خانواده ی میانگین هاست که از روی انحراف نمره ها از میانگین محاسبه می شود .

برابر با میانگین مجذور انحرافات از میانگین داده هاست و با واحد هایی معادل با مجذور واحد های داده های اصلی بیان می شود .

انحراف استاندارد :

به جذر واریانس یا به عبارت دیگر ، جذر میانگین مجذور انحرافات از میانگین ، در اصطلاح آماری انحراف استاندارد گویند .

انحراف استاندارد در حقیقت مهم ترین و معتبرترین مشخص کننده ی پراکندگی داده هاست . هر قدر بزرگتر باشد تغییر پذیری اعداد زیادتر و هر چه کوچکتر باشد تغییر پذیری آن ها کمتر است

محاسبه واریانس و انحراف استاندارد اعداد گروه بندی نشده با استفاده از نمرات انحراف

نمرات انحراف

مجذور نمرات انحراف

	X	$X - \bar{X}$	x	x^2
	24	24 - 12	+ 12	144
	15	15 - 12	+ 3	9
	12	12 - 12	0	0
	6	6 - 12	- 6	36
	3	3 - 12	- 9	81
$\sum$	60	-	0	270

مجموع مجذور نمرات انحراف

$$\bar{X} = \frac{\sum X}{N} \qquad \bar{X} = \frac{60}{5} = 12$$

$$\sum (X - \bar{X})^2 = 270$$

$$S^2 = \frac{270}{5-1} = 67/5$$

$$S^2 = \frac{\sum (X - \bar{X})^2}{N-1}$$

$$S = \sqrt{67/5} = 8/21$$

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}}$$

محاسبه ی انحراف استاندارد اعداد گروه بندی نشده با استفاده از داده های اصلی

	X	X^2
	24	576
	15	225
	12	144
	6	36
	3	9
∑	60	990

بازگشت

$$\overline{X} = \frac{\sum X}{N} \qquad \overline{X} = \frac{60}{5} = 12$$

$$S^2 = \frac{\sum (X - \overline{X})^2}{N} = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}$$

$$S^2 = \frac{990}{5} - (12)^2 \qquad S^2 = 198 - 144$$

$$S^2 = 54 \qquad S = \sqrt{54} = 7,35$$

$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{N}} = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}}$$

محاسن انحراف استاندارد

1 - مقدار آن در نمونه های مختلفی که از یک جامعه بیرون می آید ، نسبتاً پایدار است .

2 - مانند میانگین حسابی تحت تأثیر همه ی نمره ها قرار می گیرد و از آن می توان در فرمول ها و روابط جبری استفاده کرد .

3 - با استفاده از آن می توان مشخص کرد چه نسبتی از نمره ها در فاصله های مختلف نسبت به میانگین قرار گرفته است .

دو نکته در رابطه با محاسبه ی واریانس و انحراف استاندارد با استفاده از متغیر فرضی

1 - افزایش یا کاهش یک عدد ثابت به همه ی نمره های یک مجموعه ، تأثیری در واریانس و انحراف استاندارد نمره ها ندارد . (زیرا انحراف از میانگین نمره ها که پایه ی محاسبات این دو شاخص است ثابت می ماند .)

2 - چنانچه همه ی نمره ها را در عدد ثابتی ضرب کنیم ، واریانس نمره ها در مجذور آن عدد و انحراف استاندارد نمره ها در خود آن عدد ثابت ضرب می شود . (این قاعده در مورد تقسیم نیز صادق است .)

در صورتی که داده ها از نمونه ای از یک جامعه بزرگتر جمع آوری شده باشد و قصد برآورد شاخص پراکندگی جامعه را داشته باشیم در مخرج به جای n از n-1 استفاده می شود

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}(i^2)$$

- خطای طبقه بندی به دلیل اینکه انحراف ها مجذور می شوند بیشتر می شود ، هر چقدر فاصله طبقات بیشترباشد خطای بیشتری صورت می گیرد.
بنابراین زمانی که داده طبقه بندی شده و تعداد طبقه ها کمتر ۱۲ است به کمک فرمول زیر انحراف استاندارد اصلاح می شود

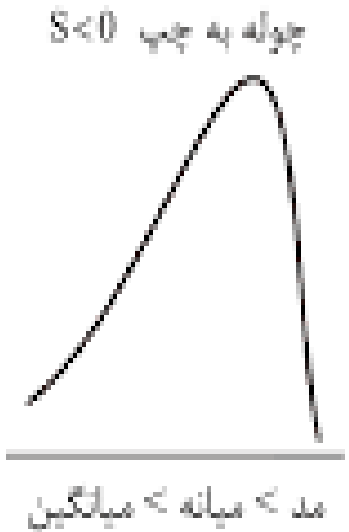
s= واریانس i= فاصله طبقات

تصحیح شپرد:

$$s_c = \sqrt{s^2 - \frac{i^2}{12}}$$

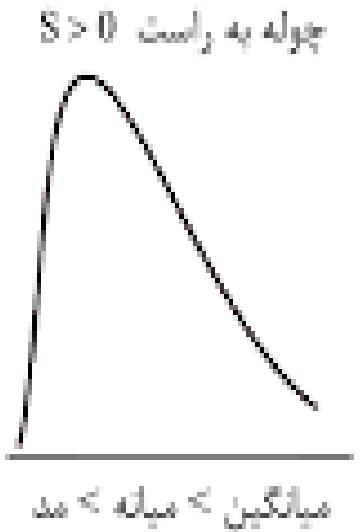
کجی مثبت و منفی

در مورد نمودارهای پیوسته تک‌مدی معمولاً یکی از سه حالت زیر اتفاق می‌افتد:



$Q_3 - Q_2 < Q_2 - Q_1$

$Q_3 - Q_2 = Q_2 - Q_1$



$Q_3 - Q_2 > Q_2 - Q_1$

$$g = \frac{\bar{x} - \text{mod}}{S}$$

هر چقدر میزان کجی به صفر نزدیکتر باشد(قدر مطلق آن کوچک باشد) .توزیع نرمال (طبیعی) است یعنی نمونه بهتر انتخاب و بررسی شده است

ضریب همبستگی

شاخصی است که جهت و مقدار رابطه ی بین دو متغیر را توصیف می کند .

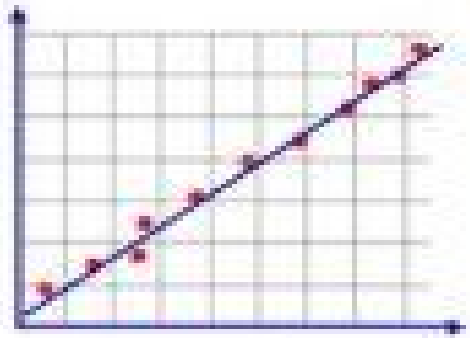
این وابستگی به معنای رابطه علت و معلولی نیست وضریب همبستگی حرفی از اینکه کدام علت و کدام معلول است به میان نمی آورد

مثال : رابطه بین هوش و پیشرفت تحصیلی

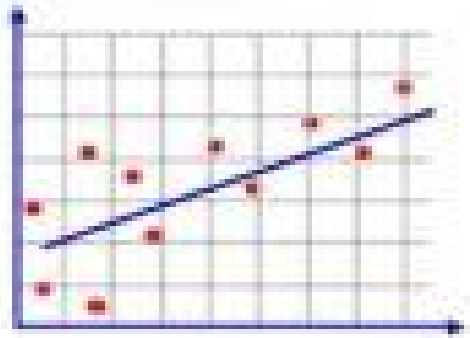
رابطه بین اضطراب و پیشرفت تحصیلی

- مستقیم و معکوس
- منفی و مثبت
- کامل و ناقص
- خطی و غیر خطی
- شدت و جهت

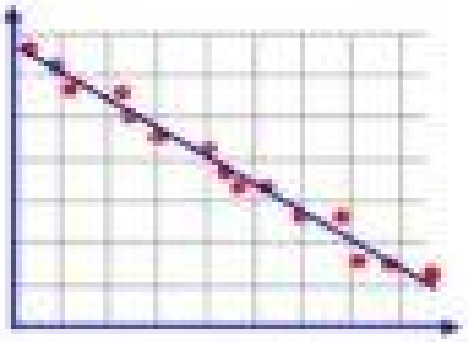
همبستگی و نمودارهای پراکندگی



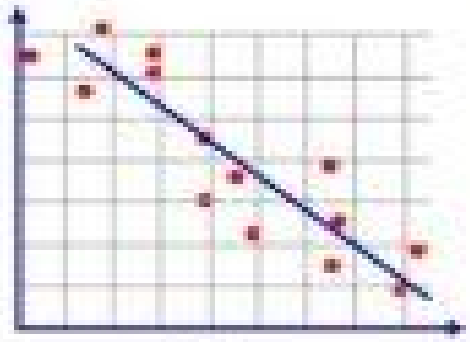
همبستگی مثبت قوی



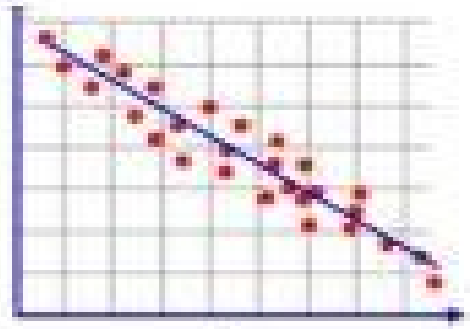
همبستگی مثبت ضعیف



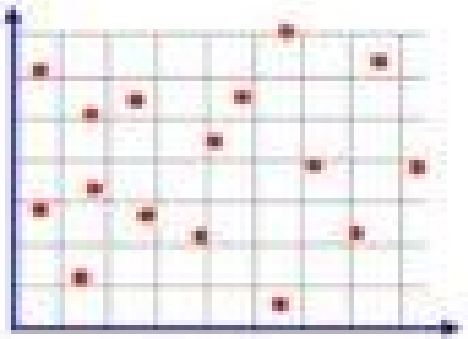
همبستگی منفی قوی



همبستگی منفی ضعیف



همبستگی منفی متوسط



همبستگی صفر

جهت همبستگی

همبستگی مثبت و کامل 1+ نشان دهنده رابطه مستقیم است (چنانچه نمره فردی در متغیر یک بالا باشد نمره او در متغیر دیگر نیز بالا خواهد بود . نمره Z در متغیر اول با نمره Z در متغیر دوم برابر است).

ضریب همبستگی منفی و کامل 1 - نشان دهنده رابطه معکوس است (چنانچه نمره فردی در متغیر یک بالا باشد نمره او در متغیر دیگر پایین است. بزرگترین نمره Z در یک متغیر با کوچکترین نمره Z در متغیر دیگر همراه خواهد بود.)

همبستگی خنثی و صفر 0 نشان دهنده عدم همبستگی می باشد

شدت همبستگی

- طور کلی ضرایب همبستگی دارای دامنه ای بین 1 - تا 1 است . هر چه ضریب همبستگی به یک نزدیکتر باشد میزان وابستگی دو متغیر بیشتر است، شدت همبستگی به وسیله قدر مطلق ضریب همبستگی مشخص می شود . مقدار یا شدت ضریب همبستگی 1+ برابر با ضریب همبستگی 1 - است ؛ تفاوت دو ضریب در جهت رابطه بین آنهاست

توجه : در صورتی که مقیاس اندازه گیری نسبتی و یا فاصله ای باشد از ضریب همیستگی گشتاوری پیرسون استفاده می شود

فرمول های محاسبه ی ضریب همبستگی گشتاوری پیرسون

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 \sum (Y - \overline{Y})^2}}$$

$$r = \frac{\sum x y}{\sqrt{\sum x^2 \sum y^2}}$$

$$r_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{\left(N \sum x^2 - (\sum x)^2 \right)\left(N \sum y^2 - (\sum y)^2 \right)}}$$

محاسبه ی ضریب همبستگی گشتاوری پیرسون با استفاده از نمرات انحراف

	X	Y	x	y	xy	x^2	y^2
	19	16	+7	+5	+35	49	25
	14	11	+2	0	0	4	0
	11	9	-1	-2	+2	1	4
	6	5	-6	-6	+36	36	36
	17	19	+5	+8	+40	25	64
	16	14	+4	+3	+12	16	9
	4	6	-8	-5	+40	64	25
	9	8	-3	-3	+9	9	9
$\sum$	96	88	0	0	174	204	172

بازگشت

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$$\overline{X} = \frac{\sum X}{N} \qquad \overline{X} = \frac{96}{8} = 12$$

$$\overline{Y} = \frac{\sum Y}{N} \qquad \overline{Y} = \frac{88}{8} = 11$$

$$r = \frac{174}{\sqrt{204 \times 172}}$$

$$r = \frac{174}{187,32} \qquad r = 0,93$$

روزهای غیبتX	میزان اضطرابy	x^2	y^2	xy
2	14	4	196	28
0	10	0	100	0
3	8	9	64	24
6	6	36	36	36
4	2	16	4	8
∑x=15	∑y=40	$\sum=x^2$ 65	$\sum=y^2$ 4 00	∑xy=96

$$r_{xy} = \frac{N \sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left(N \sum x^2 - \left(\sum x\right)^2\right)\left(N \sum y^2 - \left(\sum y\right)^2\right)}}$$

$$r_{xy} = \frac{N \sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left(N \sum x^2 - \left(\sum x\right)^2\right)\left(N \sum y^2 - \left(\sum y\right)^2\right)}}$$

$$r_{xy} = \frac{(5)(96) - (15)(40)}{\sqrt{\left((5)(65) - (15)^2\right)\left((5)(400) - (40)^2\right)}} = \frac{-120}{\sqrt{(325 - 225)(2000 - 1600)}}$$

$$r_{xy} = \frac{-120}{\sqrt{(100)(400)}} = \frac{-120}{\sqrt{40000}} = \frac{-120}{200} = -0/60$$

محاسبه ی ضریب همبستگی رتبه ای اسپیرمن

طرح	X	Y	D	D^2
الف	4	2	+2	4
ب	7	6	+1	1
ج	1	4	- 3	9
د	5	8	- 3	9
ه	2	3	- 1	1
و	8	5	+3	9
ز	3	1	+2	4
ح	6	7	- 1	1
$\sum$	-	-	0	3 8

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

$$r_s = 1 - \frac{6 \times 3\,8}{8 \times 6\,3}$$

$$r_s = 1 - \frac{228}{5\,04}$$

$$r_s = 1 - 0{,}4\,5$$

$$r_s = 0{,}5\,5$$

توجه : در صورتی که مقیاس اندازه گیری رتبه ای باشد از ضریب همیستگی اسپیرمن استفاده می شود

x	y	x_r	y_r	d	D^2
60	60	1	2	-1	1
54	68	2	1	1	1
53	40	3	11/5	8/5	72/25
49	52	4/5	3	1/5	2/25
49	51	4/5	4/5	0	0
47	38	6	14	-8	64
46	51	7	4/5	2/5	6/25
45	32	9	17	-8	64
45	39	9	13	-4	16
45	41	9	10	-1	1
43	50	11	6	5	25
41	48	12	7/5	4/5	20/25
39	36	13	16	-3	9
38	48	14	7/5	6/5	42/5
32	40	15/5	11/5	4	16
32	46	15/5	9	6/5	42/25
30	37	17	15	2	4

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

$$r_s = 1 - \frac{6 * 386/5}{17(17^2 - 1)}$$

$$r_s = 1 - \frac{2319}{4896} = 0/51$$

مفروضه های ضریب همبستگی پیرسون

1 - رابطه بین دو متغیر خطی باشد. پراکندگی آن بصورت خطی باشد (مشاهده نمودار)

2 - توزیع ها دارای شکل مشابه باشند (در صورت کجی هر دو در یک جهت باشد)

3 - نمودار پراکندگی یکسان باشد (عرض نقاط در سراسر نمودار یکسان باشد)

عوامل تأثیر گذار بر ضریب همبستگی :

1 - اساس رابطه از جامعه ای به جامعه دیگر فرق می کند

2 - پراکندگی متغیرها در جوامع مختلف متفاوت است(همبستگی در جوامع ناهمگون بیشتر است)

3 - همبستگی بین دو متغیر می تواند تحن تأثیر متغیر سوم قرار گیرد

تفسیر ضریب همبستگی :

1 - ضریب همبستگی جهت و شدت رابطه بین دو متغیر را نشان می دهد

2 - ضریب همبستگی را نمی توان بصورت درصد یا نسبت مقایسه و تفسیر کرد

3 - همبستگی به معنای رابطه علت و معلولی بین متغیرها نیست

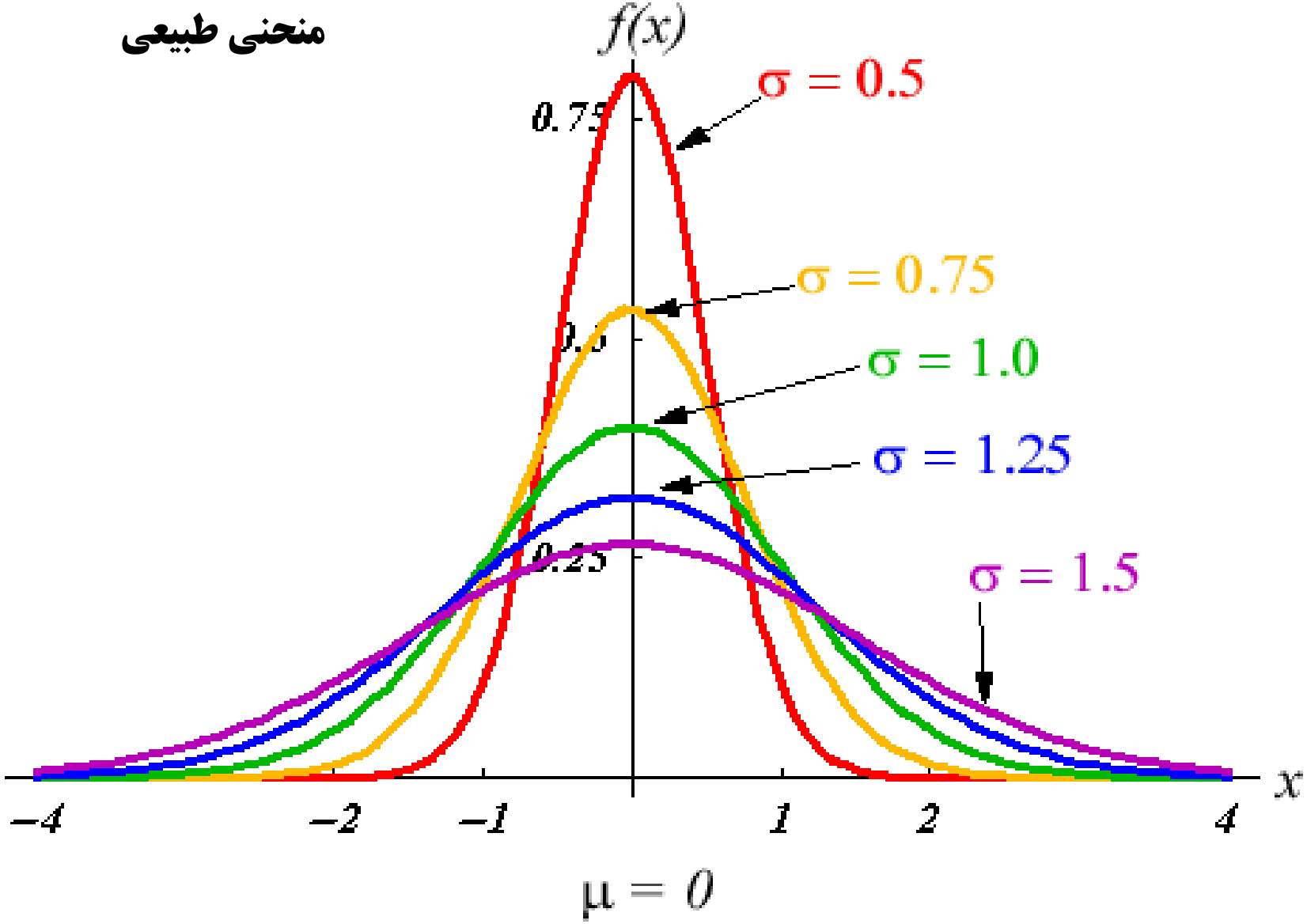
4 - همبستگی نشان می دهد که تغییر در یک متغیر هماهنگ با تغییر در متغیر دیگری هست یا خیر

- **ضریب تعیین**
- ضریب همبستگی ، اندازه همبستگی بین دو متغیر را نشان می دهد . این شاخص در باره ماهیت این ارتباط چیزی را نشان نمی دهد

$$r^2 \times 100$$

- ضریب تعیین اطلاعات کاملتری در اختیار قرار می دهد
- این ضریب نشان می دهد چند درصد از کل واریانس x ناشی از واریانس y است

منحنی طبیعی

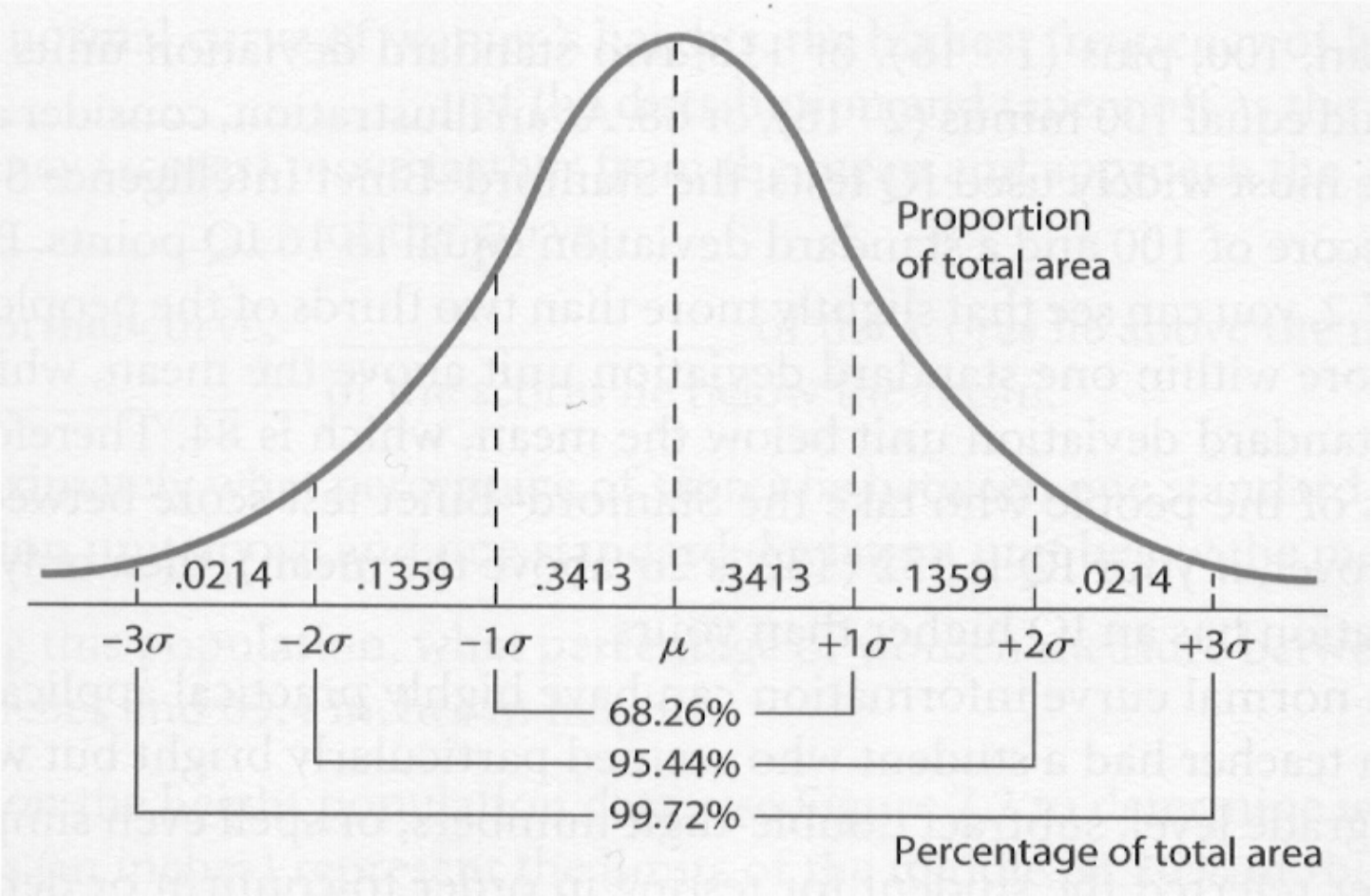


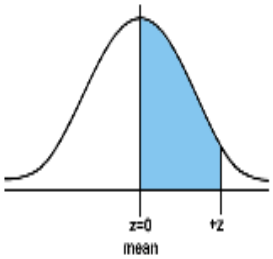
شکل توزیع طبیعی به میانگین وانحراف استاندارد بستگی دارد

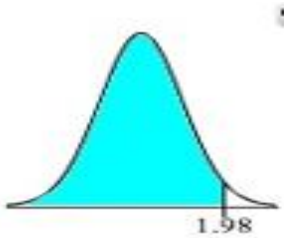
ویژگیهای منحنی طبیعی (نرمال)

- * منحنی طبیعی متقارن است و حداکثر ارتفاع آن در میانگین است .
- * در منحنی طبیعی میانگین ، میانه و نما بر روی هم قرار دارند
- * منحنی دو نقطه عطف دارد که نسبت به خطی که نشانگر میانگین است قرینه هستند(شکل منحنی در این دو نقطه تغییر می کند فاصله این دو نقطه 1+ و 1 - انحراف استاندارد است
- دنباله های منحنی با محور X موازی هستند. از ∞ - تا ∞+ ادامه دارد (در عمل از3 - تا 3+ انحراف استاندارد که با Z نشان می دهند)
- *شکل منحنی توزیع طبیعی شبیه زنگوله است

سطوح زیر منحنی طبیعی استاندارد







Z	سطح تا میانگین	سطح بزرگتر	سطح کوچکتر
0.00	.0000	.5000	.5000
0.10	.0398	.5398	.4602
0.50	.1915	.6915	.3085
0.40	.1554	.6554	.3446
1.00	.3413	.8413	.1587
1.30	.4032	.9032	.0968
1.50	.4332	.9332	.0668
1.65	.4505	.95.5	.1495
1.85	.4648	.9678	.0322
2.00	.4772	.9772	.0228
2.45	.4929	.9929	.0071
2.74	.4969	.9969	.0031
2.99	.4986	.9986	.0014
3.00	.4987	.9987	.0013
3.10	.4990	.9990	.0010
3.70	.4999	.9999	.0001

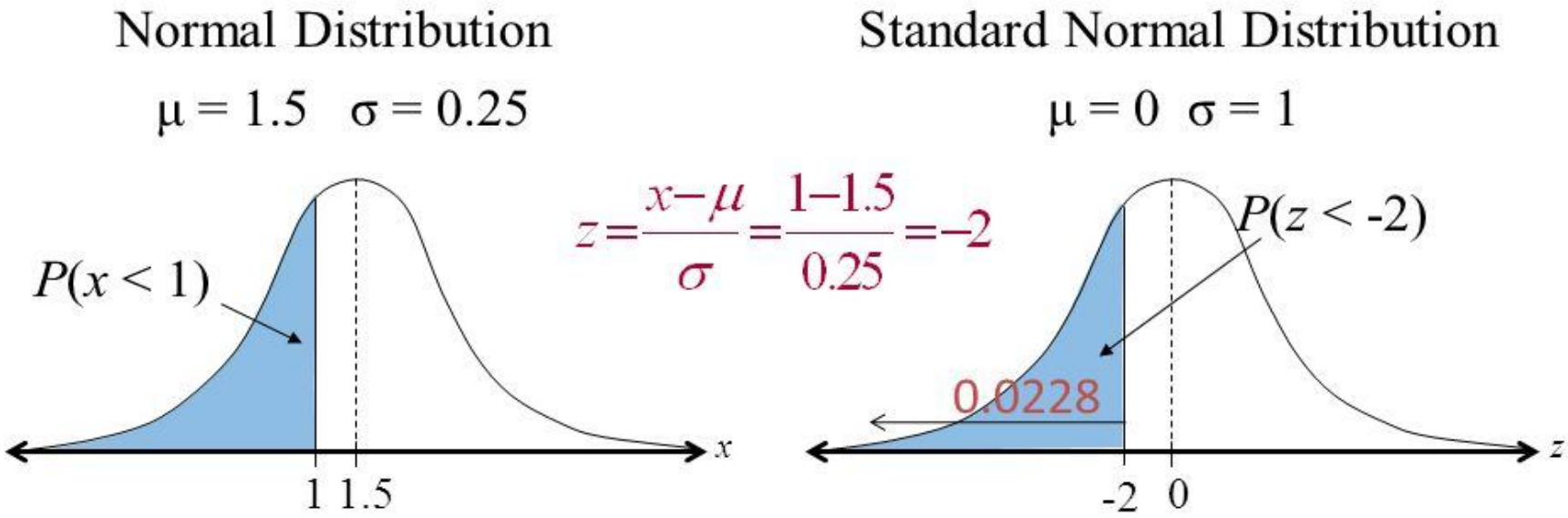
پیدا کردن رتبه درصدی معادل نمره z

1 - مقدار Z را محاسبه می کنیم

$$z = \frac{X - \bar{X}}{S}$$

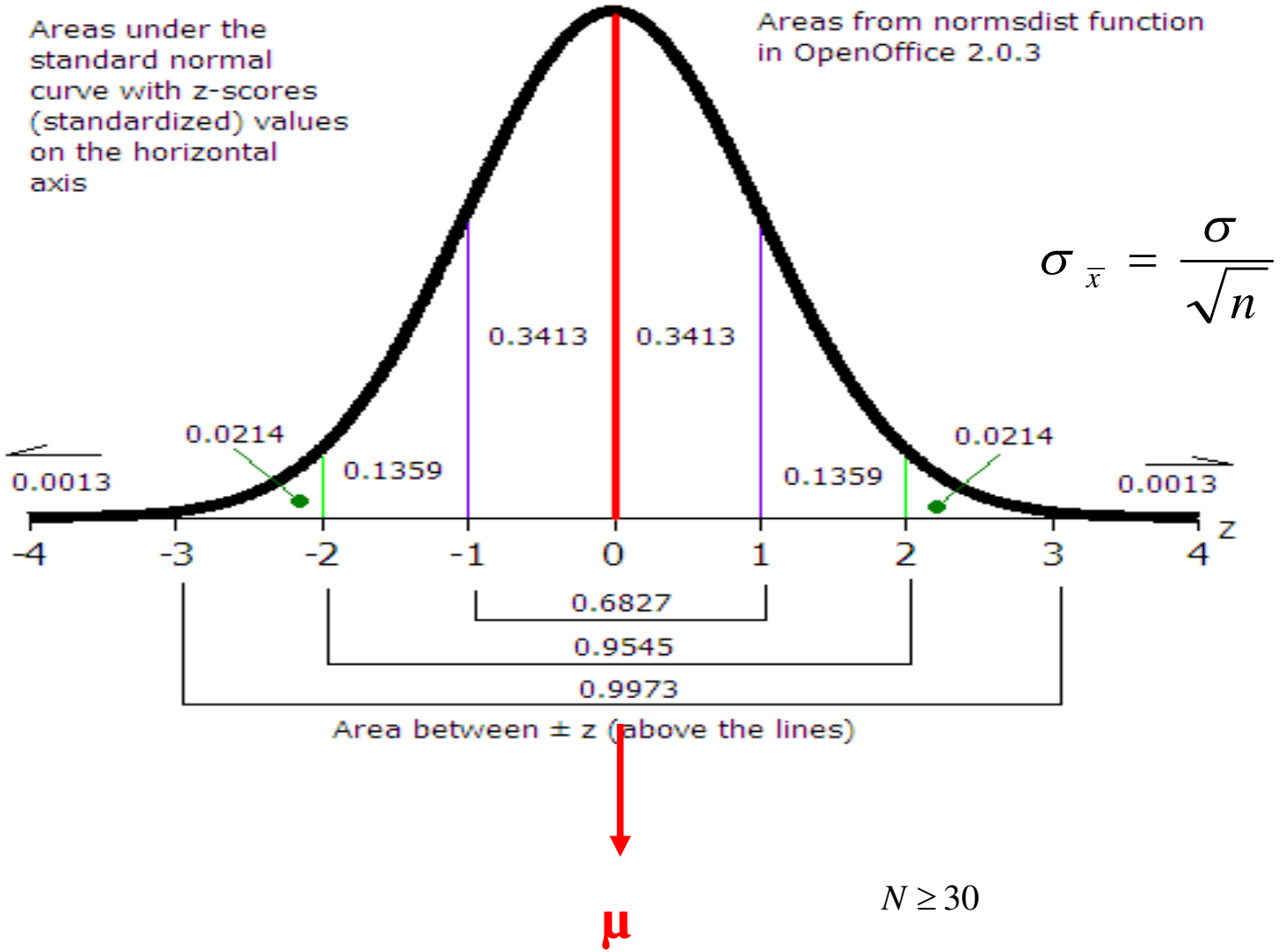
2 - در صورتی که مقدار z منفی باشد در ستون (سطح کوچکتر) و در صورت مثبت بودن در ستون (سطح بزرگتر) رتبه درصدی معادل z را تعیین می کنیم

Solution: Finding Probabilities for Normal Distributions



$$P(x < 1) = \mathbf{0.0228}$$

قضیه حد مرکزی (منحنی نرمال)



$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$N \geq 30$

μ

خطای نمونه گیری

چنانچه نمونه های مختلفی ازیک جامعه انتخاب کنیم ،این نمونه ها دارای ویژگیهای یکسانی نیستند و حتی امکان دارد ویژگیهای آنها شبیه جامعه ای که از آن انتخاب شده اند نباشد. علت این اختلاف ، وجود نوعی اشتباه است که آن را خطا یا اشتباه نمونه گیری می گویند.

چنانچه نمونه های زیادی را با حجم مساوی ، از یک جامعه انتخاب کنیم ، توزیع میانگین های این نمونه ها یک توزیع طبیعی خواهد بودو میانگین میانگینهای نمونه های انتخاب شده تقریبا برابر میانگین جامعه ای است که نمونه از آن انتخاب شده

بنابراین وقتی نمونه ای از یک جامعه انتخاب می شود و میانگین یکی از ویژگیهای آن محاسبه می شود ، میانگین محاسبه شده ، به دلیل خطای نمونه گیری ، با میانگین جامعه ای که نمونه از آن انتخاب شده است متفاوت است

انحراف استاندارد توزیع نظری میانگینها شاخصی است که به وسیله آن خطای نمونه گیری اندازه گیری می شود و آن را خطای استاندارد میانگین می نامند . خطای استاندارد میانگین با $S_{\bar{X}}$ نمایش داده می شودو به کمک فرمول زیر محاسبه می شود

$$s_{\bar{x}} = \frac{s}{\sqrt{N}} = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}{N}$$

اندازه نمونه =n

انحراف استاندارد(خطای استاندارد) =S

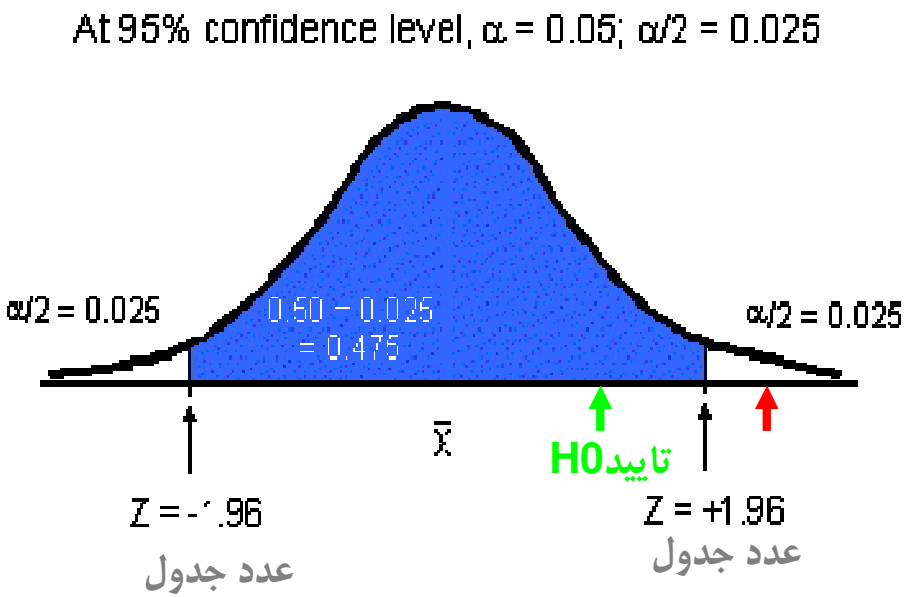
64دانش آموز را به صورت تصادفی از بین دانش آموزان مدرسه ای انتخاب کرده ایم ، چنانچه میانگین و انحراف استاندارد نمره علوم دانش آموزان به ترتیب 17 و 5/4 باشد . خطای استاندارد میانگین را محاسبه کنید.

چنانچه میانگین هوش 36 نفر کودک 116 و انحراف استاندارد هوش آنها 18 باشد ، میانگین هوش جامعه ای که از آن انتخاب شده اند با احتمال 95 درصد بین چه نمراتی قرار دارد؟

میانگین قد یک نمونه 25 نفری از دانش آموزان 165 cm و انحراف استاندارد قد آنان 15 cm است با سطح اطمینان 95% میانگین قد واقعی دانش آموزان را محاسبه کنید .

آزمونی را برای 64 نفر از دانش آموزان دبیرستانی که بطور تصادفی انتخاب کرده ایم اجرا کرده ایم . چنانچه میانگین و انحراف استاندارد این آزمون به ترتیب 25 و 10 باشد : الف)احتمال اینکه میانگین آزمون بین 26 و 24 باشد چقدر است ؟ ب) در صورتی نمونه انتخابی 49 نفر باشد، احتمال اینکه میانگین آن بین 26 و 24 باشد چقدر است ؟

مراحل عمومی آزمون فرضیه های آماری



گام اول) تعریف H_0 و H_1

تعیین ادعا و نقیض ادعا

گام دوم) تعیین مقدار آماره

همان عدد محاسبه شده از فرمولها می باشد

گام سوم) تعیین سطح بحرانی

بر اساس عدد جدول مشخص می شود

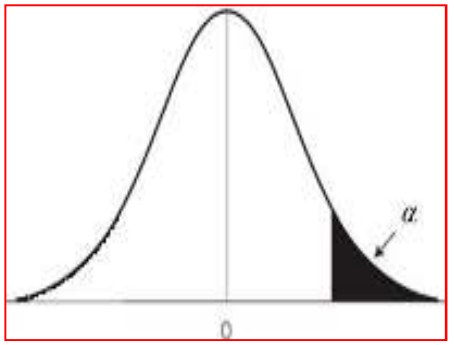
گام چهارم) تصمیم گیری

با مقایسه عدد جدول و عدد محاسبه شده

فرضیه آماری • تعیین ناحیه بحرانی α

حالت سوم

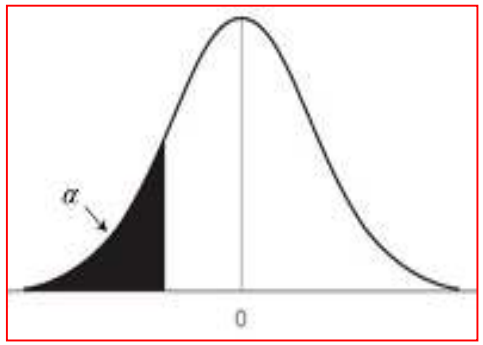
$$\begin{cases} H_0 : \mu_x \le \mu_0 \\ H_1 : \mu_x > \mu_0 \end{cases}$$



یک دامنه راست (One Tailed)

حالت دوم

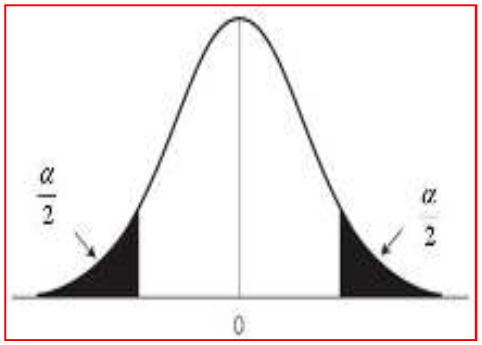
$$\begin{cases} H_0 : \mu_x \ge \mu_0 \\ H_1 : \mu_x < \mu_0 \end{cases}$$



یک دامنه چپ (One Tailed)

حالت اول

$$\begin{cases} H_0 : \mu_x = \mu_0 \\ H_1 : \mu_x \ne \mu_0 \end{cases}$$



دو دامنه (Two Tailed)

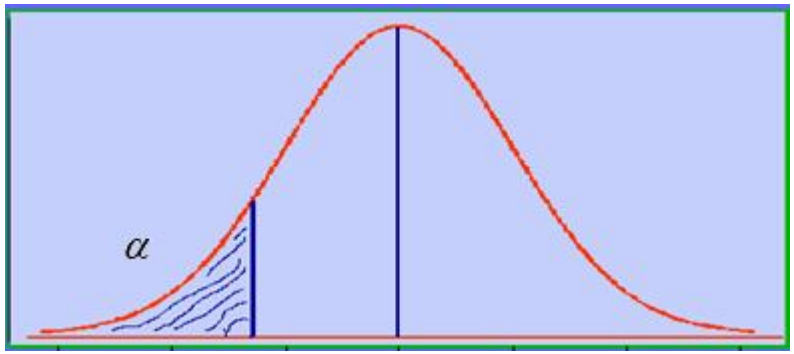
چنانچه ملاک آزمون در ناحیه بحرانی قرار بگیرد فرضیه H_0 رد می شود.

آزمون فرض یک دامنه و دو دامنه :

بستگی به تعریف فرضیه ها چنانچه فرضیه پژوهشی بدون جهت باشد دو دامنه در غیر اینصورت یک دامنه خواهد بود

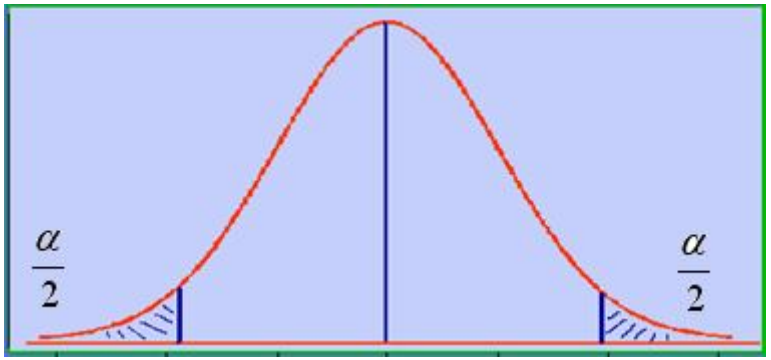
مثال 1) میانگین نمره ریاضی دانش آموزان کمتر از 15 است.

$$\begin{cases} H_0 : \mu \geq 15 \\ H_1 : \mu < 15 \end{cases}$$



مثال 2)مردان و زنان تفاوت سطح درجه هوشی ندارند.

$$\begin{cases} H_0 : \mu_w = \mu_m \\ H_1 : \mu_w \neq \mu_m \end{cases}$$

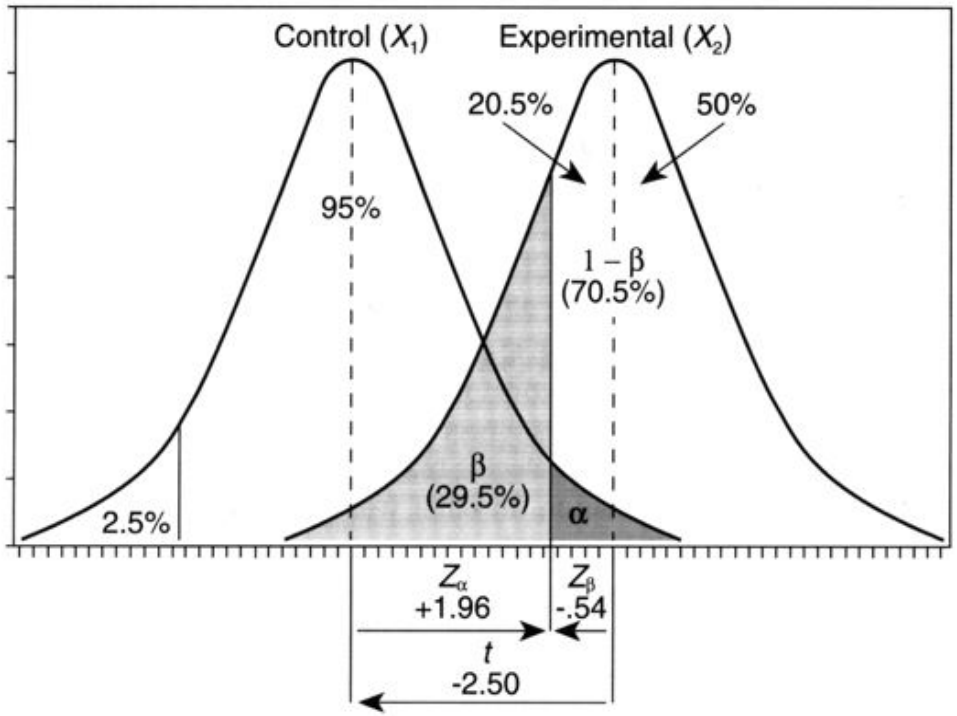


خطاهای آماری

- خطای نوع اول(α): رد کردن H_0 در حالی که H_0 درست است.
- خطای نوع دوم(β): پذیرفتن H_0 در حالی که H_0 غلط است.

گزینه های صحیح		
H_1 درست است	H_0 درست است	نتیجه گیری از نمونه
β	**تصمیم درست است**	H_0 پذیرفته می شود
تصمیم درست است	α	H_1 پذیرفته می شود

احتمال ارتکاب خطای نوع دوم به عوامل زیر بستگی دارد:



1 - سطح معنی دار بودن

2 - اندازه تاثیر متغیر مستقل

3 - مقدار پراکندگی موجود در متغیر وابسته

4 - اندازه یا حجم نمونه

-بین خطای نوع اول و دوم رابطه معکوس وجود دارد.

-اندازه خطای نوع اول با تنظیم α کاهش پیدا می کند

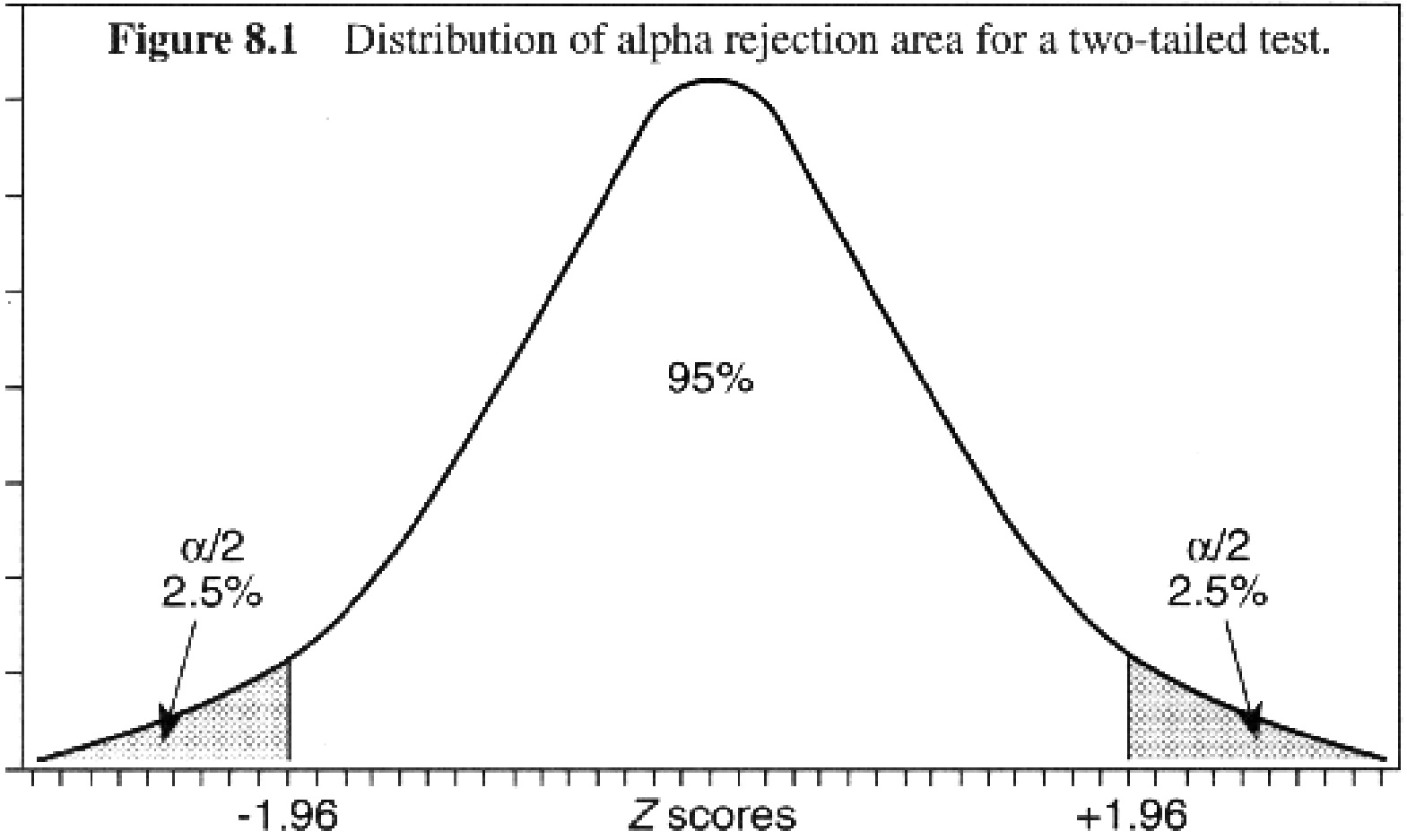
- افزایش حجم نمونه ، مقادیر α و β را همزمان کاهش می دهد

توان آزمون : احتمال رد فرض صفر هنگامی که این فرض واقعا غلط است.به عبارت دیگر احتمال قبول فرض یک وقتی این فرض واقعا درست است

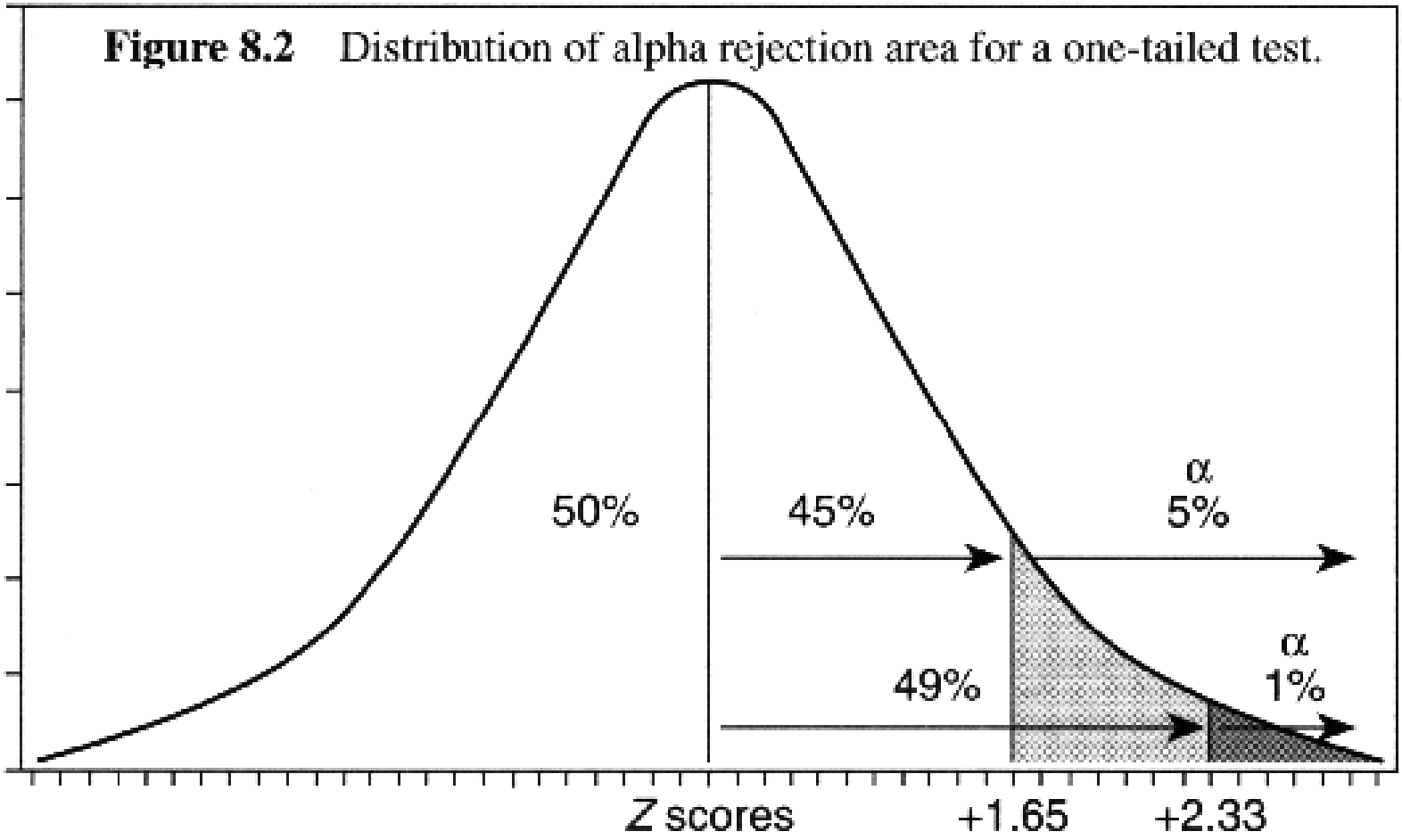
(ردH0 وقتی H0 غلط است)احتمال =β -1

سطح اطمینان (سطح معنی داری) : انتخاب این سطح اختیاری است ولی غالباً ۰/۰۵ و ۰/۰۱ انتخاب می کنند.

Two Tailed Test: Null No Difference.



One Tail Test: Null A > B. More Powerful, easier to find differences.



آزمون های t

توزیع t استودنت درجه آزادی df=n-1

وقتی به جای بررسی کل جامعه یک نمونه را برای بررسی انتخاب نموده ایم و یا اینکه مقدار جامعه کمتر 100 نفر است . N<100 - برای آزمون توزیع t در جه آزادی عبارت است از : (N-1)

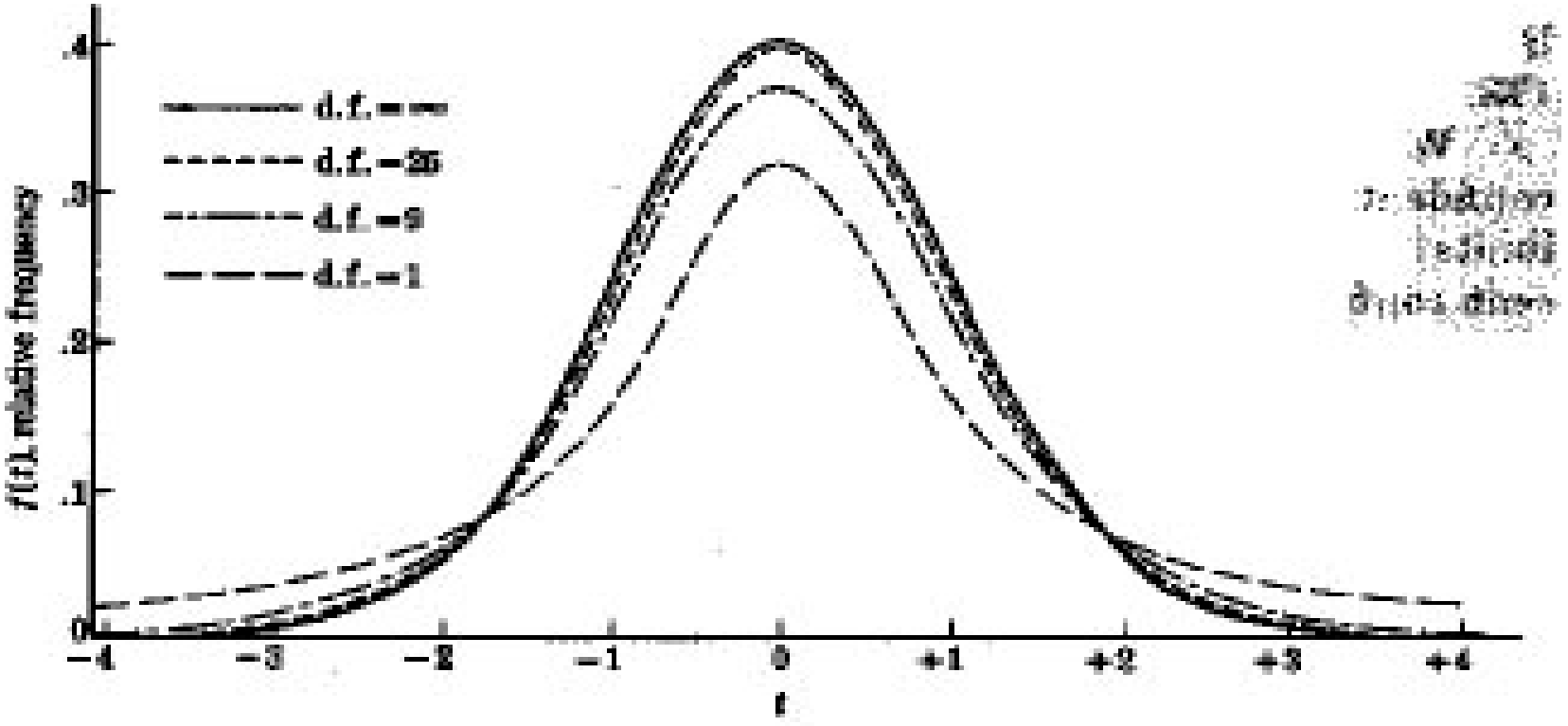


fig. 10.1 Distribution of t for various degrees of freedom. (From D. Lewis, quantitative methods in psychology, McGraw-Hill Book Company, New York, 1960.)

الف) آزمون فرضیه در باره میانگین جامعه

به منظور برآورد محل میانگین جامعه مورد استفاده قرار می گیرد. ضمنا از این آزمون می توان برای مقایسه میانگین یک نمونه با یک عدد ثابت یا ادعا استفاده کرد.

$$t = \frac{(\overline{X} - \mu)}{s_{\bar{x}}}$$

میانگین نمونه = $\bar{x}$

میانگین جامعه = μ

خطای استاندارد = $s_{\bar{x}}$

$$H_0 : \mu = 10; \quad H_1 : \mu \neq 10; \quad s_{\overline{x}} = 5; \quad N = 25$$

$$s_{\bar{x}} = \frac{s_X}{\sqrt{N}} = \frac{5}{\sqrt{25}} = 1$$

$$\overline{X} = 11 \rightarrow t = \frac{(11-10)}{1} = 1$$

$$t(.05,24) = 2.064$$

1<2.064

آزمون فرضیه در باره میانگین جامعه
(نمونه بزرگ معادل z است n>100)

$$t = \frac{(\overline{X} - \mu)}{s_{\bar{x}}} \qquad s_{\bar{x}} = \frac{s}{\sqrt{N}} = \sqrt{\frac{\sum (X - \overline{X})^2}{N - 1}}{N}$$

- **فرضیه**

$$H_0 : \mu = 10; \quad H_1 : \mu \neq 10; \quad s_X = 5; \quad N = 200$$

- خطای استاندارد

$$s_{\bar{x}} = \frac{s_X}{\sqrt{N}} = \frac{5}{\sqrt{200}} = \frac{5}{14.14} = .35$$

$$\overline{X} = 11 \rightarrow t = \frac{(11 - 10)}{.35} = 2.83;\ 2.83 > 1.96$$

ب) آزمون t برای مقایسه میانگین های دو گروه مستقل

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sum (x_1 - \bar{x}_2)^2 + \sum (x_2 - \bar{x}_2)^2}{N_1 + N_2 - 2}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$df = N_1 - 1 + N_2 - 1 = N_1 + N_2 - 2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sum x_1^2 - \frac{(\sum x_1)^2}{n_1} + \sum x_2^2 - \frac{(\sum x_2)^2}{n_2}}{N_1 + N_2 - 2}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

ب) آزمون t برای مقایسه میانگین های دو گروه مستقل

$$S_{\bar{x}} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

$$df = n_1 - 1 + n_2 - 1 = n_1 + n_2 - 2$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{S_{\bar{x}}}$$

گروه ۱		گروه ۲	
x_1	x_1^2	x_2	x_2^2
۱۶	۲۵۶	۱۰	۱۰۰
۱۲	۱۴۴	۷	۴۹
۱۱	۱۲۱	۷	۴۳
۸	۶۴	۶	۳۶
۹	۸۱	۵	۲۵
۴	۱۶		
Σx=۶۰	Σ x2=۶۸۲	Σx=۳۵	Σ x2=۲۵۹
$\bar{x}$ =۱۰	n=۶	$\bar{x}$ =۷	n=۵

$$t=\frac{10-7}{\sqrt{\frac{(6-1)(16.4)+(5-1)(3.5)}{6+5-2}\left[\frac{1}{6}+\frac{1}{5}\right]}}=\frac{3}{\sqrt{3/911}}=1/52$$

$$t=\frac{10-7}{\sqrt{\frac{82+14}{6+5-2}\left[\frac{1}{6}+\frac{1}{5}\right]}}=\frac{3}{\sqrt{\left(\frac{96}{9}\right)\left(\frac{11}{30}\right)}}=\frac{3}{\sqrt{3/911}}=1/52$$

$$t=\frac{10-7}{\sqrt{\frac{\left[682-\frac{(60)^2}{6}\right]+\left[259-\frac{(35)^2}{5}\right]}{6+5-2}\left[\frac{1}{6}+\frac{1}{5}\right]}}=\frac{3}{\sqrt{\left(\frac{82+14}{9}\right)\left(\frac{11}{30}\right)}}=\frac{3}{\sqrt{3/911}}=1.52$$

آزمون t برای مقایسه میانگین های دو گروه مستقل

$$S_{\bar{x}} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\left[\frac{N_1 + N_2}{N_1 N_2}\right]}$$

$$H_0 : \mu_1 - \mu_2 = 0;\ H_1 : \mu_1 - \mu_2 \neq 0$$

$$\bar{X}_1 = 18;\ s_1^2 = 7;\ N_1 = 5$$

$$\bar{X}_2 = 20;\ s_2^2 = 5.83;\ N_2 = 7$$

$$S_{\bar{x}} = \sqrt{\frac{4(7) + 6(5.83)}{5 + 7 - 2}\left[\frac{12}{35}\right]} = 1.47$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{S_{\bar{x}}}$$

$$t_{م} = \frac{(18-20)}{1.47} = \frac{-2}{1.47} = -1.36$$

$$t_{ج} = t(.05,10) = 2.23$$

ب) آزمون t برای مقایسه میانگین های گروههای همبسته

بیشترین کاربرد را برای برای مقایسه پیش آزمون و پس آزمون دارد . در برخی از گروههای همتا که دو گروه از بسیاری از متغیر ها وابسته هستند نیز قابل استفاده است

$$D_i = X_1 - X_2$$

روش 1

$$t = \frac{\overline{D}}{s_{\overline{D}}} = \frac{-5}{1.674} = -2.99$$

$$\overline{D} = \frac{\sum(D_i)}{N} \qquad s_D^2 = \frac{\sum(D_i - \overline{D})^2}{N-1} \qquad S_{\overline{D}} = \frac{s_D}{\sqrt{N}}$$

روش 2

$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D)^2}{n-1}}}$$

df=n-1

مثال:

	آزمون 1	آزمون 2	D	D2
محمد	10	7	3	9
رضا	5	3	2	4
حسین	6	7	-1	1
محمد	7	5	2	4
حسن	10	8	2	4
عباس	6	4	2	4
محسن	7	5	2	4
علی	8	2	6	36
احسان	6	3	3	9
باقر	5	6	-1	1
			∑D=20	∑D2=76

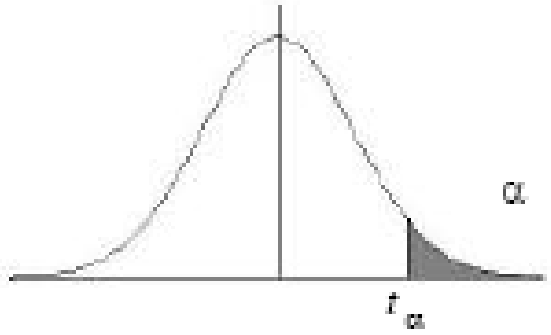
$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D)^2}{n-1}}}$$

$$t = \frac{20}{\sqrt{\frac{(10)(76) - (20)^2}{10 - 1}}} = 3.16$$

df=N-1 10-1 = 9

$t_{ج} = t(.05,9)=2.26$

Table 4: Percentage Points of the t distribution

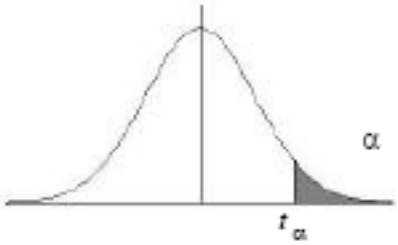


Look up α

Look up df

	α					
df	0.250	0.100	0.050	0.025	0.010	0.005
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106

Table 4: Percentage Points of the *t* distribution



	α					
df	0.250	0.100	0.050	0.025	0.010	0.005
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
			⋮			
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

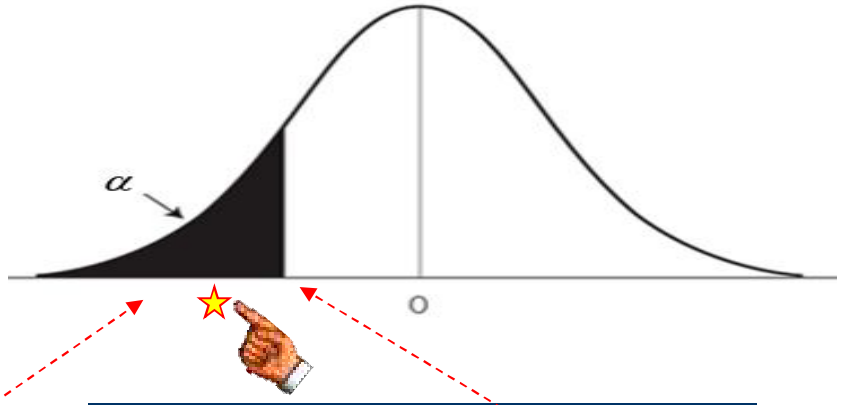
$df = \infty$ برابر با توزیع طبیعی و مساوی Z می باشد

مثال1) پژوهشگری فرضیه خود را بدین شکل صورتبندی نموده است:
"میانگین امتیاز عملکرد کارکنان در سازمان کمتر از 50 است" وی برای آزمون فرضیه فوق یک نمونه 64 تایی به شکل تصادفی انتخاب کرده که میانگین و انحراف معیار آن 45 و 16 می باشد. در سطح خطای 5% فرضیه فوق را آزمون نمایید.

نقیض ادعا

$$\begin{cases} H0 : \mu_x \geq 50 \\ H1 : \mu_x < 50 \end{cases}$$

ادعا



$$z_{\alpha} = z_{0.05} = -1.645$$

چون تعداد نمونه بیشتر از 50 می باشد لذا از فرمول زیر استفاده می کنیم:

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{45 - 50}{\frac{16}{\sqrt{64}}} = -2.5$$

تصمیم گیری:

چون ملاک آزمون(2,5 -) در ناحیه بحرانی قرار گرفته پس فرضیه صفر رد می شود و در نتیجه فرضیه پژوهش(H1) تایید می شود.

مثال2) پژوهشگری فرضیه خود را بدین شکل صورتبندی نموده است:

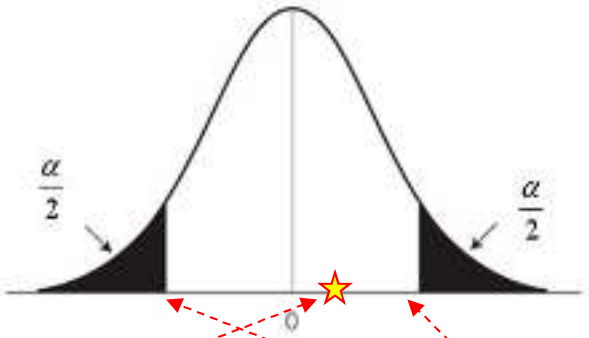
"میانگین رضایت شغلی کارکنان 55 می باشد"

ایشان یک نمونه تصادفی 12 تایی انتخاب کرده که میانگین و انحراف معیار آن 60 و 15 است سطح رضایت شغلی کارکنان را در سطح اطمینان 99 درصد آزمون کنید..

نقیض ادعا

ادعا

$$\begin{cases} H_0 : \mu_x \neq 55 \\ H_1 : \mu_x = 55 \end{cases}$$



چون تعداد نمونه کوچک می باشد لذا از فرمول زیر استفاده می کنیم:

$$t = \frac{\bar{x} - \mu 0}{s_{\bar{x}}} = \frac{60 - 55}{\frac{15}{\sqrt{12}}} = 1.155$$

$$t_{\frac{\alpha}{2},df} = t_{0.005,11} = \pm 3.106$$

تصمیم گیری: چون t آزمون(1,155) درخارج از ناحیه بحرانی قرار گرفته پس فرضیه صفر رد می شود و در نتیجه ادعای محقق تایید می گردد و فرضیه پژوهشی پذیرفته می شود.

فرضیه فوق را در سطح اطمینان **90** *درصد نیز آزمون نمایید.*

مثال)محققی به منظور مقایسه عملکرد مدیران دو سازمان که به روش های مشارکتی و دستوری اداره می شود فرضیه زیر را صورتبندی نموده است:

"بهره وری سبک مدیریت مشارکتی بهتر از سبک مدیریت دستوری است"

او بدین منظور از هر سازمان چند نمونه تصادفی انتخاب کرد که اطلاعات آنها بشرح زیر می باشد.

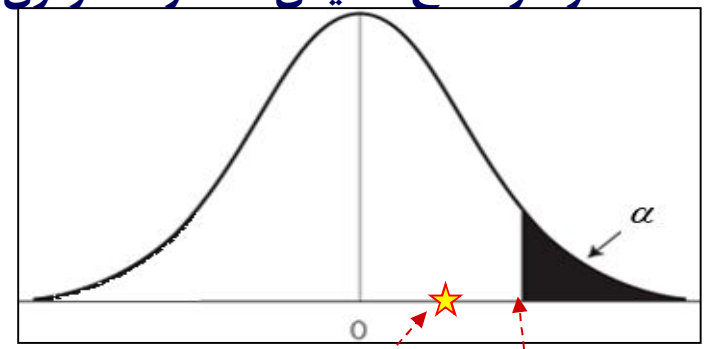
سبک مشارکتی	سبک دستوری
$n_1=10$	$n_2=15$
$Xb_1=52$	$Xb_2=45$
$S_1=12$	$S_2=8$

با فرض تساوی واریانس ها و نرمال بودن دو جامعه فرضیه داده شده را در سطح اطمینان 99 درصد آزمون نمایید.

$$\begin{cases} H_0 : \mu_1 \le \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$$

نقیض ادعا

ادعا



$$t=\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{(n_1-1)s_1^{\,2}+(n_2-1)s_2^{\,2}}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{(52-45)}{\sqrt{\frac{(14\times 64)+(9\times 144)}{15+10-2}\left(\frac{1}{10}+\frac{1}{15}\right)}}=1.757$$

$$t_{\alpha,df}=t_{0.01,23}=2.5$$

df=15+10-2=23<50

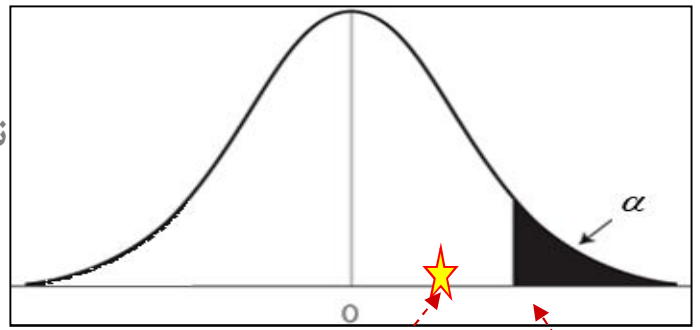
تصمیم گیری: چون ملاک آزمون(1,757) درخارج از ناحیه بحرانی قرار رفته پس فرضیه صفر رد نمی شود و در نتیجه نقیض ادعای محقق تایید می گردد و فرضیه پژوهشی رد می شود.

مثال1)نمرات وضع موجود و وضع مطلوب شایستگی های مدیریتی پنج مدیر به شکل جدول زیر می باشد:

	1	2	3	4	5
وضع موجود	50	59	50	58	50
وضع مطلوب	40	57	47	50	48

آیا فرضیه زیر در سطح اطمینان 99درصد مورد تایید است؟ "وضع موجود و وضع مطلوب شایستگی های مدیریتی مدیران تفاوت معنی داری دارد و وضع مطلوب بالاتر از وضع موجود است."

$\mu_2 > \mu_1 \rightarrow \mu_d > 0 \rightarrow$ $\begin{cases} H_0 : \mu_d \le 0 \\ H_1 : \mu_d > 0 \end{cases}$ نقیض ادعا / ادعا



	1	2	3	4	5
d_i	-10	-2	-3	-8	-2
$(d_i - \bar{d})^2$	25	9	4	9	9
d_i^2	100	4	9	64	4

$t_{\alpha,df} = t_{0.01,4} = 3.747$

تصمیم گیری: چون ملاک آزمون درخارج از ناحیه بحرانی قرار رفته پس فرضیه صفر رد نمی شود و در نتیجه نقیض ادعای محقق تایید می گردد و فرضیه پژوهشی رد می شود.

$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D)^2}{n-1}}}$$

$$t = \frac{-25}{\sqrt{\frac{5(181)(-25)^2}{5-1}}} = -2.99$$

روش دوم

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{-5}{1.674} = -2.99$$

مثال2) نمرات پیش آزمون و پس آزمون پنج فراگیر در یک دوره آموزشی به شکل زیر می باشد:

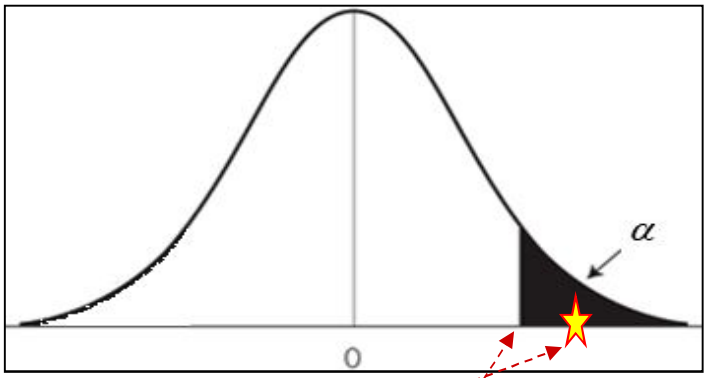
آیا فرضیه زیر در سطح اطمینان 99درصد مورد تایید است؟ "دوره آموزشی مورد نظر اثربخش می باشد"

	1	2	3	4	5
پیش آزمون	12	8	4	16	10
پس آزمون	14	13	10	18	15

$$\mu_2 > \mu_1 \rightarrow \mu_d > 0 \rightarrow \begin{cases} H_0 : \mu_d \le 0 \\ H_1 : \mu_d > 0 \end{cases}$$

نقیض ادعا

ادعا



	1	2	3	4	5
d_i	2	5	6	2	5
d_i^2	4	25	36	4	25

$$t_{\alpha,df} = t_{0.01,4} = 3.747$$

$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D)^2}{n-1}}} = 4.78$$

$$t = \frac{20}{\sqrt{\frac{5(94) - (20)^2}{5-1}}} = 4.78$$

تصمیم گیری: چون ملاک آزمون در ناحیه بحرانی قرار رفته پس فرضیه صفر رد می شود و در نتیجه ادعای محقق تایید می گردد و فرضیه پژوهشی پذیرفته می شود.

مراحل آزمون استقلال کای دو

- مرحله اول) تعریف H_0 و H_1

$$\begin{cases} H_0: \text{بین X و Y ارتباط وجود ندارد(مستقل هستند)} \\ H_1: \text{بین X و Y ارتباط وجود دارد(مستقل نیستند)} \end{cases}$$

- مرحله دوم) محاسبه آماره یا ملاک آزمون

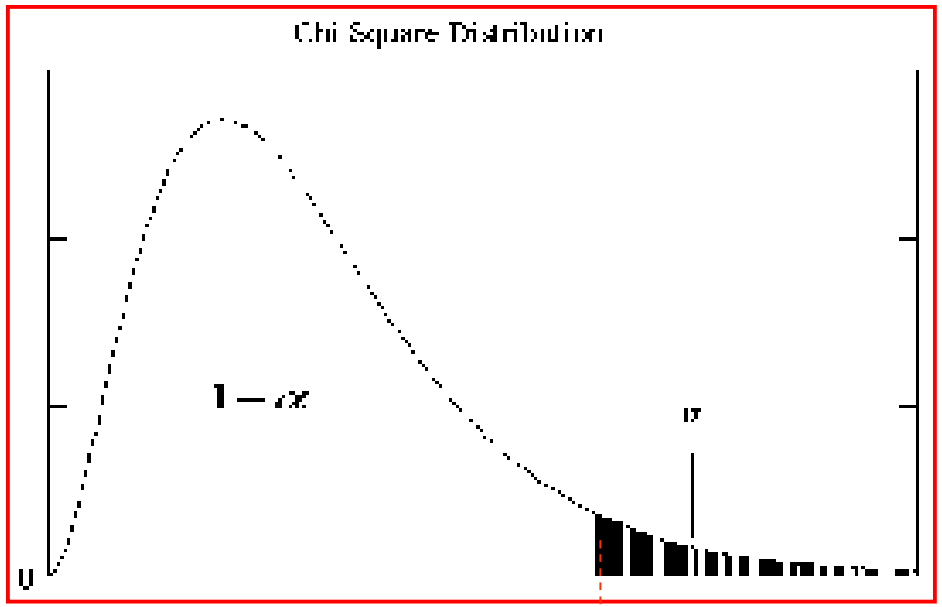
$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

$$E_i = \frac{(n_{i.} * n_{.j})}{n} = \frac{\text{مجموع سطر} * \text{مجموع ستون}}{\text{کل فراوانی}}$$

$$df = (r-1)(c-1)$$

- **تعیین ناحیه بحرانی**

با استفاده از این آزمون می فهمیم که بین گزینه ارتباط وجود دارد



$$\chi^2_{\alpha,\,df}$$

- **اگر مقدار کای دو در ناحیه بحرانی قرار بگیرد فرضیه صفر رد می شود.**

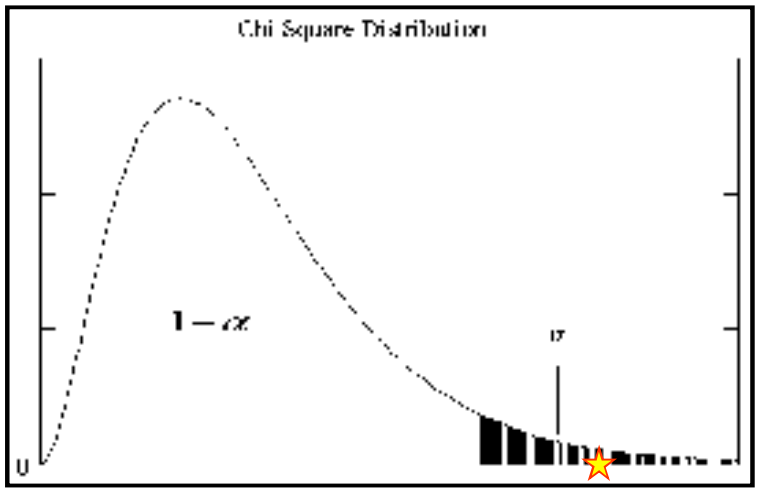
مثال) فرضیه ای بصورت زیر تدوین شده است:"بین عملکرد کارکنان و میزان رضایت شغلی آنها ارتباط وجود دارد"
برای بررسی فرضیه فوق یک نمونه 180 نفره بطور تصادفی انتخاب شده و میزان عملکرد و رضایت شغلی آنها اندازه گیری شده است که اطلاعات آن در جدول زیر ارائه شده است. در سطح خطای 5 درصد فرضیه فوق را آزمون نمایید.

رضایت / عملکرد	بالا	متوسط	پایین
خوب	۱۸	۲۰	۷
متوسط	۱۵	۳۷	۳۸
ضعیف	۷	۲۳	۱۵

H0:بین X و Y ارتباط وجود ندارد(مستقل هستند)

H1:بین X و Y ارتباط وجود دارد(مستقل نیستند)

رضایت / عملکرد	بالا		متوسط		پایین		جمع
خوب	18	10	20	20	7	15	45
متوسط	15	20	37	40	38	30	90
ضعیف	7	10	23	20	15	15	45
جمع	40		80		60		180



$$\chi^2 = \Sigma \frac{(F_0 - F_e)^2}{F_e} = \frac{(18-10)^2}{10} + \frac{(20-20)^2}{20} + \frac{(7-15)^2}{15} + \ldots + \frac{(15-15)^2}{15} = 15.62$$

$$df = (r-1)(c-1) = (3-1)(3-1) = 4$$

تصمیم گیری: چون ملاک آزمون در ناحیه بحرانی قرار گرفته لذا فرضیه صفر رد می شود لذا می توان گفت که بین عملکرد و رضایت ارتباط وجود دارد و مستقل از هم نیستند.

1 - محققی می خواهد روش تدریس مشارکتی را در کلاس درس علوم اجرا کند برای این کار یک کلاس به روش تدریس مشارکتی را به عنوان گروه تجربی و کلاس درس روش تدریس معمولی را به عنوان گروه شاهد در نظر می گیرد. کدام طرح تحقیق مناسب این مطالعه است روش تحلیل آماری مناسب را برای این طرح معرفی کنید..

2 - اگر در یک آزمون tتک نمونه ای در سطح اطمینان 99% نتیجه آزمون دو دامنه برابر 3/34 باشد چگونه باید تصمیم گیری کرد ؟

3 - در یک نمونه ای به حجم 121 نفر ، میانگین هوش 82 و انحراف استاندارد 22است با 95% اطمینان حدود میانگین هوش جامعه را بدست آورید؟

4 - معلمی علاقمند است این فرضیه را که میانگین جامعه ای مساوی 17است آزمون کند. به همین منظور به صورت تصادفی نمونه ای به اندازه 10 نفر از جامعه مورد نظر انتخاب می کند. چنانچه نمره های این 10 نفر به شرح زیر باشد، با سطح معنی داری 0/05 و با یک آزمون دو دامنه فرضیه فوق را آزمون کنید : 15 -20 -19 -14 -18 -17 -10 -11 -14 -12

5 - به منظور آزمون این فرضیه که « تدریس زبان فارسی به صورت ترکیبی بهتر از تجزیه ای است » ، دو گروه دانش آموز کلاس اول ابتدایی به صورت تصادفی انتخاب گردید. به یک گروه از آنها (گروه A) به صورت ترکیبی و به گروه دیگر (گروهB) به صورت تجزیه ای تدریس شد . پس از پایان آموزش ، آزمونی به منظور اندازه گیری پیشرفت دو گروه اجرا شد و نمره های دو گروه مطابق جدول زیر بدست آمد:

گروه A : 4 - 3 – 6 – 5 – 7 گروه B : 0 - 1 - 2 – 3 – 4

الف) فرض صفر و فرض خلاف مسئله فوق را بیان کنید.
ب) درجات آزادی مسئله مورد بحث را محاسبه کنید.
ج) مقدار t را برای داده های جمع آوری شده محاسبه کنید.
د) آیا t محاسبه شده در سطح 0/05 معنادار است ؟
هـ) آیا t محاسبه شده در سطح 0/01 معنادار است ؟
و) چه نتیجه ای می توان گرفت ؟

6 - با یک آزمون دو دامنه و در سطح 0/05 ، با نمونه ای به اندازه 16 و میانگین و انحراف استاندارد 81 و 8 این فرضیه را که میانگین جامعه معینی $\mu = 75$ است ، آزمون کنید.

7 - پژوهشگری علاقمند است تأثیر تکنیک حساسیت زدایی منظم در کاهش اضطراب امتحان دانش آموزان را مورد پژوهش قرار دهد. به همین منظور 5 نفر دانش آموزان کلاس پنجم ابتدایی را بصورت تصادفی انتخاب می کند. ابتدا به وسیله یک آزمون میزان اضطراب آنان را اندازه می گیرد. سپس به کمک تکنیک حساسیت زدایی منظم طی یکماه سعی در کاهش اضطراب آنان می نماید و در پایان مجدداً میزان اضطراب آنها را اندازه می گیرد با یک آزمون آماری مناسب و با احتمال 0/05 این فرض را که میانگین اختلاف در جامعه برابر صفر است آزمون کنید.
قبل از اجرا تکنیک : 20 - 27 - 32 - 34 - 14
بعد از اجرا تکنیک : 16 - 7 - 19 - 11 — 8

8 - روان شناسی علاقمند است که تاثیر دارویی را در کاهش افسردگی مورد پژوهش قرار دهد . اطلاعات زیر در نتیجه لجرای آزمایش بدست آمده :

گروهی که دارو مصرف کرده	۱	۲	۳	٤	٤	٥	٥	٨	٩	٩
گروهی که دارو مصرف نکرده	٤	٦	٧	٧	٨	٨	٩	١٠	١٠	١١

الف) فرض صفر و خلاف این آزمایش را بنویسید
ب) نسبت t را محاسبه کنید
ج) نتیجه آزمایش را در سطح01/ و 05/ بنویسید

9— اطلاعات زیر از دو گروه مستقل جمع آوری شده است . با یک آزمون دو دامنه این فرضیه را که توانش یادگیری دو گروه شبیه یکدیگر است مورد بررسی قرار دهید .
گروه اول : $n = 13$ میانگین : 7 مجموع مجذور انحراف از میانگین : 96
گروه دوم : $n = 19$ میانگین : 9/4 مجموع مجذوز انحراف از میانگین : 112

10 - برای نمونه ای که از 25 جفت آزمودنی تشکیل شده است ، $\sum D = 50$ و $\sum D^2 = 400$ است . آزمون t را محاسبه کنید.

داده های کیفی اسمی Categorical	داده رتبه ای و یا داده کمی غیر نرمال	داده کمی و دارای توزیع نرمال	هدف
آزمون نسبت	آزمون میانه	آزمون میانگین و انحراف معیار	توصیف یک گروه
آزمون خی – دو یا آزمون دو جمله ای	آزمون ویلکاکسون	آزمون یک نمونه ای	مقایسه یک گروه با یک مقدار فرضی
آزمون دقیق فیشر (آزمون خی دو برای نمونه های بزرگ)	آزمون من - ویتنی	آزمون برای نمونه های مستقل	مقابسه دو گروه مستقل
آزمون مک - نار	آزمون کروسکال	آزمون زوجی	مقایسه دو گروه وابسته
آزمون خی - دو	آزمون والیس	آزمون آنالیز واریانس یک راهه	مقایسه سه گروه یا بیشتر (مستقل)
آزمون کوکران	آزمون فریدمن	آزمون آنالیز واریانس با اندازه های مکرر	مقایسه سه گروه یا بیشتر (وابسته)
آزمون ضریب توافق	آزمون ضریب همبستگی اسپرمن	آزمون ضریب همبستگی پیرسون	اندازه همبستگی بین دو متغیر
آزمون رگرسیون لجستیک	آزمون رگرسیون نا پارامتریک	آزمون رگرسیون ساده یا غیر خطی	پیش بینی یک متغیر بر اساس یک یا چند متغیر