

بسمه تعالی

کارگاه تجزیه و تحلیل داده ها

(ویژه محققین دانشگاه)

آبان ماه 1395

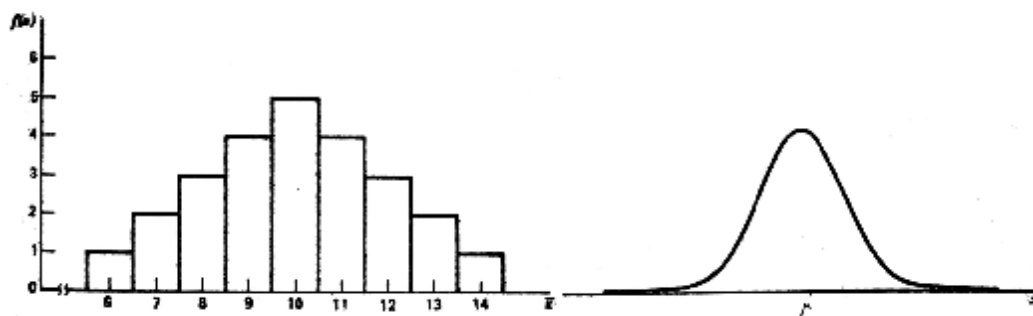
دکتر راهب قربانی

بسم الله الرحمن الرحيم

(Normal) توزیع نرمال (طبیعی، بهنجار)

مثال:

مقدار افت فشار	فراوانی ساده	نسبی	درصد
5/5-6/5	1	0/04	4
6/5-7/5	2	0/08	8
7/5-8/5	3	0/12	12
8/5-9/5	4	0/16	16
9/5-10/5	5	0/20	20
10/5-11/5	4	0/16	16
11/5-12/5	3	0/12	12
12/5-13/5	2	0/08	8
13/5-14/5	1	0/04	4
جمع	25	1	100

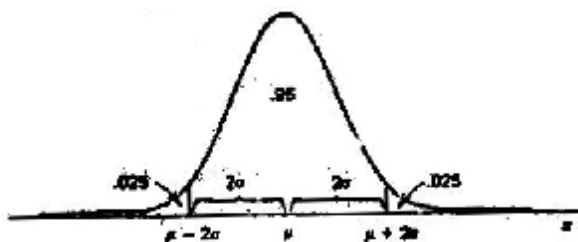


ویژگی ها:

- 1- توزیع نرمال حول میانگین (m) متقارن است.
- 2- میانگین، میانه و مد مساوی (برهم منطبق) هستند.
- 3- سطح کل زیر منحنی نرمال و بالای محور X ها برابر 1 می باشد.

ویژگی توزیع نرمال:

در هر توزیع نرمال 95/44% داده ها در فاصله دو انحراف معیار دو طرف میانگین قرار دارند:



توزیع میانگین نمونه ($\bar{X}$)

قضیه حد مرکزی:

اگر جمعیتی با هر شکل تابعی دارای میانگین m و واریانس s^2 باشد، در نمونه برداری با حجم بزرگ ($n > 30$)، $\bar{X}$ تقریباً دارای توزیع نرمال با میانگین m و واریانس $\frac{s^2}{n}$ می باشد.

برآورد فاصله ای یا فاصله اطمینان

(Confidence Interval = CI)

فاصله اطمینان 95 درصد برای میانگین جامعه از رابطه

$$\bar{X} \pm 1.96 \frac{s}{\sqrt{n}} \text{ به دست می آید.}$$

اگر s نامعلوم باشد، به جای آن از انحراف معیار نمونه (S یعنی s) استفاده می کنیم. در این صورت توزیع Z به t تبدیل می شود. و رابطه به صورت زیر در می آید:

$$\bar{X} \pm t_{1-\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$

آزمون فرض

- جامعه
- پارامتر (*Parameter*)
- آماره (*Statistic*)
- فرضیه های H_0 و H_1

آزمون فرضیه

قاعده ای است که به وسیله آن مشخص می شود که آیا نمونه مورد مطالعه از نظر منطقی با فرضیه مورد نظر مطابقت دارد یا خیر.

مثال از فرضیه H_0 :

در انجام آزمون آماری در معرض دو خطا قرار داریم :

تصمیم پژوهشگر بر اساس نمونه		واقعیت
H_0 رد می شود	H_0 رد نمی شود	
خطای نوع اول	تصمیم گیری درست	H_0 درست
تصمیم گیری درست	خطای نوع دوم	H_0 نادرست

سطح معنی داری

سطح معنی داری یا سطح اشتباه همان احتمال خطای نوع اول است که با a نشان می دهیم. $1-a$ ضریب اطمینان می باشد.

احتمال خطای نوع دوم

احتمال قبول یک فرضیه نادرست که با b نشان می دهیم.

توان آزمون

احتمال رد یک فرضیه نادرست. به عبارتی دیگر $1-b$ توان آزمون است.

p-value

احتمال این که اختلاف مشاهده شده ناشی از شانس باشد.

آزمون تساوی میانگین دو جامعه (t-student)

- نرمال بودن توزیع صفت در دو جامعه 1 و 2 (؟)

- مستقل بودن دو نمونه

فرضیه های H_0 و $H_1(H_A)$

$$H_0 : m_1 = m_2$$

$$H_1 : m_1 \neq m_2$$

در این صورت یکی از دو صورت زیر رخ می دهد:

1) **انحراف معیار صفت در دو جامعه مساوی باشد:**

محاسبه میانگین و انحراف معیار صفت در دو نمونه، محاسبه p -value و تصمیم گیری

2) **انحراف معیار صفت در دو جامعه مساوی نباشد:**

محاسبه میانگین و انحراف معیار صفت در دو نمونه، محاسبه p -value و تصمیم گیری

سؤال:

چگونه به فهمیم که واریانس ها را مساوی بگیریم یا نامساوی؟

آزمون تساوی واریانس های دو جامعه

$$H_0 : s_1^2 = s_2^2$$

$$H_1 : s_1^2 \neq s_2^2$$

تقسیم واریانس های دو نمونه، محاسبه **p-value** و تصمیم گیری

$$F = \frac{S_1^2}{S_2^2} \text{ است}$$

- اگر $F > F_{1-\frac{\alpha}{2}, n_1-1, n_2-1}$ باشد H_0 رد می شود.

آزمون نرمال بودن داده ها

کلموگروف اسمیرنف (مقایسه توزیع تجمعی داده ها و مقایسه آن با توزیع تجمعی نرمال) ، محاسبه **p-value** و تصمیم گیری

آزمون من ویتنی (Mann-Whitney)

موارد کاربرد:

نرمال نبودن توزیع صفت دردو جامعه در آزمون تی دو گروه مستقل

مثال:

مقایسه سطح منواکسید کربن (ppm) در سیگاری ها و غیر سیگاری

سیگاری		غیر سیگاری	
سطح منو	رتبه	سطح منو	رتبه
11	12	6	5/5
12	13	5	4
8	9	7	7
16	14	8	9
10	11	3	2/5
17	15	2	1
8	9	3	2/5
6	5/5	-	-

با استفاده از جمع رتبه ها و محاسبه آماره مربوطه ، p -value محاسبه و تصمیم گیری می شود.

آزمون تی زوجی (Paired t-test)

- نرمال بودن توزیع d

$$H_0 : m_d = 0$$

$$H_1 : m_d \neq 0$$

که $d = X_1 - X_2$ مشاهده قبل و X_2 مشاهده بعد از همان

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} \quad \text{واحد آماری می باشد. آماره آزمون}$$

محاسبه p -value و تصمیم گیری.

مثال: ارزیابی کاهش سطح کلسترول سرم توسط رژیم غذایی معین

شماره	قبل	بعد	d_i
1	201	200	-1
2	231	236	+5
3	221	216	-5
4	260	233	-27
5	228	224	-4
6	237	216	-21
7	326	296	-30
8	235	195	-40
9	240	207	-33

آزمون ویلکسون (Wilcoxon)

این آزمون زمانی که در t زوجی (Paired t-test) توزیع d نرمال نباشد.

مثال: ارزیابی قندخون قبل و بعد از مداخله دارویی

رتبه	d	بعد	قبل
11	-30	110	140
8	-7	132	139
6	-5	105	110
2	+2	170	168
4/5	-4	154	158
10	-9	149	158
1	-1	179	180
3	-3	190	193
12	-11	164	175
4/5	-4	175	179
9	-8	181	189
7	+6	180	174

$$T = 7 + 2 = 9$$

تعیین توزیع احتمال T و محاسبه p -value و تصمیم گیری

موفق باشید