جلسه اول — تحلیل آماری — ۹۱/۱۲/۲۲

آقای دکتر عالم تبریزی

* کتاب آمار و احتمالات و آمار کاربردی (تحلیل آماری) انتشارات صفار و اشراقی (روبروی سینما بهمن) نویسنده: دکتر عالم تبریزی

فرمولها و جداول سر امتحان داده می‌شود

* سه نوع آمار داریم:
  - توصیفی ⇐ اگر بخواهیم همه افراد جامعه را سرشماری کنیم (می‌خواهیم کالا با هم چیزی باشیم)
  - احتمالات ⇐ بحث ما نیست
  - استنباطی ⇐ اگر دسترسی به همه افراد جامعه غیرممکن باشد، نمونه‌گیری می‌کنیم.

* کل آمار ۳ مشخصه دارد: میانگین، نسبت، انحراف معیار (یا واریانس)

(نمادهای این 3 مشخصه در جامعه و نمونه متفاوت است)

این مشخصه‌ها در مورد یک جامعه یا دو جامعه یا n جامعه است.

| | جامعه | نمونه |
|---|---|---|
| میانگین | $\mu$ | $\bar{x}$ |
| نسبت | $P$ | $p$ |
| انحراف معیار | $\sigma$ | $S$ |
| حجم | $N$ | $n$ |
| داده | $X_i$ | $x_i$ |

یک جامعه یا دو جامعه یا n جامعه

* برای بررسی جامعه 3 نوع داده داریم:
1. داده‌های بدون فراوانی و بدون فاصله طبقات
2. داده‌های با فراوانی و بدون فاصله طبقات
3. داده‌های با فراوانی و با فاصله طبقات

$$\mu = \frac{\sum X_i}{N}$$

پراکندگی داده‌ها حول میانگین

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

$$P = \frac{X}{N}$$

$X$: تعداد مورد نظر، $N$: تعداد کل داده‌ها

ضریب تغییرات $CV = \frac{\sigma}{\mu} \times 100$

1. داده‌های بدون فراوانی و بدون فاصله طبقات:

مثال: میانگین سنی آقایان ارشد صنعتی پردیس بهشتی آنها چقدر است؟ انحراف معیار را هم محاسبه کنید.

سن آقایان: 28 35 42 36 40 44 47 36

ابتدا سن آقایان را از کوچک به بزرگ مرتب می‌کنیم

| $X_i$ | $(X_i-\mu)^2$ |
|---|---|
| 28 | $(28-38.5)^2 = 110.25$ |
| 35 | $(35-38.5)^2 = 12.25$ |
| 36 | $(36-38.5)^2 = 6.25$ |
| 36 | $(36-38.5)^2 = 6.25$ |
| 40 | $(40-38.5)^2 = 2.25$ |
| 42 | $(42-38.5)^2 = 12.25$ |
| 44 | $(44-38.5)^2 = 30.25$ |
| 47 | $(47-38.5)^2 = 72.25$ |
| 308 | 252 |

$$\mu = \frac{308}{8} = 38.5$$

$$\sigma = \sqrt{\frac{252}{8}} = 5.61$$

داده‌ها به طور متوسط 5.61 از میانگین منحرف هستند

نکته: انحراف معیار، پراکندگی داده ها حول میانگین را نشان می دهد

نکته: اگر یک جامعه داشته باشیم، مانند مثال قبل عمل می کنیم. ولی اگر بخواهیم دو یا چند جامعه را با هم مقایسه کنیم، مانند مثال زیر عمل می کنیم.

مثال: میانگین و انحراف معیار 3 جامعه داده شده است. توضیح دهید که کدام یک از جوامع نسبت به میانگین متقارن تر است.

| جامعه | $\mu$ | $\sigma$ |
|---|---|---|
| اول | 38.5 | 4.36 |
| دوم | 36 | 3.5 |
| سوم | 32 | 2.5 |

$$CV_1 = \frac{4.36}{38.5} \times 100 = 11.3\%$$

$$CV_2 = \frac{3.5}{36} \times 100 = 9.7\%$$

$$CV_3 = \frac{2.5}{32} \times 100 = 7.8\%$$

جامعه سوم چون ضریب تغییرات کوچکتری دارد، نسبت به میانگین متقارن تر است.

* شاخص های مکانی داده های بدون فراوانی و بدون فاصله طبقاتی:

اگر تعداد داده ها فرد باشد، داده وسط را انتخاب می کنیم. ⟸ برای میانه

از داده های مثال قبل استفاده می شود. اگر تعداد داده ها زوج باشد، دو داده وسط استفاده می شود. می توان مانند زیر هم عمل کرد.

میانه $M_d = \frac{36+40}{2} = 38$

برای محاسبه میانه: $\frac{N}{2} + \frac{1}{2} = \frac{8}{2} + \frac{1}{2} = 4.5$ — $M_d = 38 = Q_2$ میانه (چارک دوم) — 50% افراد کمتر از 38 سال و 50% افراد بیشتر از 38 سال دارند

برای محاسبه چارک اول: $\frac{N}{4} + \frac{1}{2} = \frac{8}{4} + \frac{1}{2} = 2.5$ — $\frac{35+36}{2} = 35.5 = Q_1$ چارک اول — 25% افراد کمتر از 35.5 سال، و 75% افراد بیشتر از 35.5 سال دارند

برای محاسبه چارک سوم: $\frac{3N}{4} + \frac{1}{2} = \frac{3\times8}{4} + \frac{1}{2} = 6.5$ — $\frac{42+44}{2} = 43 = Q_3$ چارک سوم — 75% افراد کمتر از 43 سال و 25% افراد بیشتر از 43 سال سن دارند

مد یا نما $M_o$ داده ای که بیشتر از همه تکرار شده باشد — $M_o = 36$ — اکثریت افراد 36 سال دارند

* اگر دو تا داده داشته باشیم که بیش از یک بار تکرار شده باشد، 2 تا نما داریم ولی اگر 3 تا داده داشته باشیم که بیش از یک بار تکرار شده باشد، نما نداریم

2- داده های با فراوانی و بدون فاصله طبقات:

اگر داده ها تنوعشان کم باشد ولی تعدادشان زیاد باشد، از این نوع استفاده می کنیم

مثال: به کارمندان یک سازمان 80 نفرند. تعداد فرزندشان از 0 تا 5 است. میانگین و انحراف معیار و شاخص های مکانی را محاسبه کنید.

در این مسئله تنوع کم است ولی تعداد داده ها زیاد است.

F= frequency

| $X_i$ | $F_i$ | $F_i X_i$ | $(X_i - \mu)^2$ | $F_i(X_i-\mu)^2$ | $f_i$ فراوانی نسبی | $P_i$ درصد فراوانی نسبی | $F_c$ کمتر از | $F_c$ بیشتر از | $f_{c,i}$ کمتر از | $f_{c,i}$ بیشتر از | $P_{c,i}$ کمتر از | $P_{c,i}$ بیشتر از |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 8 | 0 | $(0-2.18)^2 = 4.7524$ | $8(2.18)^2 = 38.0192$ | 0.1 | 10 | 8 | 80 | 0.1 | 1 | 10 | 100 |
| 1 | 12 | 12 | 1.3924 | 16.7088 | 0.15 | 15 | 20 | 72 | 0.25 | 0.9 | 25 | 90 |
| 2 | 30 | 60 | 0.0324 | 0.972 | 0.375 | 37.5 | 50 | 60 | 0.625 | 0.75 | 62.5 | 75 |
| 3 | 20 | 60 | 0.6724 | 13.448 | 0.25 | 25 | 70 | 30 | 0.875 | 0.375 | 87.5 | 37.5 |
| 4 | 8 | 32 | 3.3124 | 26.4992 | 0.1 | 10 | 78 | 10 | 0.975 | 0.125 | 97.5 | 12.5 |
| 5 | 2 | 10 | 7.9524 | 15.9048 | 0.025 | 2.5 | 80 | 2 | 1 | 0.025 | 100 | 2.5 |
| | N=80 | 174 | | 111.552 | | | | | | | | |

فراوانی تجمعی — فراوانی نسبی تجمعی — درصد فراوانی تجمعی

از معیار $f_i$ نشان می دهد — از معیار $P_i$ نشان می دهد

$$\mu = \frac{\sum F_i X_i}{N} \qquad \sigma = \sqrt{\frac{\sum F_i (X_i - \mu)^2}{N}}$$

$$\mu = \frac{174}{80} = 2.175 \simeq 2.18 \qquad \sigma = \sqrt{\frac{111.552}{80}} = 1.18$$

تعبیر سطر 5

8 نفر از 80 نفر تعداد فرزندانشان، 4 نفر است.

10٪ افراد، تعداد فرزندانشان، 4 نفر است.

78 نفر از 80 نفر تعداد فرزندانشان کمتر یا مساوی 4 نفر است.

10 نفر از 80 نفر تعداد فرزندانشان بیشتر یا مساوی 4 نفر است.

97.5 درصد افراد، تعداد فرزندانشان کمتر یا مساوی 4 نفر است.

12.5 درصد افراد، تعداد فرزندانشان بیشتر یا مساوی 4 نفر است.

نما $M_o = 2$ — چون این طبقه بیشترین فراوانی را دارد

$\frac{N}{2} = 40$ — درستون فراوانی تجمعی کمتر از، اگر خودش بود، همان را انتخاب می‌کنیم ولی اگر بین دو طبقه قرار گرفت، طبقه پائین‌تر را انتخاب می‌کنیم.

میانه $M_d = 2$

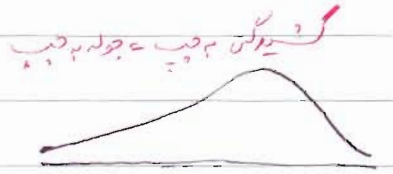
$\frac{N}{4} = 20$ — درستون فراوانی تجمعی کمتر از، اگر بین دو طبقه قرار گرفت، طبقه پائین‌تر را انتخاب می‌کنیم

چارک اول $Q_1 = 1$

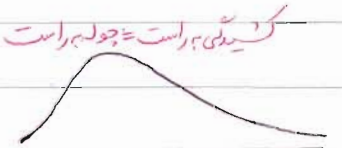
$\frac{3N}{4} = 60$ — درستون فراوانی تجمعی کمتر از، اگر بین دو طبقه قرار گرفت، طبقه پائین‌تر را انتخاب می‌کنیم

چارک سوم $Q_3 = 3$

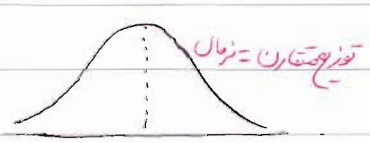
نکته ★ رابطه بین میانگین و میانه و نما:




$\mu < M_d < M_o$

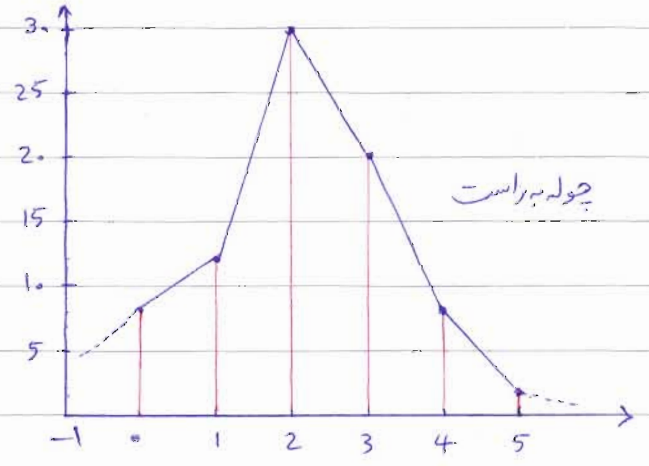



$\mu > M_d > M_o$




$\mu = M_o = M_d$

★ نمودار داده‌های با فراوانی و بدون فاصله طبقات:




نمودار مقابل یک مقداری کشیدگی به راست دارد چون در مسئله فوق

$\mu > M_d > M_o$

معمولاً در ادارات محور افقی را سال و محور عمودی را مثلاً میزان عملکرد، غیبت، میزان گندم و ... قرار می‌دهند.

## 3 داده‌های با فراوانی و با فاصله طبقات:

تنوع داده‌ها بسیار زیاد است. به همین دلیل فاصله طبقات در نظر می‌گیریم.

فرمول استورجس

معمولاً تعداد طبقات بین 5 تا 10 می‌باشد. می‌توان از فرمول زیر استفاده کرد

تعداد طبقات $K = 1 + 3.3 \log N$

دامنه تغییرات $R = x_{max} - x_{min}$

فاصله طبقات $C = \frac{R}{K}$

داده‌ها دو نوعند:
- پیوسته: جدول فراوانی باید پیوسته باشد. مانند بارش باران در مازندران
- گسسته: جدول فراوانی می‌تواند گسسته یا پیوسته شود.

مثال: بارش باران در مازندران در 60 روز به صورت می‌باشد. میانگین و انحراف معیار و شاخص‌های مکانی را محاسبه کنید.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12.1 | 0.5 | 2 | 4.1 | 0 | 3.2 | 0 | 1.8 | 10 | 5.3 | 0.4 | 3.7 |
| 8.6 | 4.6 | 0 | 3.2 | 7.5 | 8.6 | 3.4 | 4.9 | 5.7 | 4.6 | 0 | 0 |
| 3.8 | 10.5 | 11.1 | 3.7 | 0 | 0 | 4.3 | 7.5 | 8.6 | 5.7 | 0 | 0 |
| 1.5 | 1.8 | 3.7 | 5 | 6.5 | 7.3 | 0 | 0 | 4.8 | 9 | 2 | 0 |
| 4.1 | 8.6 | 7.3 | 0 | 0 | 4.5 | 8.9 | 10 | 0 | 0 | 6.4 | 1.5 |

$k = 1 + 3.3 \log 60 = 6.87 \simeq 7$ به بالا گرد می‌شود.

$R = x_{max} - x_{min} = 12.1 - 0 = 12.1$

$C = \frac{12.1}{7} = 1.73$

حد بالای طبقه
حد پایین طبقه
فاصله طبقات

| X | چوب خط | $F_i$ (فراوانی) | $X_i$ (نماینده طبقه) | $F_iX_i$ | $(X_i-M)^2$ | $F_i(X_i-M)^2$ | $F_c$ |
|---|---|---|---|---|---|---|---|
| 0-1.73 | ~~\|\|\|\|~~ ~~\|\|\|\|~~ ~~\|\|\|\|~~ ~~\|\|\|\|~~ | 20 | $\frac{0+1.73}{2} = 0.865$ | 17.3 | 11.8 | 236 | 20 |
| 1.73-3.46 | ~~\|\|\|\|~~ \|\| | 7 | 2.595 | 18.165 | 2.907 | 20.349 | 27 |
| 3.46-5.19 | ~~\|\|\|\|~~ ~~\|\|\|\|~~ \|\|\| | 13 | 4.325 | 56.225 | 0.000625 | 0.008 | 40 |
| 5.19-6.92 | ~~\|\|\|\|~~ | 5 | 6.055 | 30.275 | 3.08 | 15.4 | 45 |
| 6.92-8.65 | ~~\|\|\|\|~~ \|\|\| | 8 | 7.785 | 62.28 | 12.145 | 97.16 | 53 |
| 8.65-10.38 | \|\|\|\| | 4 | 9.515 | 38.06 | 27.2 | 108.8 | 57 |
| 10.38-12.11 | \|\|\| | 3 | 11.245 | 33.735 | 48.233 | 144.7 | 60 |
| | | 60 | | 256.04 | | 622.417 | |

$M = \frac{256.04}{60} = 4.27 \simeq 4.3$

$\sigma = \sqrt{\frac{622.417}{60}} = 3.22$

حد پایین طبقه مددار

اختلاف فراوانی طبقه مددار با طبقه قبل

$$M_o = L_i + \frac{d_1}{d_1 + d_2}(C)$$

فاصله طبقات

اختلاف فراوانی طبقه مددار با طبقه بعد

طبقه‌ای که فراوانی بیشتر دارد انتخاب می‌کنیم.

$$M_o = 0 + \frac{(20-0)}{(20-0)+(20-7)} \times 1.73 = 1.05$$

در فراوانی تجمعی کمتر از، طبقه میانه را مشخص می‌کنیم.

$$M_d = Q_2 = L_i + \frac{\frac{N}{2} - F_{c-1}}{F_i} \times (C)$$

در این مثال، طبقه میانه طبقه سوم است.

$$\frac{N}{2} = \frac{60}{2} = 30 \qquad M_d = 3.46 + \frac{30-27}{13}(1.73) = 3.86$$

50٪ روزها بارندگی از 3.86 کمتر شده و 50٪ روزها بارندگی از این مقدار بیشتر شده است.

$$Q_1 = L_i + \frac{\frac{N}{4} - F_{c-1}}{F_i} \times (C)$$

$$\frac{N}{4} = \frac{60}{4} = 15 \qquad Q_1 = 0 + \frac{15-0}{20} \times 1.73 = 1.3$$

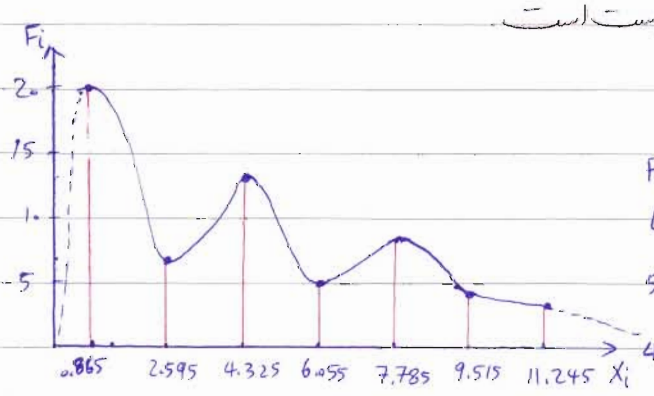
25٪ روزها بارندگی از 1.3 میلیمتر کمتر بوده و 75٪ روزها با بارندگی از این مقدار بیشتر بوده است.

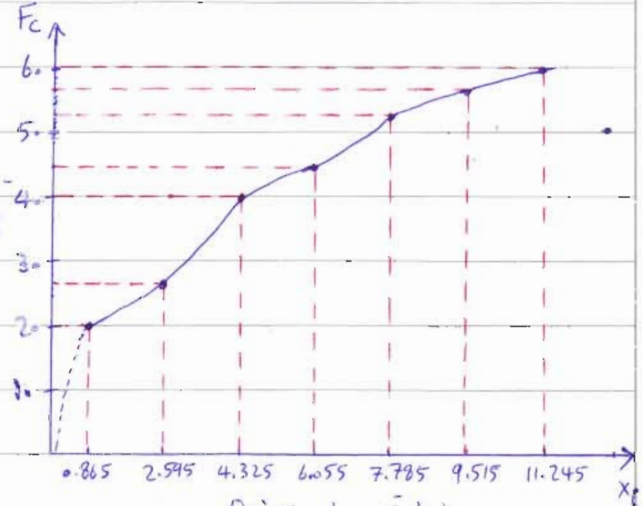
$$Q_3 = L_i + \frac{\frac{3N}{4} - F_{c-1}}{F_i} \times (C)$$

$$\frac{3N}{4} = \frac{3 \times 60}{4} = 45 \qquad Q_3 = 5.19 + \frac{45-40}{5} \times 1.73 = 6.92$$

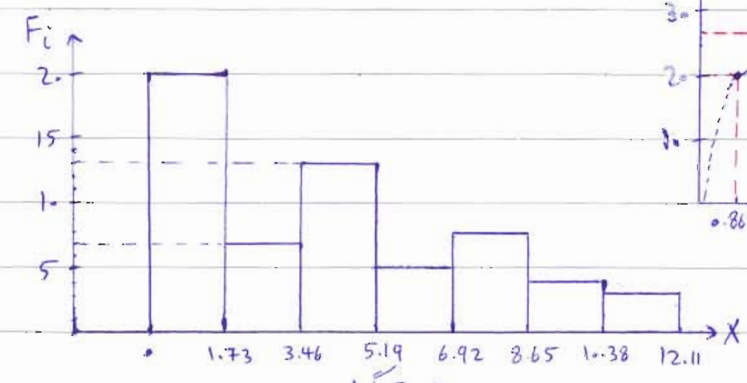
75٪ روزها بارندگی از 6.92 میلیمتر کمتر بوده و 25٪ روزها بارندگی از این مقدار بیشتر بوده است.

در این مثال $\mu > M_d > M_o$ پس نمودار چوله به راست است.




نمودار تجمعی یا Ogive




هیستوگرام

در آمار توصیفی با سه نوع شاخص مواجه هستیم:
- شاخص مرکزی
  - میانگین حسابی
  - میانگین هندسی ← زمانی که بخواهیم رشد را محاسبه کنیم
  - میانگین هارمونیک
- شاخص مکانی
  - میانه
  - چارک اول
  - چارک سوم
- شاخص پراکندگی

$$M > H > G_M$$

میانگین حسابی ($M$) — میانگین هارمونیک ($H$) — میانگین هندسی ($G_M$)

**میانگین هندسی:** هرگاه بخواهیم میانگین رشد را بدست آوریم (میانگین افزایش، میانگین کاهش یا روند)، از میانگین هندسی استفاده می‌کنیم.

$$G_M = \sqrt[N]{x_1 \cdot x_2 \cdots x_n} \Rightarrow$$ برای داده‌های بدون فراوانی

$$G_M = \sqrt[N]{x_1^{F_1} \cdot x_2^{F_2} \cdots x_n^{F_n}} \Rightarrow$$ برای داده‌های با فراوانی

**مثال:** تولیدات کارخانه‌ای از سال 1384 تا 1391 به صورت زیر است. میانگین رشد را محاسبه کنید.

$G_M$ یعنی میانگین هر سال به سال قبل یا هر تاریخ نسبت به تاریخ قبل

$M$ یعنی به طور متوسط در این سالها

| سال | تولید $\times 1000$ | تغییرات |
|---|---|---|
| 1384 | 2.5 | — |
| 1385 | 2.8 | $2.8/2.5 = \%12 = 1.12$ |
| 1386 | 3 | $3/2.8 = \%7.1 = 1.071$ |
| 1387 | 2.5 | $2.5/3 = -\%16.7 = 0.833$ |
| 1388 | 2.6 | $2.6/2.5 = \%4 = 1.04$ |
| 1389 | 3.2 | $3.2/2.6 = \%23.1 = 1.231$ |
| 1390 | 3 | $3/3.2 = -\%6.25 = 0.9375$ |
| 1391 | 3.6 | $3.6/3 = \%20 = 1.2$ |

$M=8$ ولی اولی حساب نمی‌شود

برای اینکه جایگاه منفی‌ها حفظ شود، همه را با 1 جمع کرده و در آخر از جواب 1 کم می‌کنیم

$$G_M = \sqrt[7]{\frac{2.8}{2.5} \times \frac{3}{2.8} \times \frac{2.5}{3} \times \cdots \times \frac{3.6}{3}} = \sqrt[7]{\frac{3.6}{2.5}} - 1 = 5.3\%$$

$$G_M = \sqrt[7]{1.12 \times 1.071 \times 0.833 \times \cdots} - 1 = 5.3\%$$

$$M = \%6.3$$

تغییرات امسال به پارسال: $\frac{\text{امسال}}{\text{پارسال}} - 1 = \frac{120}{100} - 1 = \%20$

تغییرات پارسال به امسال: $\frac{\text{پارسال}}{\text{امسال}} - 1 = \frac{100}{120} - 1 = -\%16.7$

★ اگر یکی از سالها نبود ۰، روند در نظر می‌گیریم.

**میانگین هارمونیک:**

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \Rightarrow$$ برای داده‌های بدون فراوانی

$$H = \frac{N}{\frac{F_1}{x_1} + \frac{F_2}{x_2} + \cdots + \frac{F_n}{x_n}} \Rightarrow$$ برای داده‌های با فراوانی

مثال: سهام 3 شرکت آلفا و بتا و گاما را می‌خریم. میانگین قیمتی که برای هر سهم پرداخت کردیم، چقدر است؟

| شرکت | سهم | قیمت |
|---|---|---|
| آلفا | 100 | 200 |
| بتا | 150 | 300 |
| گاما | 200 | 150 |

$$M \text{ میانگین حسابی} = \frac{100\times 200 + 150\times 300 + 200\times 150}{450} = 211.11$$

مثال: سهام شرکت آلفا را می‌خریم. بابت هر سهم شرکت آلفا چقدر پرداخت کرده‌ایم؟

| شرکت | سهم | قیمت |
|---|---|---|
| آلفا | 100 | 200 |
| | 150 | 300 |
| | 200 | 150 |

$$H \text{ میانگین هارمونیک} = \frac{450}{\frac{100}{200} + \frac{150}{300} + \frac{200}{150}} = 192.8$$

شاخص پراکندگی:

مثال: چهار شرکت داریم که میانگین تولید محصول و انحراف معیار آن را داریم. توزیع کدام شرکت مناسب‌تر است؟

توزیع شرکتی مناسب‌تر است که میانگین آن بالا باشد و پراکندگی پایینی باشد.

| شرکت | $\bar{x}$ | S |
|---|---|---|
| A | 20 | 3 |
| B | 24 | 5 |
| C | 28 | 6 |
| D | 23 | 5.5 |

شرکت D حذف می‌شود. چون میانگین آن از شرکت B پایین‌تر و انحراف معیار آن نسبت به شرکت B بالاتر است.

در بقیه شرکت‌ها ضریب تغییرات (CV) را محاسبه می‌کنیم

$$CV = \frac{S}{\bar{x}} \times 100$$

$$CV_1 = \frac{3}{20}\times 100 = \%15 \qquad CV_2 = \frac{5}{24}\times 100 = \%20.8 \qquad CV_3 = \frac{6}{28}\times 100 = \%21.4$$

توزیع شرکت A بهتر از شرکت B و توزیع شرکت B بهتر از شرکت C می‌باشد.

بدون فراوانی: $\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{N}}$ ($\sigma$: جامعه، $x_i$: ارزش داده)

با فراوانی: $\sigma = \sqrt{\frac{\sum F_i(x_i-\mu)^2}{N}}$ ($\sigma$: جامعه، $x_i$: نماینده طبقه) — وقتی از جامعه به جامعه می‌رسیم

بدون فراوانی: $S = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$ (S: نمونه‌گیری)

با فراوانی: $S = \sqrt{\frac{\sum F_i(x_i-\bar{x})^2}{n-1}}$ (S: نمونه‌گیری) — وقتی از جامعه به نمونه می‌رسیم

با فراوانی: $\bar{x} = \frac{\sum F_i x_i}{n}$

بدون فراوانی: $\bar{x} = \frac{\sum x_i}{n}$

متوسط انحراف هر داده از میانگین: $M_D = \frac{\sum |x_i-\bar{x}|}{n}$

مثال: ده لوله از هر دستگاه برش انتخاب می‌کنیم. میانگین، انحراف معیار و متوسط انحراف هر داده از میانگین را محاسبه کنید.
(داده‌های بدون فراوانی)

قطر لوله: 20.2   21.3   20.4   20.5   21   21.1   20.7   20.4   20.2   20

چون از هر دستگاه، 10 لوله انتخاب شده، پس نمونه‌گیری است.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{20.2 + 21.3 + \cdots + 20}{10} = 20.58$$

| $(x_i-\bar{x})^2$ | $\lvert x_i-\bar{x}\rvert$ |
|---|---|
| $(20.2-20.58)^2=0.1444$ | 0.38 |
| $(21.3-20.58)^2=0.5184$ | 0.72 |
| $(20.4-20.58)^2=0.0324$ | 0.18 |
| $(20.5-20.58)^2=0.0064$ | 0.08 |
| $(21-20.58)^2=0.1764$ | 0.42 |
| $(21.1-20.58)^2=0.2704$ | 0.52 |
| $(20.7-20.58)^2=0.0144$ | 0.12 |
| $(20.4-20.58)^2=0.0324$ | 0.18 |
| $(20.2-20.58)^2=0.1444$ | 0.38 |
| $(20-20.58)^2=0.3364$ | 0.58 |
| 1.676 | 3.56 |

$$S=\sqrt{\frac{1.676}{10-1}}=0.43$$

$$M_D=\frac{3.56}{10}=0.356$$

تفاوت انحراف معیار با متوسط انحراف هر داده از میانگین:

انحراف معیار، پراکندگی داده ها حول میانگین را نشان می دهد ولی متوسط انحراف هر داده از میانگین، اختلاف هر داده از میانگین را نشان می دهد.

$\underbrace{\text{}}$ داده ها حول میانگین، چقدر از یکدیگر دور یا نزدیک هستند.

مثال: در جدول زیر زمان تأخیر مدیر طی ۵۰ روز آمده است. میزان میانگین و انحراف معیار را حساب کنید.
(داده های با فراوانی و بدون فاصله طبقات)

در اینجا $x_i$ می گیریم

در اینجا که می خواهیم تا میانه طبقات را حساب کنیم، $x_i$ می گیریم

| زمان تأخیر به دقیقه | F | $x_i$ | $F_i x_i$ | $(x_i-\bar{x})^2$ | $F_i(x_i-\bar{x})^2$ |
|---|---|---|---|---|---|
| 0-10 | 8 | 5 | 40 | 253.44 | 2027.52 |
| 10-20 | 12 | 15 | 180 | 77.44 | 929.28 |
| 20-30 | 15 | 25 | 375 | 1.44 | 21.6 |
| 30-40 | 10 | 35 | 350 | 125.44 | 1254.4 |
| 40-50 | 3 | 45 | 135 | 449.44 | 1348.32 |
| 50-60 | 2 | 55 | 110 | 973.44 | 1946.88 |
| | 50 | | 1190 | | 7528 |

$$\bar{x}=\frac{1190}{50}=23.8$$

$$S=\sqrt{\frac{7528}{50-1}}=12.4$$

مثلاً وقتی تعداد فرزندان کارمندان را در جدول می نویسیم، $x_i$ می گذاریم.

| تعداد فرزندان کارمندان $x_i$ | $F_i$ |
|---|---|
| 1 | 13 |
| 2 | 12 |
| 3 | 3 |
| 4 | 1 |
| 5 | 1 |

مثال: سهام 3 شرکت آلفا و بتا و گاما را می‌خریم. میانگین قیمتی که برای هر سهم پرداخت کرده‌ایم، چقدر است؟

| شرکت | سهم | قیمت |
|---|---|---|
| آلفا | 100 | 200 |
| بتا | 150 | 300 |
| گاما | 200 | 150 |

$$\text{میانگین حسابی } M = \frac{100\times 200 + 150\times 300 + 200\times 150}{450} = 211.11$$

مثال: سهام شرکت آلفا را می‌خریم. بابت هر سهم شرکت آلفا چقدر پرداخت کرده‌ایم؟

| شرکت | سهم | قیمت |
|---|---|---|
| آلفا | 100 | 200 |
| | 150 | 300 |
| | 200 | 150 |

$$\text{میانگین هارمونیک } H = \frac{450}{\frac{100}{200} + \frac{150}{300} + \frac{200}{150}} = 192.8$$

شاخص پراکندگی:

مثال: چهار شرکت داریم که میانگین تولید محصول و انحراف معیار آن را داریم. توزیع کدام شرکت مناسب‌تر است؟

| شرکت | $\bar{x}$ | $S$ |
|---|---|---|
| A | 20 | 3 |
| B | 24 | 5 |
| C | 28 | 6 |
| D | 23 | 5.5 |

توزیع شرکتی مناسب‌تر است که میانگین آن بالا باشد و پراکندگی پایین باشد.

شرکت D حذف می‌شود. چون میانگین آن از شرکت B پایین‌تر و انحراف معیار آن نسبت به شرکت B بالاتر است.

در بقیه شرکت‌ها، ضریب تغییرات (CV) را محاسبه می‌کنیم

$$CV = \frac{S}{\bar{x}} \times 100$$

$CV_1 = \frac{3}{20}\times 100 = \%15$ $\quad CV_2 = \frac{5}{24}\times 100 = \%20.8$ $\quad CV_3 = \frac{6}{28}\times 100 = \%21.4$

توزیع شرکت A بهتر از شرکت B و توزیع شرکت B بهتر از شرکت C می‌باشد.

بدون فراوانی $\underset{\text{جامعه}}{\sigma} = \sqrt{\frac{\sum (\overset{\text{ارزش داده}}{x_i} - \mu)^2}{N}}$ $\qquad$ با فراوانی $\underset{\text{جامعه}}{\sigma} = \sqrt{\frac{\sum F_i (\overset{\text{نماینده طبقه}}{x_i} - \mu)^2}{N}}$ $\qquad$ وقتی از جامعه به جامعه می‌رسیم

بدون فراوانی $\underset{\text{نمونه‌گیری}}{S} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ $\qquad$ با فراوانی $\underset{\text{نمونه‌گیری}}{S} = \sqrt{\frac{\sum F_i (x_i - \bar{x})^2}{n-1}}$ $\qquad$ وقتی از جامعه به نمونه می‌رسیم

با فراوانی $\bar{x} = \frac{\sum F_i x_i}{n}$

بدون فراوانی $\bar{x} = \frac{\sum x_i}{n}$ $\qquad$ متوسط انحراف هر داده از میانگین $M_D = \frac{\sum |x_i - \bar{x}|}{n}$

مثال: ده لوله از هر دستگاه برش انتخاب می‌کنیم. میانگین، انحراف معیار و متوسط انحراف هر داده از میانگین را محاسبه کنید.
(داده‌های بدون فراوانی)

قطر لوله: 20.2 21.3 20.4 20.5 21 21.1 20.7 20.4 20.2 20

چون از هر دستگاه، 10 لوله انتخاب شده، پس نمونه‌گیری است.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{20.2 + 21.3 + \cdots + 20}{10} = 20.58$$

## احتمالات :

در آمار توصیفی ، اگر داده هایی داریم ، فقط در مورد همان داده ها قضاوت می کنیم .

در احتمالات ، بر اساس گذشته چون با متغیر تصادفی مواجه هستیم

متغیر تصادفی : می تواند هر یک از ارزش ها را با احتمالات مختلف دریافت کند .

مثال : تولید در خط تولید یک کارخانه به صورت جدول زیر آمده است . احتمال هر خط تولید محاسبه شود .

| تولید $x_i$ | روز $F$ | احتمال $P_i$ |
|---|---|---|
| 15 | 10 | 0.2 |
| 18 | 12 | 0.24 |
| 20 | 13 | 0.26 |
| 25 | 15 | 0.3 |
| | 50 | |

احتمال $P = \frac{x}{N}$ → حالات مورد نظر / کل پیشامد

مثال : در یک کلاس 35 نفره ، تعداد دانشجویان خانم ، 10 نفر می باشد . احتمال در یک انتخاب خانم انتخاب شود و احتمال اینکه در یک انتخاب آقا انتخاب شود ، چقدر است ؟

احتمال انتخاب خانم $P(A) = \frac{10}{35}$ احتمال انتخاب آقا $P(B) = \frac{25}{35}$

نتیجه : 2 اصل مهم در احتمالات :

1) $0 \leq P(A) \leq 1$

2) $P(A) + P(B) = 1$

مثال : از 1000 محصول ، 20 تای آن خراب است . اگر کسی که این محصولات را می خرد ، محصول خراب را بخرد .

$N = 1000$ $x = 20$ $P(A) = \frac{20}{1000} = \%2$

* سه اصل مهم در احتمالات :

1. اصل ضرب ، 2. اصل جمع ، 3. قضیه شرطی .

1. اصل ضرب . اگر بخواهیم احتمال توأم دو واقعه را با هم بسنجیم ، اصل ضرب است .

$P(A \cap B) = P(A)\,p(B)$ اگر مستقل باشند ، از رابطه روبرو استفاده می شود .

$P(A \cap B) = P(A)\,p(B|A)$ اگر وابسته باشند ، از رابطه روبرو استفاده می شود .

پدیده های مستقل مثل زاس ، سکه و ... اما پدیده های اجتماعی وابسته اند چون معمولاً تحت تأثیر واقعه قبل خود قرار می گیرد .

مثال: ۱۰٪ آقایان و ۱۵٪ خانم‌ها نازا هستند. احتمال اینکه یک خانم و آقا که با هم ازدواج می‌کنند نازا باشند چقدر است؟

$P(A) = ۱۰\%$

$P(B) = 15\%$ $\Longrightarrow P(A \cap B) = \%۱۰ \times \%15 = 0.015 = 1.5\%$ نازا بودن خانم و آقا از هم مستقل است

مثال: 3٪ خانم‌ها تحصیلکرده‌اند، 20٪ آنها مو بلوندند، 15٪ آنها تک فرزندند، 20٪ آنها پولدارند. احتمال اینکه یک آقا با خانمی ازدواج کند که تحصیلکرده، مو بلوند، تک فرزند و پولدار باشد، چقدر است؟

$P(A \cap B \cap C \cap D) = 0.3 \times 0.2 \times 0.15 \times 0.2 = 0.0018$

* وابسته زمانی است که از یک جمعی بدون جایگذاری، انتخاب کنیم.

مثال: در یک اداره ۱۰ کارمند است که 2 نفر آنها زن و شوهر است. می‌خواهیم 2 کارمند انتخاب کنیم که به سفر زیارتی بروند. احتمال اینکه زن و شوهر انتخاب شوند، چقدر است؟

$P(A \cap B) = \frac{2}{10} \times \frac{1}{9} = \frac{1}{45}$

مثال: در یک مهمانی 20 نفر دعوتند. در هنگام برگشت هوا تاریک است. احتمال اینکه در هنگام پوشیدن کفش‌ها، کفش درست انتخاب شود؟

$P(A \cap B) = \frac{2}{40} \times \frac{1}{39}$

2 اصل جمع:

$P(A \cup B) = P(A) + P(B)$ اگر ناسازگار باشند

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ اگر سازگار باشند

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

مثال: احتمال اینکه یک محصول بگیرید ۱۰۰٪ سالم باشد، 30٪ است. احتمال اینکه یک محصول بگیرید که در حد قابل قبول باشد، 40٪ است. احتمال اینکه معیوب باشد، 30٪ است. یک محصول انتخاب می‌کنید. احتمال اینکه ۱۰۰٪ سالم یا قابل قبول باشد، چقدر است؟

$P(A) = 30\%$ $P(B) = 40\%$ $P(C) = 30\%$

$P(A \cup B) = 0.3 + 0.4 = 0.7$

مثال: 70٪ ایرانی‌ها در طول عمرشون به خارج می‌روند. 30٪ آنها سفرهای زیارتی می‌روند. 20٪ آنها هم سفر زیارتی و هم خارجی می‌روند. چند درصد ایرانی‌ها به هیچ کدام از این سفرها نمی‌روند؟

$P(A \cup B) = 0.7 + 0.3 - 0.2 = 0.8 \Longrightarrow$ یا سفر زیارتی می‌روند یا سفر خارجی

$1 - P(A \cup B) = 1 - 0.8 = 0.2 = 20\% = P(A \cup B)'$

3 قضیه شرطی: اگر پدیده‌ای به پدیده دیگری وابسته باشد، احتمال وقوع این پدیده در صورت وقوع اولی چقدر است.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

مثال: آقای خانه ۸۰٪ فوتبالها را می بیند، خانم خانه ۳۰٪ فوتبالها را می بیند. احتمال اینکه هر دو با هم فوتبال ببینند ۱۵٪ است. اگر امروز آقا در حال تماشای فوتبال باشد، احتمال اینکه خانم هم بیاید، چقدر است؟

$P(A) = \%80 \qquad P(B) = \%30 \qquad P(A \cap B) = \%15$

$$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{0.15}{0.8} = 0.1875$$

مثال: رئیس جمهور در ۸۰٪ جلسات هیئت دولت حضور دارد. معاونش هم در ۹۰٪ جلسات حضور دارد و در ۷۰٪ جلسات هم هر دو شرکت دارند: الف) احتمال اینکه در جلسات یکی از این دو نفر باشند، چقدر است؟ ب) احتمال اینکه در جلسه ای که رئیس جمهور شرکت می کند، معاونش هم باشد، چقدر است؟

$P(A) = \%80 \qquad P(B) = \%90 \qquad P(A \cap B) = \%70$

الف) $P(A\cup B) = P(A) + P(B) - P(A\cap B) = 0.8 + 0.9 - 0.7 = 1 = 100\%$

ب) $P(B|A) = \dfrac{P(A\cap B)}{P(A)} = \dfrac{0.7}{0.8} = 0.875$ یعنی در ۸۷.۵٪ جلساتی که رئیس جمهور است، معاونش هم حضور دارد.

امید ریاضی:

$E(x) = \sum x.p(x)$ میانگین

$V(x) = E(x^2) - (E(x))^2$ واریانس

مثال: در یک کارخانه نتایج حاصل از بررسی مسئول کنترل کیفیت به شرح زیر است. متوسط ضایعات مورد انتظار چقدر است؟ پراکندگی را هم محاسبه کنید.

ضایعات

| $x$ | $P(x)$ |
|---|---|
| 0.01 | 0.2 |
| 0.02 | 0.3 |
| 0.03 | 0.35 |
| 0.04 | 0.15 |

$E(x) = 0.01\times0.2 + 0.02\times0.3 + 0.03\times0.35 + 0.04\times0.15 = 0.0245$

$E(x^2) = 0.01^2\times0.2 + 0.02^2\times0.3 + 0.03^2\times0.35 + 0.04^2\times0.15 = 0.000695$

$V(x) = 0.000695 - 0.0245^2 = 0.00009475$

$S(x) = 0.0097$

دو نوع توزیع داریم
- توزیع گسسته: دوجمله ای، پواسون، هندسی، فوق هندسی، دوجمله منفی، یکنواخت
- توزیع پیوسته: نرمال، نمایی، وایبال

★ اگر تعداد پدیده ها از 3 بیشتر شد، دیگر سراغ اصل ضرب و جمع نمی رویم. از توزیع ها استفاده می کنیم.

توزیع هایی که مورد استفاده ما هستند، دوجمله ای، پواسون و نرمال می باشد.

ما از این توزیع ها میانگین، نسبت و انحراف معیار را می خواهیم.

1 توزیع دوجمله‌ای: توزیع دوجمله‌ای نسبت را نشان می‌دهد.

مثال: 3٪ ازدواج های گروهی به طلاق منجر می‌شود. 100 نمونه انتخاب می‌کنیم. احتمال اینکه بیش از 10 تای آن به طلاق منجر شود، چقدر است؟

$$p(n=100,\ p=\%30,\ x>10)\ \begin{cases} \longrightarrow np\leq 5 & \text{پوآسون} \\ \longrightarrow np>5 & \text{نرمال} \end{cases}$$

در جدول n=100 نداریم.
اگر n از 30 بیشتر شده، np را می‌سنجیم

2 توزیع پوآسون:

مثال: روزانه به طور متوسط در تهران 2 نفر خودکشی می‌کنند. 1) احتمال اینکه فردا بیش از 3 نفر در تهران خودکشی کنند، چقدر است؟

2) احتمال اینکه در یک ماه آینده، بیش از 30 نفر خودکشی کنند، چقدر است؟

1) $p(\lambda=2,\ x>3)$

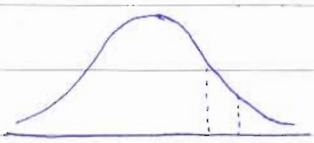
2) $p(\lambda=2\times 30,\ x>30)$ $\longrightarrow$ $\lambda>10$ نرمال

اگر $\lambda\leq 10$ بود خودش را حساب می‌کنیم.
اگر $\lambda>10$ شد، نرمال حساب می‌کنیم

نکته: اگر در توزیع دوجمله‌ای $np>5$ و در توزیع پوآسون $\lambda>10$ شود، هر دو توزیع به نرمال تبدیل می‌شود. پس توزیع گسسته به پیوسته تبدیل می‌شود

3 توزیع نرمال:

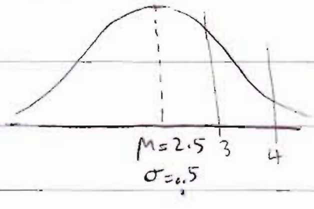
$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



در این توزیع برای بدست آوردن احتمال، باید از $f(x)$ انتگرال بگیریم که بسیار سخت است.

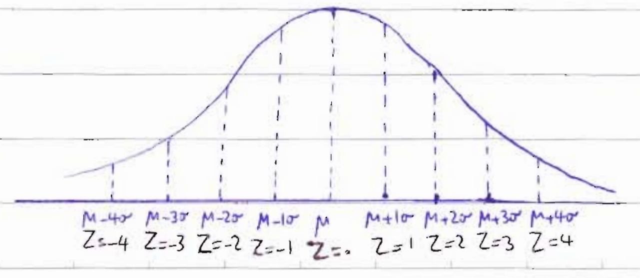
$$\int_a^b f(x)\,dx=\Big[F(x)+c\Big]_a^b=F(b)-F(a)$$

مثال: میانگین مقاومت یک نوع پلاستیک در مقابل ضربه 2.5 کیلوگرم با انحراف معیار 0.5 کیلوگرم می‌باشد. چند درصد این پلاستیک‌ها 3 تا 4 کیلوگرم مقاومت دارند؟




$$\int_3^4 \frac{1}{\sqrt{2\pi}\times 0.5}\,e^{\frac{-(x-2.5)^2}{2\times 0.5^2}}\,dx$$

انتگرال گرفتن از رابطه بالا بسیار سخت است. پس واحدی درست کردند به نام واحد Z استاندارد نرمال. هر چی داده داریم را به Z تبدیل می‌کنیم. کل منحنی به 8 قسمت تقسیم می‌شود که هر قسمت با قسمت بعد 1σ فاصله دارد که به آن 1σ، Z می‌گوییم.




$$Z=\frac{x-\mu}{\sigma}$$

یعنی اختلاف ما از میانگین، چند درصد انحراف معیار است.

آمار استنباطی:
- پارامتریک ⟸ نرمال است یا تابع توزیع مشخص است.
- ناپارامتریک ⟸ غیرنرمال است.

پارامتریک باید تصادفی باشد. چهار نوع تصادفی داریم:

1. تصادفی ساده.
2. تصادفی سیستماتیک
3. تصادفی خوشه‌ای (برای مدیریت صنعتی بیشتر استفاده می‌شود)

بایدخودمون در شرکت یا کارخانه یا سازمان حضور داشته باشیم و اطلاعات دست اول بگیریم.

4. تصادفی طبقه‌ای (بیشتر برای مدیریت مالی استفاده می‌شود)

استاد 800 — 10,000 — 2500

$N = 13,300$ $\quad n = 500$

$\frac{10,000}{13,300} \times 500 \qquad \frac{2500}{13,300} \times 500 \qquad \frac{800}{13,300} \times 500$

قضیه: اگر جامعه نرمال باشد، نمونه تصادفی در هر حجمی نرمال است.

نکته: اگر از جامعه اطلاعاتی نداشته باشیم و $\underbrace{n>30}_{\text{حجم نمونه}}$ باشد، به شرط تصادفی بودن نمونه، توزیع نرمال می‌شود.

نرمال:
- جامعه نرمال ⟸ نمونه در هر حجمی نرمال
- اطلاعی از جامعه نرمال نداریم $\xRightarrow{n>30}$ نمونه به شرط تصادفی بودن، نرمال است.
- $\chi^2$ (پارامتریک)
- کالموگراف - اسمیرنف (Kalmograph - Smirinov) (ناپارامتریک)

* می‌خواهیم از جامعه به نمونه برسیم. در بسیاری از شرایط، هدف اصلی تخمین ویژگی شمارش جامعه آماری، به عبارتی پارامترها است.

پارامتر: اطلاعاتی که از جامعه بدست می‌آید. $\begin{cases} \mu \\ \sigma \\ P \end{cases}$

آماره: اطلاعاتی که از نمونه بدست می‌آید. $\begin{cases} \bar{x} \\ s \\ p \end{cases}$

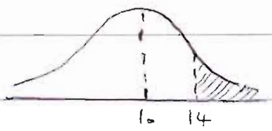
هدف در آمار: چگونه از آماره به پارامتر برسیم یعنی چگونه می‌توانیم آماره را به پارامتر تعمیم دهیم.

مثال: میانگین خواب دانشجویان ۱۰ ساعت و انحراف معیار آن 2 ساعت است. احتمال اینکه یک نفر بالای ۱۴ ساعت بخوابد؟

از ۵۰۰ نفر چند نفر بالای 14 ساعت می‌خوابد؟

$\mu = 10 \qquad \sigma = 2$

$p(x>14)=p\left(Z>\frac{14-10}{2}\right)=p(Z>2)=1-p(Z<2)=1-0.9772=0.0228$




از ۱۰۰۰۰ نفر، 228 نفر بالای 14 ساعت می‌خوابند.

مثال: از جامعه‌ای به حجم ۱۰۰۰۰، نمونه‌ای ۱۰۰ تایی انتخاب می‌کنیم. به چند طریق امکان‌پذیر است؟

$N=10{,}000 \qquad n=100$

$$C^{10{,}000}_{100}=\frac{10000!}{9900!\,100!}>1{,}000{,}000{,}000$$

توزیع نمونه‌گیری $\bar{x}$ - نرمال:

$\mu_x=\mu_{\bar{x}}$

غیرنرمال $\sigma_{\bar{x}}=\frac{\sigma_x}{\sqrt{n}}\times\sqrt{\frac{N-n}{N-1}}$ نرمال $\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$ $\quad n\geq 30$

مثال: 4 نفر داریم با سنین زیر. میانگین سن جامعه 4 نفری و میانگین نمونه‌های 2 تایی از آنها را به دست آورید.

30 35 38 41

در این مسئله $\mu=\frac{30+35+38+41}{4}=\frac{144}{4}=36$

$C^4_2=6$ نمونه‌های دوتایی

| | | | |
|---|---|---|---|
| 30 | 35 | | 32.5 |
| 30 | 38 | | 34 |
| 30 | 41 | $\xrightarrow{\text{از هر نمونه دوتایی میانگین می‌گیریم}}$ | 35.5 |
| 35 | 38 | | 36.5 |
| 35 | 41 | | 38 |
| 38 | 41 | | 39.5 |
| | | | 216 |

$\mu_{\bar{x}}=\frac{216}{6}=36$

اگر نمونه سه‌تایی هم داشته باشیم به همین صورت عمل می‌کنیم.

$$Z=\frac{x-\mu}{\sigma}\Rightarrow Z=\frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\Rightarrow \boxed{Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}$$

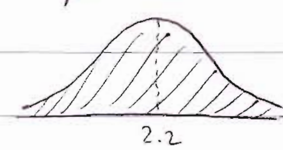
این فرمول کلی‌تر است.

مثال: میانگین مطالعه دانشجویان ایرانی در روز 2.2 ساعت و انحراف معیار آن 0.8 ساعت است.

الف) یک دانشجو به تصادف انتخاب می‌کنیم. احتمال اینکه کمتر از 3 ساعت مطالعه داشته باشد، چقدر است؟

ب) یک نمونه به حجم 16 دانشجو انتخاب می‌کنیم. احتمال اینکه میانگین زمان مطالعه آنها کمتر از 3 ساعت باشد، چقدر است؟

$\mu=2.2 \qquad \sigma=0.8$




الف) $n=1\Rightarrow p(x<3)=p\left(Z<\frac{3-2.2}{0.8}\right)=p(Z<1)=0.8413$

ب) $n=16\Rightarrow p(\bar{x}<3)=p\left(Z<\frac{3-2.2}{\frac{0.8}{\sqrt{16}}}\right)=p(Z<4)\simeq 1$

یعنی ۱۰۰٪ احتمال دارد که میانگین این 16 نفر، 3 شود

مثال: میانگین قیمت سهام شرکت صنعت A در سال گذشته ۱۲۰۰ تومان با انحراف معیار ۶۰ تومان بوده است. یک سهم از این صنعت به تصادف انتخاب می‌کنیم.

الف) احتمال اینکه کمتر از ۱۰۰۰ تومان قیمت گذاری شده باشد، چقدر است؟

ب) یک نمونه به حجم ۹ سهم به تصادف انتخاب می‌کنیم. احتمال اینکه میانگین قیمت آن سهم‌ها کمتر از ۱۰۰۰ تومان شده باشد، چقدر است؟

$\mu = 1200 \qquad \sigma = 60$

الف) $n=1 \Rightarrow p(x<1000) = p\left(Z<\frac{1000-1200}{60}\right) = p(Z<-3.33) = 0.0005$

ب) $n=9 \Rightarrow p(\bar{x}<1000) = p\left(Z<\frac{1000-1200}{\frac{60}{\sqrt{9}}}\right) = p(Z<-10) \simeq 0$

نکته: اگر توزیع دوجمله‌ای باشد

$p(n, p, x) \xrightarrow{n>30}$ $np \leq 5$ تقریب پواسون

$\qquad\qquad\qquad\quad$ $np>5$ نرمال

$\mu = np \qquad \sigma = \sqrt{npq}$

$Z = \frac{x-\mu}{\sigma} \Longrightarrow Z = \frac{x-np}{\sqrt{npq}}$

مثال: ۲۰ درصد ایرانی‌ها طرفدار آقای A هستند. ۵۰۰۰ نفر انتخاب می‌کنیم. ۱) احتمال اینکه بیشتر از ۸۰۰ نفر به ایشان وفادار باشند، چقدر است؟ ۲) احتمال اینکه دقیقاً ۸۰۰ نفر طرفدار آقای A باشند، چقدر است؟

1) $p(n=5000, p=20\%, x>800) \xrightarrow{n>30 \Rightarrow np=5000\times 0.2=1000>5 \text{ نرمال}}$

$p(x>800) = p\left(Z>\frac{800-1000}{\sqrt{5000\times 0.2\times 0.8}}\right) = p(Z>-7.07) = 1-p(Z<-7.07) = 1-0 = 1$

۱۰۰٪ بیشتر از ۸۰۰ نفر طرفدار آقای A هستند

2) $p(x=800) = p(795<x<805) = p\left(\frac{795-1000}{\sqrt{5000\times 0.2\times 0.8}} < x < \frac{805-1000}{\sqrt{5000\times 0.2\times 0.8}}\right) = p(Z<-6.89) - p(Z<-7.24)$

$= 0 - 0 = 0$

نکته: اگر دهدهی بود، ۰.۵ اضافه و کم می‌کنیم. اگر صدصدی بود، ۵ اضافه و کم می‌کنیم و به همین ترتیب.

توزیع نمونه‌گیری P نرمال:

$\mu_p = P$

غیرنرمال $\sigma_p = \sqrt{\frac{pq}{n}} \times \sqrt{\frac{N-n}{N-1}}$ $\qquad$ نرمال $\sigma_p = \sqrt{\frac{pq}{n}}$

مثال: ۲۰٪ سهام‌های پذیرفته شده امسال ضرر دادند. یک نمونه به حجم ۲۰۰ سهم انتخاب می‌کنیم.

اولاً) احتمال اینکه بیشتر از ۳۰ سهم دچار ضرر شده باشند، چقدر است؟

ثانیاً) احتمال اینکه بیشتر از ۳۰٪ سهام‌ها دچار ضرر شده باشد، چقدر است؟

$P = 20\%$ $\qquad n = 200$

تعداد $\qquad$ نسبت

اولاً) $p(x>30)$ یا $p(P>0.15) = p\left(Z > \dfrac{30-200\times 0.2}{\sqrt{200\times 0.2\times 0.8}}\right)$ یا $p\left(Z > \dfrac{0.15-0.2}{\sqrt{\dfrac{0.2\times 0.8}{200}}}\right) = p(Z>-1.77)$

$= 1 - p(Z<-1.77) = 1 - 0.0384 = 0.9616$ $\qquad$ ٪96.16 احتمال دارد از 200 تا بیشتر از 30 تا یا 15٪ ضرر ده باشد

$= P(x>60)$

ثانیاً) $p(P>0.3) = p\left(Z > \dfrac{0.3-0.2}{\sqrt{\dfrac{0.2\times 0.8}{200}}}\right) = p\left(Z > \dfrac{60-200\times 0.2}{\sqrt{200\times 0.2\times 0.8}}\right) = p(Z>3.54) = 1 - p(Z<3.54)$

$= 1 - 0.9998 = 0.0002$

جمع بندی:

از فصل 7 ، 4 موضوع را باید بلد باشیم:

1 نمونه گیری تصادفی: ساده، سیستماتیک، خوشه ای، طبقه ای

2 $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ $\qquad$ $\sigma_p = \sqrt{\dfrac{pq}{n}}$

3 $Z = \dfrac{\bar{x}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$ $\qquad$ $Z = \dfrac{p-P}{\sqrt{\dfrac{pq}{n}}}$

4 قضیه حد مرکزی $\Leftarrow$ نرمال $\qquad$ جامعه نرمال، $n \geq 30$

برآوردها:

ناپارامتریک

پارامتریک

در میانگین هم پارامتریک و هم ناپارامتریک داریم
باید به هردو وارد شویم.

- میانگین
  - یک جامعه
    - برآورد
    - آزمون
  - دو جامعه
    - برآورد
    - آزمون
  - چند جامعه
    - آزمون
- نسبت
  - یک جامعه
    - برآورد
    - آزمون
  - دو جامعه
    - برآورد
    - آزمون
  - چند جامعه
    - آزمون
- انحراف معیار
  - یک جامعه
    - برآورد
    - آزمون
  - دو جامعه
    - برآورد
    - آزمون
  - چند جامعه
    - آزمون

برآورد
- نقطه ای: $\mu = \bar{x}$ $\qquad$ $P = p$ $\qquad$ $\sigma = S$
- فاصله ای: $\mu = \bar{x} \pm e$ $\qquad$ $P = p \pm e$ $\qquad$ $< \sigma^2 <$ ?

$\bar{x} - e < \mu < \bar{x} + e$ $\qquad$ $p - e < P < p + e$

**نکته:**

$M \pm 1\sigma$ : %68

$M \pm 2\sigma$ : %95

$M \pm 3\sigma$ : %99.73

* اگر از $3\sigma$ استفاده کنیم، خوب است. چون خطا 2.7 در هزار می‌شود.
* ژاپنی‌ها از $6\sigma$ استفاده می‌کنند.

$M \pm 6\sigma$ — 3.4 در میلیون خطا

$\bar{x} = 12$ $\quad \sigma = 0.6$ $\quad \bar{x} \pm 3\sigma \to 10.2,\ 13.8$

$\bar{x} = 12$ $\quad \sigma = 0.1$ $\quad \bar{x} \pm 6\sigma \to 11.4,\ 12.6$

اگر ژاپنی‌ها از $\sigma = 0.6$ استفاده کنند، $\bar{x} \pm 6\sigma$ می‌شود $\to 8.4,\ 15.6$

پس اختلاف $6\sigma$ زیاد می‌شود، پس انحراف معیار را کاهش می‌دهند.

* تغییرپذیری را چگونه کنترل کنیم؟ ?

---

## جلسه چهارم

92,2,3

**برآورد میانگین جامعه:** پس از اینکه نمونه تصادفی از جامعه انتخاب شده، میانگین نمونه محاسبه گردیده، سپس اگر انحراف معیار جامعه مشخص باشد، از رابطه اول، و مشخص نباشد و حجم نمونه بالای 30 باشد، از رابطه دوم و کمتر از 30 باشد، از رابطه سوم استفاده می‌کنیم. هر سه رابطه برای مواقعی است که داده‌ها نرمال باشند ولی اگر نرمال نباشد، از رابطه چهارم یعنی چبی‌شف که غیر پارامتریک است، استفاده می‌شود.

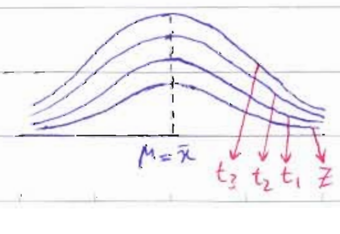
نرمال:

انحراف معیار جامعه مشخص است: $\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < M < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

انحراف معیار نمونه مشخص باشد:
$$\begin{cases} \bar{x} - Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < M < \bar{x} + Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \\ \bar{x} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < M < \bar{x} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \end{cases}$$

غیر نرمال — چبی‌شف: $\bar{x} - \sqrt{\frac{1}{\alpha}} \frac{S}{\sqrt{n}} < M < \bar{x} + \sqrt{\frac{1}{\alpha}} \frac{S}{\sqrt{n}}$

| | | | |
|---|---|---|---|
| $M$ | میانگین جامعه | $n$ | حجم نمونه |
| $\bar{x}$ | میانگین نمونه | $\alpha$ | درصد خطا |
| $\sigma$ | انحراف معیار جامعه | $1-\alpha$ | درصد اطمینان |
| $S$ | انحراف معیار نمونه | $Z_{\frac{\alpha}{2}}$ | متغیر استاندارد نرمال |
| | | $t_{\frac{\alpha}{2}}$ | همان متغیر استاندارد نرمال و برای $n < 30$ — $t_{student}$ |




* هرچه $n$ کمتر باشد، نمودار بلندتر است. پس پراکندگی بیشتر است.

$n_1 > n_2 > n_3$

مثال: می‌خواهیم میانگین وزن خانم‌های بین 20 تا 30 سال را برآورد کنیم. یک نمونه به حجم 35 خانم انتخاب می‌کنیم. میانگین وزن 54 با پراکندگی 1.2 کیلوگرم شده است. در سطح اطمینان 95%، میانگین وزن تمام خانم‌ها را برآورد کنید.

$n=35 \quad \bar{x}=54 \quad S=1.2 \quad 1-\alpha=0.95 \quad \alpha=0.05 \quad \frac{\alpha}{2}=0.025$

$$54-1.96\frac{1.2}{\sqrt{35}} < \mu < 54+1.96\frac{1.2}{\sqrt{35}} \Longrightarrow 53.6<\mu<54.4$$

* در اینجا نگفتیم جامعه نرمال است. چون حجم نمونه از 30 بیشتر است.

نکته: اگر همه داده‌ها برابر باشد، انحراف معیار صفر است.

حفظ شود:

| $1-\alpha$ | $Z_{\frac{\alpha}{2}}$ |
|---|---|
| 90% | 1.645 |
| 95% | 1.96 |
| 99% | 2.576 |

نکته:

$\sigma=0.05 \qquad \mu\pm6\sigma \qquad$ 3.4 در میلیون

$\sigma=0.3 \qquad \mu\pm3\sigma \qquad$ 2.7 در هزار

مثال: در نظر داریم میانگین قطر لوله‌های تولیدی یک کارخانه را برآورد نماییم. یک نمونه به حجم 25 محصول انتخاب می‌کنیم. میانگین 25 میلیمتر با انحراف معیار 0.8 میلیمتر شده است. با فرض توزیع نرمال، میانگین را برآورد نمائید.

نکته: برای اینکه به جدول $t_{student}$ مراجعه نمائیم، می‌بایستی درجه آزادی ($df=n-1$) را بدست آورده، با در اختیار بودن $\frac{\alpha}{2}$ به جدول $t_{student}$ مراجعه نمائیم.

* اگر در مسئله $\alpha$ قید نشود، همان 5% است.

$n=25 \quad \bar{x}=25mm \quad S=0.8\,mm \quad \alpha=0.05 \quad \frac{\alpha}{2}=0.025$

$df=n-1=25-1=24 \qquad t_{24,0.025}=2.064$

$$25-2.064\frac{0.8}{\sqrt{24}} < \mu < 25+2.064\frac{0.8}{\sqrt{24}} \Longrightarrow \underbrace{24.67}_{\text{ضایعات}} < \mu < \underbrace{25.33}_{\text{بازیافت}}$$

در خط تولید کمتر، ضایعات و بیشتر، بازیافت است.

مثال: برای اینکه متراژ منزل دانشجویان پردیس را برآورد کنیم، یک نمونه به حجم 15 دانشجو انتخاب می‌کنیم. سطح زیربنای منزلشان در جدول آمده است. الف) با فرض توزیع نرمال و ب) بدون هیچ پیش فرضی، در سطح $\alpha=1\%$ برآورد کنید.

* (در آمار بهتر است داده‌های پرت حذف شود) ولی در اینجا این کار را نمی‌کنیم.

| $x_i$ | $(x_i-\bar{x})^2$ |
|---|---|
| 140 | 6.25 |
| 250 | 12656.25 |
| 100 | 1406.25 |
| 110 | 756.25 |
| 85 | 2756.25 |
| 90 | 2256.25 |
| 220 | 6806.25 |
| 160 | 506.25 |
| 107 | 930.25 |
| 110 | 756.25 |
| 150 | 156.25 |
| 150 | 156.25 |
| 56 | 6642.25 |
| 180 | 1806.25 |
| 155 | 306.25 |
| 2063 | 37903.75 |

$\frac{\alpha}{2}=0.005 \qquad df=14 \qquad t_{14,\,0.005}=2.977$

$\bar{x}=\frac{2063}{15}=137.5$

$S=\sqrt{\frac{37903.75}{15-1}}=52.03$

الف) $137.5-2.977\frac{52.03}{\sqrt{15}}<\mu<137.5+2.977\frac{52.03}{\sqrt{15}}$

$\Rightarrow 97.5<\mu<177.5$

ب) $137.5-\sqrt{\frac{1}{0.01}}\,\frac{52.03}{\sqrt{15}}<\mu<137.5+\sqrt{\frac{1}{0.01}}\,\frac{52.03}{\sqrt{15}}$

$\Rightarrow 3.2<\mu<271.8$

مثال: پراکندگی زمان توقف در ایستگاه های اتوبوس، طبق آمار اتوبوسرانی 5 دقیقه اعلام شده است. اگر 50 مسافر به تصادف انتخاب نمائیم و میانگین زمان توقف آنها 25 دقیقه باشد، در سطح اطمینان 90٪ حدود توقف را برآورد نمائید و اگر این داده ها مربوط به 25 نمونه بوده باشد، آیا در حدود بدست آمده تغییری ایجاد می شود؟

$\sigma=5 \qquad \bar{x}=25$

$n>30$

$n=50 \Rightarrow 25-1.645\frac{5}{\sqrt{50}}<\mu<25+1.645\frac{5}{\sqrt{50}} \Rightarrow 23.84<\mu<26.16$

$n=25 \Rightarrow 25-1.645\frac{5}{\sqrt{25}}<\mu<25+1.645\frac{5}{\sqrt{25}} \Rightarrow 23.355<\mu<26.645$

چون واریانس جامعه داده شده است، در $n=25$ هم از $Z$ استفاده می کنیم.

مثال: در یک نمونه به حجم 25، انحراف معیار 3.2 و حدود برآورد $23.5<\mu<26.5$ شده است. سطح اطمینان و میانگین را بدست آورید.

$n=25 \qquad S=3.2$

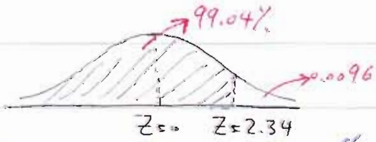
$\bar{x}=\frac{23.5+26.5}{2}=25$

$\bar{x}-e=23.5$

$\bar{x}+e=26.5 \quad \Rightarrow e=1.5$

$Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}=1.5 \Rightarrow Z_{\frac{\alpha}{2}}\frac{3.2}{\sqrt{25}}=1.5 \Rightarrow Z_{\frac{\alpha}{2}}=\frac{7.5}{3.2}=2.34 \Rightarrow Z_{2.34}=0.9904$

$\frac{\alpha}{2}=0.0096 \Rightarrow \alpha=0.0192 \Rightarrow 1-\alpha=0.9808$




مثال: میانگین برآورد شده از یک نمونه 64 لاستیک که در مورد عمر لاستیک ها بوده، به صورت $40000<\mu<50000$ با سطح اطمینان 95٪ حاصل شده است. میانگین و انحراف معیار نمونه را بدست آورید.

$\bar{x}=\frac{50000+40000}{2}=45000$

$$45{,}000 + e = 50{,}000 \Rightarrow e = 5{,}000 \Rightarrow Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} = 5{,}000 \Rightarrow 1.96 \frac{S}{\sqrt{64}} = 5{,}000 \Rightarrow S = \frac{40{,}000}{1.96} = 20408$$

**فرمول محاسبه حجم نمونه:** برای محاسبه حجم نمونه، جداولی از طرف آقای مورگان ارائه شده که با مشخص بودن n جامعه می‌توانیم حجم نمونه مناسب را تعیین کنیم. اما از طرف دیگر آقای کوکران برای مواقعی که n جامعه مشخص باشد یا نباشد، رابطه‌های زیر را ارائه داد.

N: جامعه مشخص نباشد
$$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = e \Rightarrow \sqrt{n}\, e = Z_{\frac{\alpha}{2}} \sigma \Rightarrow \sqrt{n} = \frac{Z_{\frac{\alpha}{2}} \sigma}{e} \Rightarrow n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{e}\right)^2$$

N: جامعه مشخص باشد
$$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = e \Rightarrow Z_{\frac{\alpha}{2}} \sigma \sqrt{\frac{N-n}{N-1}} = \sqrt{n}\, e \Rightarrow Z^2_{\frac{\alpha}{2}} \sigma^2 \frac{N-n}{N-1} = n e^2$$
$$\Rightarrow n e^2 (N-1) = Z^2_{\frac{\alpha}{2}} \sigma^2 (N-n) \Rightarrow n\left[e^2(N-1) + Z^2_{\frac{\alpha}{2}} \sigma^2\right] = Z^2_{\frac{\alpha}{2}} \sigma^2 N$$

* اگر e را ندهند، $e = \alpha \bar{x}$ ...

**مثال:** برای اینکه حجم نمونه مناسب جهت بررسی میزان جرایم خودروها را تعیین نمائیم، یک نمونه مقدماتی به حجم 25 خودرو انتخاب می‌کنیم. میانگین جرایم آنها 35,000 تومان با پراکندگی 15,000 تومان بوده است. در سطح $\alpha = \%5$، حجم نمونه مناسب را تعیین کنید.

$$e = \alpha \bar{x} = 0.05 \times 35{,}000 = 1750$$

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{e}\right)^2 = \left(\frac{1.96 \times 15{,}000}{1750}\right)^2 = 283$$

ارزیابی کردیم

$283 - 25 = 258$   257 نمونه دیگر باید انتخاب کنیم

**مثال:** در دانشگاه شهید بهشتی، 900 هیئت علمی وجود دارد که می‌خواهیم میزان درآمدهای آنها را احساب کنیم. یک نمونه به حجم 50 به تصادف انتخاب می‌کنیم. پراکندگی درآمد آنها 2,500,000 تومان با دقت $e = \pm 500{,}000$ می‌باشد. حجم نمونه مناسب را انتخاب کنید.

$N = 900 \qquad n = 50 \qquad S = 2{,}500{,}000 \qquad e = \pm 500{,}000$

$$n = \left(\frac{1.96 \times 2{,}500{,}000}{500{,}000}\right)^2 \simeq 97$$

$$n = \frac{1.96^2 \times 2{,}500{,}000^2 \times 900}{500{,}000^2 (900-1) + 1.96^2 \times 2{,}500{,}000^2} \simeq 87$$

n در این روش معمولاً کمتر می‌شود

---

92/2/10    جلسه پنجم

**برآورد نسبت یک جامعه:** به همین منظور یک نمونه تصادفی به حجم n انتخاب و پس از بدست آوردن نسبت نمونه، از رابطه زیر نسبت جامعه تعیین می‌گردد...

$$p - Z_{\alpha/2} \sqrt{\frac{pq}{n}} < P < p + Z_{\alpha/2} \sqrt{\frac{pq}{n}}$$

$$p = \frac{x}{n}$$
$$q = 1 - p$$

**مثال:** می‌خواهیم نسبت طرفداران رئیس جمهور را برآورد نمائیم. یک نمونه به حجم 1000 نفر انتخاب می‌کنیم. 200 نفر از طرفداران ایشان بودند. در سطح $\alpha = \%5$، این نسبت را در جامعه برآورد کنید.

$p = \frac{200}{1000} = 0.2$ $\qquad q = 1-p = 1-0.2 = 0.8$

$$0.2 - 1.96\sqrt{\frac{0.2\times0.8}{1000}} < P < 0.2 + 1.96\sqrt{\frac{0.2\times0.8}{1000}} \Rightarrow 0.175 < P < 0.225$$

اگر ~~۱۰۰~~ بار نمونه‌گیری کنیم، در سطح $\alpha = 5\%$، ۹۵ بار در این سطح قرار می‌گیرد.

مثال: از نمونه‌ای به حجم n از افرادی که طی 3 سال اخیر ازدواج نموده‌اند و به تصادف انتخاب نمودیم. در سطح اطمینان ۹۵٪، نسبت جامعه بین ۰.۲٪ و ۰.۳٪ تعیین شده است. اولاً) نسبت نمونه، ثانیاً) حجم نمونه را تعیین کنید.

$0.2 < P < 0.3$

اولاً) $p = \frac{0.2+0.3}{2} = 0.25$

ثانیاً) $p \pm e \begin{cases} 0.3 \\ 0.2 \end{cases} \Rightarrow 0.25 + e = 0.3 \Rightarrow e = 0.05$

$$1.96\sqrt{\frac{0.25\times0.75}{n}} = 0.05 \Rightarrow \boxed{n \simeq 289}$$

فرمول محاسبه حجم نمونه (برای جامعه محدود):

$$Z_{\alpha/2}\sqrt{\frac{pq}{n}} = e \Rightarrow Z^2_{\alpha/2}\frac{pq}{n} = e^2 \Rightarrow n = \frac{Z^2_{\alpha/2}\,pq}{e^2} \Rightarrow n = \left(\frac{Z_{\alpha/2}}{e}\right)^2 p.q$$

مثال: برای اینکه نسبت ضایعات کارخانه‌ای را برآورد نمائیم، چه حجم نمونه‌ای را می‌بایستی با دقت ۳٪ در سطح اطمینان ۹۵٪ تعیین نمائیم؟

$e = 0.03 \qquad \alpha = 0.05$

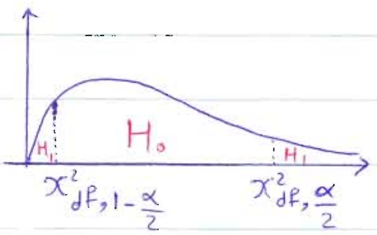
$$n = \left(\frac{1.96}{\pm 0.03}\right)^2 \frac{1}{2} \times \frac{1}{2} = 1068$$

نکته: در برآورد نسبت یک جامعه اگر e مشخص نباشد، $e = \alpha$ می‌شود.

نکته: در برآورد نسبت یک جامعه اگر p مشخص نباشد، $p = q = \frac{1}{2}$ می‌شود.

برآورد واریانس یک جامعه: برای اینکه میزان تغییرپذیری یا ریسک یک پدیده را بررسی نمائیم، از این برآورد استفاده می‌نمائیم. ابتدا یک نمونه تصادفی به حجم n انتخاب می‌نمائیم و پس از بدست آوردن واریانس آن، از رابطه زیر واریانس جامعه را برآورد می‌نمائیم.

نکته قابل ذکر اینکه توزیع میانگین‌ها و نسبت‌های نمونه‌های یک جامعه از توزیع نرمال برخوردار بوده اما توزیع واریانس نمونه‌های یک جامعه از توزیع کای دو یا خی دو (کای مربع یا $\chi^2$) که یک توزیع کشیده به راست می‌باشد، استفاده می‌نماید. به طوری که اگر به جای نمونه n تایی، نمونه ۱ واحدی انتخاب نمائیم، توزیع کای دو با $Z^2$ برابر می‌شود ($\chi^2 = Z^2$).




$$\frac{(n-1)S^2}{\chi^2_{df,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{df,1-\frac{\alpha}{2}}} \qquad df = n-1$$

مثال: می‌خواهیم واریانس هزینه کادوهایی که امشب به خانمها پرداخت می‌شود را برآورد کنیم. یک نمونه به حجم 25 آقا انتخاب نمودیم. میانگین ۲۵,۰۰۰ تومان با پراکندگی ۵,۰۰۰ تومان شده است. در سطح اطمینان ۹۵٪ پراکندگی را برآورد نمائید. (با فرض توزیع نرمال)

$\bar{x} = 25{,}000 \qquad S = 5{,}000 \qquad n = 25$

$$\frac{(25-1)5{,}000^2}{\chi^2_{24,0.025} = 39.3641} < \sigma^2 < \frac{(25-1)5{,}000^2}{\chi^2_{24,0.975} = 12.4011} \Rightarrow 15242314.70 < \sigma^2 < 4838280.475 \Rightarrow 39041.4 < \sigma < 69557.7$$

سمت چپ به ۵,۰۰۰ نزدیکتر است. چون کشیدگی به راست دارد.

نکته: درصورتی که درجه آزادی 1 باشد ($df=1$)، آنگاه $Z^2 = \chi^2$

$df=1$ و $\alpha = 0.05$ $\qquad Z^2 = 1.96^2$ ، $\chi^2 = 3.84$

مثال: می خواهیم ریسک خرید یک نوع سهام را برآورد کنیم. در 30 روزی که سهام معامله شده، میانگین 1500 با پراکندگی 80 تومان بوده است. در سطح $\alpha = 1\%$، ریسک سهام را برآورد کنید.

داده ها حتماً نرمالند. چون $n=30$ مرز نرمال است، در اینجا ذکر شده که داده ها نرمالند.

$n=30 \quad \bar{x}=1500 \quad S=80 \quad \alpha=0.01 \quad \frac{\alpha}{2}=0.005 \quad \chi^2_{29,\,0.005}=52.3356 \quad \chi^2_{29,\,0.995}=13.1211$

$$\frac{(30-1)\,80^2}{\chi^2_{29,\,0.005}=52.3356} < \sigma^2 < \frac{(30-1)\,80^2}{\chi^2_{29,\,0.995}=13.1211} \Rightarrow 3546.32 < \sigma^2 < 14145.15 \Rightarrow 59.55 < \sigma < 118.9$$

ریسک سهام بین این مقادیر است.

آزمون ها: هرگاه در مورد یکی از سه مشخصه فوق یعنی میانگین، نسبت و واریانس ادعایی صورت گیرد، می بایستی آزمون انجام شود. (اگر چه برآورد ابزاری قوی تر از آزمون است، زیرا در فاصله برآورد، هر نوع ادعایی تأیید می گردد. اما در پایان نامه و مراحل دیگر، آزمون ها کاربرد بیشتری دارد).

در هر آزمونی مراحل چندگانه زیر دنبال می شود:

1 یک دامنه یا دو دامنه بودن آزمون: هرگاه ادعا به صورت مساوی یا مخالف ($=$، $\neq$) مطرح شود، آزمون دو دامنه است ولی اگر ادعا به صورت کمتر و بیشتر، حداقل و حداکثر ($<$، $>$، $\geq$، $\leq$) باشد، آزمون یک دامنه می گردد.

2 تعیین فرض $H_0$ و $H_1$: در آزمون های دو دامنه، معمولاً فرض محقق $H_0$ بوده اما در آزمون های یک دامنه، فرض محقق $H_1$ می باشد.

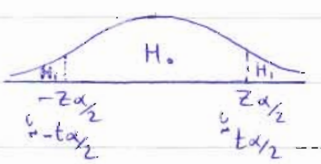
3 تعیین مقدار بحرانی جدول: همانند برآوردها به یکی از 3 جدول $Z$ و $t$ و $\chi^2$ مراجعه می کنیم.

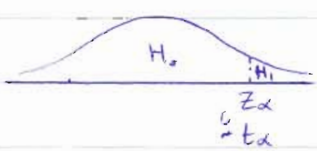
4 تعیین آماره آزمون مناسب که همان $Z$ و $t$ و $\chi^2$ می باشد.

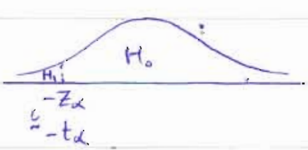
5 محاسبه آماره آزمون و مقایسه نتایج آن با مقدار بحرانی جدول در جهت پذیرش فرض $H_0$ و $H_1$

6 تفسیر نتایج

1 آزمون فرض میانگین یا آزمون تک متغیره (آزمون پارامتریک): هرگاه در مورد میانگین یک جامعه ادعایی صورت گیرد، ضمن رعایت مراحل آزمون ها می بایستی از آماره $Z$ یا $t$ استفاده نمائیم. اگر انحراف معیار جامعه مشخص باشد، بدون توجه به حجم نمونه، از رابطه 1 استفاده می کنیم ولی اگر انحراف معیار جامعه مشخص نباشد و حجم نمونه بیشتر از 30 باشد از رابطه 2 ولی اگر کمتر از 30 باشد، از رابطه 3 استفاده می کنیم.




$H_0: \mu=\mu_0,\ H_1: \mu\neq\mu_0$ یا $\mu-\mu_0=0,\ \mu-\mu_0\neq 0$

$H_0: \mu=\mu_0,\ H_1: \mu>\mu_0$ یا $\begin{cases}\mu\leq\mu_0\\ \mu>\mu_0\end{cases}$ $\begin{cases}\mu-\mu_0\leq 0\\ \mu-\mu_0>0\end{cases}$

$H_0: \mu=\mu_0,\ H_1: \mu<\mu_0$ یا $\begin{cases}\mu\geq\mu_0\\ \mu<\mu_0\end{cases}$ $\begin{cases}\mu-\mu_0\geq 0\\ \mu-\mu_0<0\end{cases}$

(1) $Z=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ $\qquad$ (2) $Z=\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}}$ $\qquad$ (3) $t=\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}}$

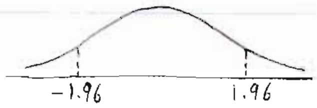
مثال: ادعا شده است که میانگین تماشای تلویزیون در ایران 2.5 ساعت می باشد. 100 نفر نمونه به تصادف انتخاب نمودیم. میانگین تماشای تلویزیون 2.2 با پراکندگی 0.3 شده است. در سطح $\alpha = 5\%$، ارزیابی شما از ادعا چیست؟

1 چون ادعای کمتر و بیشتر و حداقل و حداکثر نداریم، 2 دامنه است.

2
$H_0: \mu = 2.5$ فرض محقق

$H_1: \mu \neq 2.5$ فرض مقابل

3 چون حجم نمونه بالای 30 است، پس از Z استفاده می کنیم.




$Z_{0.025} = 1.96$

4 چون انحراف معیار را از نمونه گرفتیم، رابطه (2) است.
$$Z = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

$$Z = \frac{2.2 - 2.5}{\frac{0.3}{\sqrt{100}}} = -10$$

5 چون 10- در منطقه $H_1$ است، پس فرض $H_1$ تأیید می شود.

6 در سطح اطمینان 95%، ادعای مؤسسه یا محقق تأیید می شود.
یا
در سطح اطمینان 95%، دلیلی برای تأیید فرض $H_0$ وجود ندارد.
یا
در سطح اطمینان 95%، دلیلی برای رد فرض $H_1$ وجود ندارد.

⇐ یکی از این 3 تفسیر را می توانیم بنویسیم.

مثال: می خواهیم ادعایی مبنی بر اینکه هر دانشجوی ارشد بیشتر از 40000 تومان در هر فیش پرداخت می نماید را بررسی نمائیم. یک نمونه به حجم 15 دانشجو به تصادف انتخاب (با فرض توزیع نرمال) و داده های جدول حاصل شده است. در سطح $\alpha = 5\%$ ارزیابی شما چیست؟

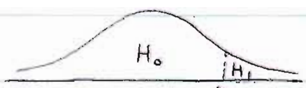
| $x$ ($\times 1000$) | $(x-\bar{x})^2$ |
|---|---|
| 39 | 324 |
| 60 | 9 |
| 75 | 324 |
| 140 | 6889 |
| 20 | 1369 |
| 34 | 529 |
| 82 | 625 |
| 50 | 49 |
| 35 | 484 |
| 42 | 225 |
| 79 | 484 |
| 48 | 81 |
| 42 | 225 |
| 32 | 625 |
| 77 | 400 |
| 855 | 12642 |

1) چون گفته بیشتر از، پس 1 دامنه است.

2) $H_0: \mu = 40$

$H_1: \mu > 40$

3) چون حجم نمونه کمتر از 30 است، پس از t استفاده می کنیم.




$t_{14, 0.05} = 1.761$

4) $t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$

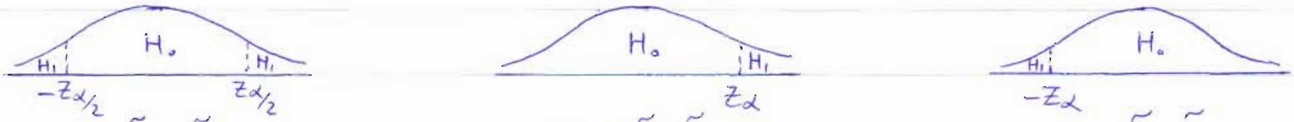
5) $\bar{x} = \frac{855}{15} = 57$

$S = \sqrt{\frac{12642}{15-1}} \cong 30$

$t = \frac{57 - 40}{\frac{30}{\sqrt{15}}} = 2.2$ فرض $H_1$ تأیید می شود.

6) در سطح اطمینان 95%، ادعای مؤسسه یا محقق تأیید می شود.
یا
در سطح اطمینان 95%، دلیلی برای تأیید فرض $H_0$ وجود ندارد.
یا
در سطح اطمینان 95%، دلیلی برای رد فرض $H_1$ وجود ندارد.

**2 آزمون علامت (آزمون ناپارامتریک):** هرگاه داده‌ها نرمال نباشند، از آزمون علامت استفاده می‌کنیم. ابتدا داده‌ها را با مقدار ادعا شده مقایسه می‌کنیم. بیشتر باشد، یک علامت مثبت می‌گذاریم؛ کمتر باشد، یک علامت منفی می‌گذاریم و مساوی باشند حذف می‌کنیم. سپس تعداد علامتهای مثبت را جمع می‌کنیم. اگر n بیشتر از ۱۰ باشد، از تقریب نرمال استفاده نموده، ضمن رعایت مراحل آزمونها، از آماره Z استفاده نموده، ارزیابی صورت می‌گیرد. ولی اگر n کمتر از ۱۰ باشد، به جدول توزیع دوجمله‌ای مراجعه، با n داده شده، $P=\%50$ و x محاسبه شده، مقدار احتمال را تعیین و از 1 کسر می‌نمائیم. چنانچه کمتر از $\alpha$ قرار گیرد، فرض $H_1$ تأیید می‌شود و اگر بیشتر از $\alpha$ شود، فرض $H_0$ تأیید می‌شود.




$H_0: \tilde{M} = \tilde{M}_0$ $\quad$ $H_1: \tilde{M} \neq \tilde{M}_0$

$H_0: \tilde{M} = \tilde{M}_0$ $\quad$ $H_1: \tilde{M} > \tilde{M}_0$

$H_0: \tilde{M} = \tilde{M}_0$ $\quad$ $H_1: \tilde{M} < \tilde{M}_0$

$$Z = \frac{x - np}{\sqrt{npq}} \qquad p = q = \frac{1}{2}$$

**مثال:** مثال قبل را بدون پیش فرض نرمال محاسبه نمائید.

| $x$ ($\times 1000$) | علامت |
|---|---|
| 39 | − |
| 60 | + |
| 75 | + |
| 140 | + |
| 20 | − |
| 34 | − |
| 82 | + |
| 50 | + |
| 35 | − |
| 42 | + |
| 79 | + |
| 48 | + |
| 42 | + |
| 32 | − |
| 77 | + |

$x = 10$ $\qquad$ $Z_{0.05} = 1.645$

$$Z = \frac{10 - 15 \times 1/2}{\sqrt{15 \times 1/2 \times 1/2}} = 1.29$$

فرض $H_0$ تأیید می‌شود. $\quad 1.29 < 1.645$

* صرورتاً آزمون پارامتریک و ناپارامتریک هم سو نیستند.

وقتی $n>10$ باشد از رابطه Z استفاده می‌شود فقط برای نشان دادن روش حل توزیع دوجمله‌ای از این روش استفاده کردیم

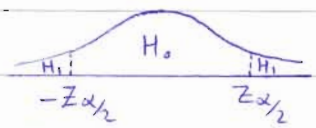
$$P(n=15,\ p=\tfrac{1}{2},\ x>10) = 1 - P(n=15,\ p=\tfrac{1}{2},\ x\leq 10) = 1 - 0.941 = 0.059$$

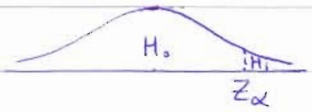
$\alpha = 0.05$ $\qquad$ $0.059 > 0.05$

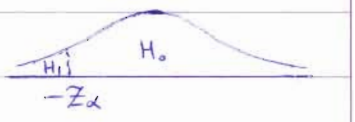
فرض $H_0$ تأیید می‌شود.

* اگر در جدول مقابل یک 40 داشتیم، آن را حذف می‌کردیم، n از 15 یکی کم می‌کردیم و $n=14$ می‌شد.

آزمون نسبت یک جامعه: هرگاه بخواهیم در مورد نسبت یک جامعه ارزیابی انجام دهیم، از آن جامعه نمونه‌های تصادفی انتخاب می‌کنیم و پس از بدست آوردن نسبت نمونه، ضمن رعایت مراحل آزمونها، از آماره $Z$ استفاده کرده، ارزیابی صحیح به عمل می‌آوریم.




$H_0: P = P_0$ $\qquad$ $H_0: P = P_0$ $\qquad$ $H_0: P = P_0$

$H_1: P \neq P_0$ $\qquad$ $H_1: P > P_0$ $\qquad$ $H_1: P < P_0$

$$Z = \frac{P - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$$

مثال: می‌گویند مردم از وضعیت اقتصادی ناراضیند. به همین منظور یک نمونه به حجم 1000 شهروند انتخاب می‌کنیم. 600 نفر ناراضی بودند. در سطح $\alpha = \%5$ ارزیابی شما چیست؟

$H_0: P = \%50$ $\qquad$ وقتی می‌گوییم ناراضیند یعنی باید بالای $\%50$ ناراضی باشند.

$H_1: P > \%50$

$$P = \frac{600}{1000} = 0.6$$

$$Z = \frac{0.6 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 6.32$$ فرض $H_1$ پذیرفته می‌شود

نکته: $\alpha$ با بیشتر از 4 رقم اعشار که 4 رقم اول آنها صفر است را صفر می‌گیریم.

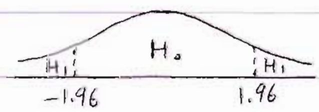
Pvalue: در خروجی نرم افزار، برای اینکه مشخص شود فرض $H_0$ یا $H_1$ کدام یک پذیرفته می‌شود، اگر Pvalue داده شده کمتر از $\alpha$ (یک دامنه) یا $\frac{\alpha}{2}$ (دو دامنه) باشد، فرض $H_1$ پذیرفته می‌شود و اگر بیشتر از $\alpha$ یا $\frac{\alpha}{2}$ باشد، فرض $H_0$ پذیرفته می‌شود. اما برای اینکه Pvalue را بدست بیاوریم، به جدولهای توزیع نرمال مراجعه می‌نماییم.

مثال: یک سکه را 100 بار انداختیم. 54 بار شیر آمده است. آیا سکه سالم است؟

$H_0: P = \%50$

$H_1: P \neq \%50$

$$P = \frac{54}{100} = 0.54$$



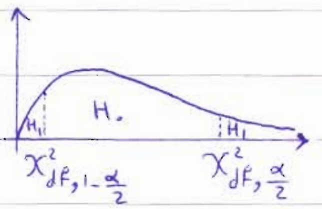

$$Z = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{100}}} = 0.8$$ فرض $H_0$ تأیید می‌شود.

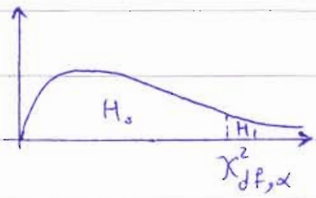
نمی‌توانیم بگوییم که سکه خراب بوده است.

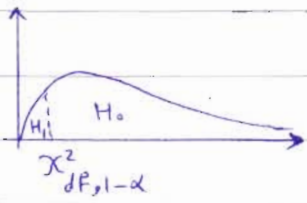
$Z_{0.8} = 0.7881$ $\qquad$ $Pvalue = \frac{1 - 0.7881}{2} = \frac{0.2119}{2} = 0.10595 > 0.025$

چون دو دامنه است بر 2 تقسیم می‌شود. اگر یک دامنه بود، $Pvalue = 0.2119$ می‌شد.

**آزمون واریانس یک جامعه:** هرگاه بخواهیم میزان ریسک، تغییرپذیری و پراکندگی یک جامعه را بررسی نمائیم، از این آزمون استفاده می کنیم. برای استفاده از این آزمون ابتدا نمونه تصادفی از جامعه ای انتخاب می کنیم، واریانس آن را محاسبه می نمائیم و سپس ضمن رعایت مراحل آزمونها، از آماره $\chi^2$ استفاده نموده، ارزیابی صورت می گیرد ضمن اینکه حتماً می بایستی داده ها نرمال باشد.



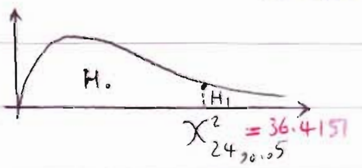

$H_0: \sigma^2 = \sigma_0^2$ $\quad$ $H_0: \sigma^2 = \sigma_0^2$ $\quad$ $H_0: \sigma^2 = \sigma_0^2$

$H_1: \sigma^2 \neq \sigma_0^2$ $\quad$ $H_1: \sigma^2 > \sigma_0^2$ $\quad$ $H_1: \sigma^2 < \sigma_0^2$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

**مثال:** می خواهیم پراکندگی زمان تأخیر کارکنان دستگاهی را برآورد نمائیم. از 25 کارمندی که به تصادف انتخاب نمودیم، میانگین زمان تأخیر 2.5 ساعت با پراکندگی 0.5 ساعت شده است. 1) ادعای دستگاه منوط بر اینکه پراکندگی بیشتر از 0.4 ساعت است را ارزیابی کنید. 2) ادعای دستگاه منوط بر اینکه پراکندگی 0.4 ساعت است را ارزیابی کنید.

(1)

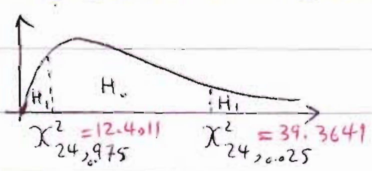



$H_0: \sigma^2 = 0.4^2$

$H_1: \sigma^2 > 0.4^2$ فرض $H_1$ تأیید می شود

$$\chi^2 = \frac{(25-1)\,0.5^2}{0.4^2} = 37.5$$

(2)

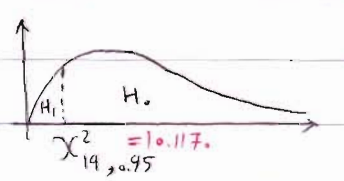



$H_0: \sigma^2 = 0.4^2$

$H_1: \sigma^2 \neq 0.4^2$ فرض $H_0$ تأیید می شود.

**مثال:** مدیرعامل شرکت آلفا ادعا نموده که ریسک خرید سهام آلفا کمتر از 40 تومان می باشد. برای ارزیابی طی 20 روز کاری که معامله صورت گرفته است، میانگین قیمت سهام 800 تومان با پراکندگی 50 تومان شده است. ارزیابی شما چیست؟ (با فرض توزیع نرمال)

$n = 20 \qquad \bar{x} = 800 \qquad S = 50$




$$\chi^2 = \frac{(20-1)\,50^2}{40^2} = 29.68$$ فرض $H_0$ تأیید می شود

$H_0: \sigma^2 = 40^2$

$H_1: \sigma^2 < 40^2$

* چون گفته کمتر از 40 و پراکندگی ما 50 است، می دانیم که تأیید می شود.
اگر گفته بود کمتر از 40 و پراکندگی ما 35 بود، شک می کردیم.

(پارامتریک)

برآورد اختلاف میانگین دو جامعه (دو گروه مستقل): برای اینکه اختلاف میانگین دو جامعه را برآورد کنیم، از هر کدام نمونه‌های تصادفی انتخاب می‌کنیم و میانگین نمونه‌ها را بدست می‌آوریم. اگر انحراف معیار دو جامعه مشخص باشد، با شرط توزیع نرمال بدون توجه به حجم نمونه از رابطه (1) بدست می‌آید. اگر انحراف معیار دو جامعه مشخص نباشد ولی حجم نمونه‌ها بیشتر از 30 باشد که طبیعتاً نرمال خواهند گردید از رابطه (2) استفاده می‌کنیم ولی چنانچه حجم نمونه‌ها کمتر از 30 باشد ولی نرمال باشد، از رابطه (3) استفاده می‌شود ولی چنانچه حجم نمونه‌ها کمتر از 30 باشد و لی نرمال نباشد، از رابطه (4) یعنی چبی شف استفاده می‌شود.

نرمال:

$$(1)\quad (\bar{x}_1-\bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

بدون توجه به حجم نمونه‌ها با شرط اینکه انحراف معیار جامعه مشخص باشد.

$$(2)\quad (\bar{x}_1-\bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

انحراف معیار نمونه مشخص باشد. $n_1, n_2 > 30$

$$(3)\quad (\bar{x}_1-\bar{x}_2) - t_{\alpha/2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + t_{\alpha/2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

انحراف معیار نمونه مشخص باشد. $n_1, n_2 < 30$

غیرنرمال:

$$(4)\quad (\bar{x}_1-\bar{x}_2) - \sqrt{\frac{1}{\alpha}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + \sqrt{\frac{1}{\alpha}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

انحراف معیار نمونه مشخص باشد. $n_1, n_2 < 30$

$$S_p = S_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}} \qquad df = n_1+n_2-2$$

مثال: برای مقایسه میانگین سپرده‌های قرض الحسنه دو بانک تجارت و ملت، 100 دفترچه از بانک تجارت به تصادف انتخاب نمودیم. میانگین موجودی 150,000 با پراکندگی 40,000 می‌باشد. از بانک ملت، 80 دفترچه انتخاب نمودیم. میانگین موجودی 125,000 با پراکندگی 30,000 می‌باشد. در سطح $\alpha = 1\%$ ارزیابی شما چیست؟

$n_1 = 100 \qquad n_2 = 80$

$\bar{x}_1 = 150,000 \qquad \bar{x}_2 = 125,000$ رابطه (2) انتخاب می‌شود.

$S_1 = 40,000 \qquad S_2 = 30,000$

* هر کدام از نمونه‌ها را که اول می‌آوریم، ادعای ما روی آن است.

$$(150,000-125,000) - 2.576\sqrt{\frac{40,000^2}{100}+\frac{30,000^2}{80}} < \mu_1-\mu_2 < (150,000-125,000) + 2.576\sqrt{\frac{40,000^2}{100}+\frac{30,000^2}{80}}$$

$$11,553 < \mu_1-\mu_2 < 38,447$$

چون دو سمت هم علامت هستند، پس اختلاف داریم.

ادعا: میانگین سپرده‌های قرض الحسنه بانک تجارت از 11,553 الی 38,447، از بانک ملت بیشتر است.

نکته: اگر در هر دو سمت برآورد، علامتها یکسان باشد، نشان دهنده وجود اختلاف است. ولی اگر علامتها یکسان نباشد، عدم اختلاف را نشان می‌دهد.

مثال: برای اینکه میانگین طول عمر دو نوع لامپ را مقایسه کنیم، از شرکت اول 25 نمونه انتخاب می‌کنیم. میانگین بر حسب ماه، 12 ماه با پراکندگی 3 ماه می‌باشد. از دومی 30 لامپ انتخاب می‌کنیم. میانگین 14 ماه با پراکندگی 4 ماه شده است. در سطح $\alpha = 1\%$ ارزیابی شما چیست؟

رابطه (2) انتخاب می‌شود.

$$df = 25 + 30 - 2 > 30$$

$$t_{53,\,0.005} \simeq Z_{0.005} = 2.576$$

$$(12-14) - 2.576\sqrt{\frac{3^2}{25} + \frac{4^2}{30}} < \mu_1 - \mu_2 < (12-14) + 2.576\sqrt{\frac{3^2}{25} + \frac{4^2}{30}}$$

$$-4.44 < \mu_1 - \mu_2 < 0.44$$

چون علامتها یکسان نیستند، پس اختلاف نداریم.

ادعا: میانگین عمر هر دو برند یکی است.

مثال: می‌خواهیم مقایسه‌ای بین میانگین سن دانشجویان دختر و پسر بدست آوریم. از هر گروه به شرح جدول زیر نمونه‌های تصادفی انتخاب نمودیم. اختلاف را برآورد و ارزیابی نمائید. الف) با فرض توزیع نرمال. ب) بدون فرض توزیع نرمال.

| x خانم | $(x-\bar{x})^2$ |
|---|---|
| 30 | 4 |
| 37 | 81 |
| 28 | 0 |
| 24 | 16 |
| 34 | 36 |
| 25 | 9 |
| 25 | 9 |
| 24 | 16 |
| 24 | 16 |
| 29 | 1 |
| 280 | 188 |

| $(x-\bar{x})^2$ | x آقا |
|---|---|
| 4 | 28 |
| 0 | 30 |
| 1 | 31 |
| 16 | 26 |
| 4 | 28 |
| 64 | 38 |
| 9 | 27 |
| 25 | 35 |
| 49 | 37 |
| 16 | 26 |
| 9 | 27 |
| 9 | 27 |
| 206 | 360 |

$$\bar{x} = \frac{280}{10} = 28 \qquad \bar{x} = \frac{360}{12} = 30$$

$$S = \sqrt{\frac{188}{10-1}} = 4.57 \qquad S = \sqrt{\frac{206}{12-1}} = 4.33$$

الف) رابطه (3) انتخاب می‌شود.

$$df = 10 + 12 - 2 = 20 \qquad t_{20,\,0.025} = 2.086$$

$$(28-30) - 2.086 \times 4.44\sqrt{\frac{1}{10} + \frac{1}{12}} < \mu_1 - \mu_2 < (28-30) + 2.086 \times 4.44\sqrt{\frac{1}{10} + \frac{1}{12}}$$

$$-5.97 < \mu_1 - \mu_2 < 1.97$$

اختلاف معناداری بین سن آقایان و خانمها وجود ندارد.

$$S_p = S_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{(10-1)\,4.57^2 + (12-1)\,4.33^2}{10+12-2}} = 4.44$$

ب) رابطه (4) انتخاب می‌شود.

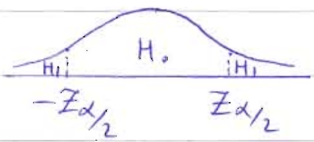
$$(28-30) - \sqrt{\frac{1}{0.05}} \times 4.44\sqrt{\frac{1}{10} + \frac{1}{12}} < \mu_1 - \mu_2 < (28-30) + \sqrt{\frac{1}{0.05}} \times 4.44\sqrt{\frac{1}{10} + \frac{1}{12}}$$

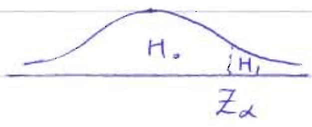
$$-10.5 < \mu_1 - \mu_2 < 6.5$$

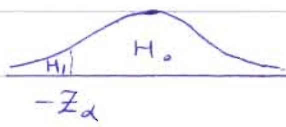
اختلاف معناداری بین سن آقایان و خانمها وجود ندارد.

★ 21 خرداد امتحان میان ترم از فصل 7 و 8

**آزمون میانگین دو گروه مستقل (ناپارامتریک) Mann-Whitney (من-ویتنی):** برای این منظور ابتدا رتبه هر یک از داده‌ها را بدست آورده، سپس آماره $U_1$ یا $U_2$ را محاسبه کرده و در ادامه میانگین و انحراف معیار $U_1$ را بدست آورده ضمن رعایت مراحل آزمون‌ها، از آماره $Z$ استفاده می‌کنیم و ارزیابی صحیح به عمل می‌آوریم. لازم به یادآوری است که اگر آزمون‌ها دو دامنه باشد، بین اینکه $U_1$ یا $U_2$ کدام را انتخاب کنیم، تفاوتی وجود نداشته ولی در آزمون‌های یک دامنه سعی شود تنها $U_1$ یا $U_2$ که ادعا بر روی آن واقع است، آن را ملاک قرار دهیم.




$H_0: \tilde{\mu}_1 = \tilde{\mu}_2$ | $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$ | $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$

$H_1: \tilde{\mu}_1 \neq \tilde{\mu}_2$ | $H_1: \tilde{\mu}_1 > \tilde{\mu}_2$ | $H_1: \tilde{\mu}_1 < \tilde{\mu}_2$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1 \qquad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2$$

$$E(U_1) = E(U_2) = \frac{n_1 n_2}{2} \qquad \sigma(U_1) = \sigma(U_2) = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$$

$$Z_1 = \frac{U_1 - E(U_1)}{\sigma(U_1)} \qquad Z_2 = \frac{U_2 - E(U_2)}{\sigma(U_2)}$$

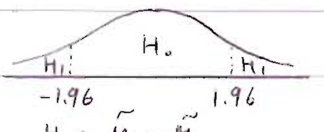
یا از $Z_1$ استفاده می‌شود یا از $Z_2$

**مثال:** در نظر داریم مقایسه‌ای بین عمر دو نوع محصول به عمل بیاوریم. از اولی 10 محصول و از دومی 12 محصول انتخاب کردیم. در شرایط آزمایشگاهی مورد سنجش قرار گرفتند. با توجه به عمر آنها بر حسب ماه، آیا برابری عمر دو محصول را می‌توان ادعا نمود؟

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | جمع |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | 20 | 28 | 32 | 30 | 35 | 25 | 28 | 26 | 29 | 30 | — | — | |
| رتبه | 1 | 8 | 15.5 | 12.5 | 21.5 | 3.5 | 8 | 5 | 10.5 | 12.5 | | | 98 |
| $x_2$ | 24 | 25 | 27 | 32 | 33 | 34 | 28 | 29 | 31 | 35 | 34 | 33 | |
| رتبه | 2 | 3.5 | 6 | 15.5 | 17.5 | 19.5 | 8 | 10.5 | 14 | 21.5 | 19.5 | 17.5 | 155 |

چون آزمون دو دامنه است، فرقی نمی‌کند که از کدام روش استفاده کنیم ولی برای دیدن یکی شدن جواب‌ها از هر دو استفاده می‌کنیم.

برای چک کردن اینکه جواب‌ها درست است یا نه از تصاعد حسابی استفاده می‌کنیم $S_n = \frac{n}{2}(a_1 + a_n) = \frac{22}{2}(1+22) = 253$ تصاعد حسابی




$H_0: \tilde{\mu}_1 = \tilde{\mu}_2$

$H_1: \tilde{\mu}_1 \neq \tilde{\mu}_2$

$$U_1 = 10 \times 12 + \frac{10(10+1)}{2} - 98 = 77$$

$$U_2 = 10 \times 12 + \frac{12(12+1)}{2} - 155 = 43$$

$$E(U_1) = E(U_2) = \frac{10 \times 12}{2} = 60$$

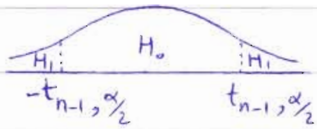
$$\sigma(U_1) = \sigma(U_2) = \sqrt{\frac{10 \times 12(10+12+1)}{12}} = 15.16$$

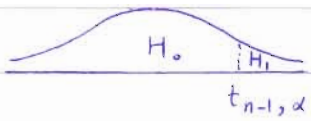
$$Z_1 = \frac{77-60}{15.16} = 1.12 \qquad Z_2 = \frac{43-60}{15.16} = -1.12$$

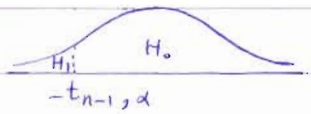
فرض $H_0$ تأیید می‌شود و میانگین عمر آنها یکسان است.

## آزمون میانگین دو گروه وابسته

الف) پارامتریک (آزمون مقایسات زوجی): هرگاه از یک گروه قبل از تغییرات و از همان گروه بعد از تغییرات، ارزیابی صورت گیرد و بخواهیم بررسی نمائیم که آیا تغییرات تأثیری بر روند داشته است یا خیر، دو گروه وابسته نامیده می‌شوند و چنانچه نرمال باشند، از این آزمونها استفاده می‌کنیم. ابتدا داده‌ها را دو به دو مقایسه نموده، اختلاف آنها را بدست می‌آوریم. سپس انحراف معیار آنها را تعیین می‌کنیم و پس از آن ضمن رعایت مراحل آزمونها، از آماره $t$ استفاده نموده، ارزیابی به عمل می‌آوریم. همچنین بحث برآوردها نیز قابل انجام می‌باشد.




$H_0: \mu_d = 0$ $\qquad$ $H_0: \mu_d = 0$ $\qquad$ $H_0: \mu_d = 0$

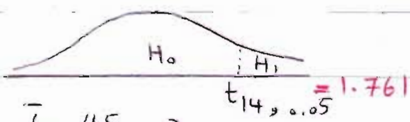
$H_1: \mu_d \neq 0$ $\qquad$ $H_1: \mu_d > 0$ $\qquad$ $H_1: \mu_d < 0$

$$t = \frac{\bar{d}}{S_d/\sqrt{n}} \qquad \bar{d} - t_{n-1,\alpha/2} \cdot \frac{S_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{n-1,\alpha/2} \cdot \frac{S_d}{\sqrt{n}}$$

مثال: می‌خواهیم بررسی نمائیم آیا آمدن آقای هاشمی سبب هیجان بیشتر شده است یا خیر؟ پرسشنامه‌ای تنظیم نمودیم و نمرات آنها را به صورت قبل و بعد در جدول آمده است. در سطح $\alpha = 5\%$، آیا میزان هیجان بعد از آمدن ایشان بیشتر شده است یا خیر؟

| فرد | $x_1$ (قبل) | $x_2$ (بعد) | $d$ | $(d-\bar{d})^2$ |
|---|---|---|---|---|
| 1 | 45 | 48 | 3 | 0 |
| 2 | 50 | 50 | 0 | 9 |
| 3 | 70 | 80 | 10 | 49 |
| 4 | 65 | 60 | -5 | 64 |
| 5 | 60 | 60 | 0 | 9 |
| 6 | 56 | 62 | 6 | 9 |
| 7 | 70 | 75 | 5 | 4 |
| 8 | 61 | 59 | -2 | 25 |
| 9 | 48 | 57 | 9 | 36 |
| 10 | 55 | 58 | 3 | 0 |
| 11 | 64 | 64 | 0 | 9 |
| 12 | 38 | 42 | 4 | 1 |
| 13 | 49 | 53 | 4 | 1 |
| 14 | 50 | 54 | 4 | 1 |
| 15 | 72 | 76 | 4 | 1 |
| | | | 45 | 218 |

آزمون راستی است. درجه آزادی 14، $\alpha = 0.05$




$$\bar{d} = \frac{45}{15} = 3$$

$$S_d = \sqrt{\frac{218}{15-1}} = 3.95$$

$$t = \frac{3}{\frac{3.95}{\sqrt{15}}} = 2.94$$

فرض $H_1$ تأیید می‌شود

اگر خواستند برآورد کنیم:

$$t_{14,0.025} = 2.145$$

$$3 - 2.145 \times \frac{3.95}{\sqrt{15}} < \mu_d < 3 + 2.145 \times \frac{3.95}{\sqrt{15}}$$

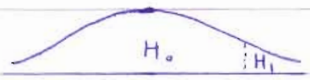
$$0.81 < \mu_d < 5.19$$

هیجان از 0.81 تا 5.19 بیشتر شده است

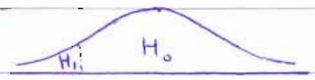
## ب) ناپارامتریک

**ب-1) آزمون علامت زوج نمونه ای :** در این گونه مواقع هنگامی که هر دو داده نظیر را با هم مقایسه می نمائیم ، اگر دومی از اولی بیشتر باشد ، یک مثبت ؛ کمتر باشد ، یک منفی ؛ مساوی باشد ، حذف می نمائیم. پس از آن تعداد مثبت ها را محاسبه و از آماره $Z$ استفاده می نمائیم که در این رابطه $P$ همیشه $\frac{1}{2}$ و $q$ نیز $\frac{1}{2}$ می باشد.


$-Z_{\alpha/2}$ $\quad Z_{\alpha/2}$




$Z_\alpha$




$-Z_\alpha$

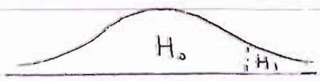
$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$ $\qquad$ $H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$ $\qquad$ $H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$

$H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2$ $\qquad$ $H_1 : \tilde{\mu}_1 > \tilde{\mu}_2$ $\qquad$ $H_1 : \tilde{\mu}_1 < \tilde{\mu}_2$

$$Z = \frac{x - np}{\sqrt{npq}}$$

مثال: ادامه مثال ص 30

| فرد | $x_1$ (قبل) | $x_2$ (بعد) | d | علامت |
|---|---|---|---|---|
| 1 | 45 | 48 | 3 | + |
| 2 | 50 | 50 | 0 | x |
| 3 | 70 | 80 | 10 | + |
| 4 | 65 | 60 | -5 | − |
| 5 | 60 | 60 | 0 | x |
| 6 | 56 | 62 | 6 | + |
| 7 | 70 | 75 | 5 | + |
| 8 | 61 | 59 | -2 | − |
| 9 | 48 | 57 | 9 | + |
| 10 | 55 | 58 | 3 | + |
| 11 | 64 | 64 | 0 | x |
| 12 | 38 | 42 | 4 | + |
| 13 | 49 | 53 | 4 | + |
| 14 | 50 | 54 | 4 | + |
| 15 | 72 | 76 | 4 | + |
| | | | 45 | |



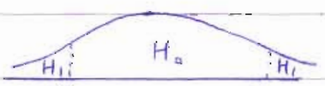

$Z_{0.05} = 1.645$

$$Z = \frac{10 - 12 \times \frac{1}{2}}{\sqrt{\underbrace{12}_{\text{x ها حذف می شوند}} \times \frac{1}{2} \times \frac{1}{2}}} = 2.31$$

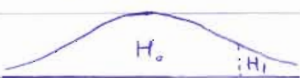
ادعا تایید می شود

یعنی فرض $H_1$ تایید می شود

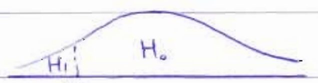
**ب-2) آزمون Wilcoxon :** در این آزمون هنگامیکه داده ها را با هم مقایسه می نمائیم ، رتبه آنها را بدون توجه به علامت مشخص می نمائیم. سپس جمع رتبه های مثبت را در نظر میگیریم و آن $T^+$ می نامیم. پس از آن میانگین و انحراف معیار رتبه ها را از رابطه های زیر بدست آورده ، از آماره $Z$ استفاده نموده ، ارزیابی صحیح به عمل می آوریم.




$-Z_{\alpha/2}$ $\quad Z_{\alpha/2}$




$Z_\alpha$




$-Z_\alpha$

$H_0 : \tilde{\mu} = \tilde{\mu}_0$ $\qquad$ $H_0 : \tilde{\mu} = \tilde{\mu}_0$ $\qquad$ $H_0 : \tilde{\mu} = \tilde{\mu}_0$

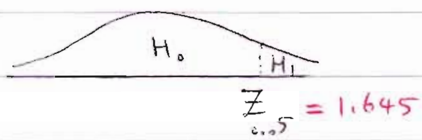
$H_1 : \tilde{\mu} \neq \tilde{\mu}_0$ $\qquad$ $H_1 : \tilde{\mu} > \tilde{\mu}_0$ $\qquad$ $H_1 : \tilde{\mu} < \tilde{\mu}_0$

$$E(T) = \frac{n(n+1)}{4} \qquad \sigma(T) = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$Z = \frac{T^{+} - E(T)}{\sigma(T)}$$

مثال: ادامه مثال ص 30

| فرد | قبل $x_1$ | بعد $x_2$ | $d$ | علامت | اولویت داده‌ها بدون علامت | رتبه |
|---|---|---|---|---|---|---|
| 1 | 45 | 48 | 3 | + | 2 | 1 |
| 2 | 50 | 50 | 0 | X | 3 | 2.5 |
| 3 | 70 | 80 | 10 | + | 3 | 2.5 |
| 4 | 65 | 60 | -5 | − | 4 | 5.5 |
| 5 | 60 | 60 | 0 | X | 4 | 5.5 |
| 6 | 56 | 62 | 6 | + | 4 | 5.5 |
| 7 | 70 | 75 | 5 | + | 4 | 5.5 |
| 8 | 61 | 59 | -2 | − | 5 | 8.5 |
| 9 | 48 | 57 | 9 | + | 5 | 8.5 |
| 10 | 55 | 58 | 3 | + | 6 | 10 |
| 11 | 64 | 64 | 0 | X | 9 | 11 |
| 12 | 38 | 42 | 4 | + | 10 | 12 |
| 13 | 49 | 53 | 4 | + | — | |
| 14 | 50 | 54 | 4 | + | — | |
| 15 | 72 | 76 | 4 | + | — | |
| | | | 45 | | | |




$$T^{+} = 2.5 + 2.5 + 5.5 + 5.5 + 5.5 + 5.5 + 8.5 + 10 + 11 + 12 = 68.5$$

با صفرها کاری نداریم

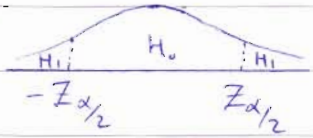
$$E(T) = \frac{12(12+1)}{4} = 39$$

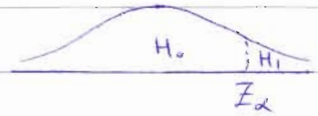
$$\sigma(T) = \sqrt{\frac{12(12+1)(2\times12+1)}{24}} = 12.75$$

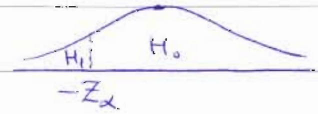
$$Z = \frac{68.5 - 39}{12.75} = 2.31$$

ادعا تأیید می‌شود یعنی فرض $H_1$ تأیید می‌شود

**آزمون نسبت دو جامعه:** هرگاه بخواهیم درمورد برابری یا عدم برابری نسبت دو جامعه ارزیابی صورت دهیم، نمونه‌های تصادفی از دو جامعه انتخاب و پس از بدست آوردن نسبت نمونه‌ها، ضمن رعایت مراحل آزمونها از آماره $Z$ استفاده نموده، ارزیابی به عمل می‌آوریم.




$H_0: P_1 = P_2$ 
$H_1: P_1 \neq P_2$

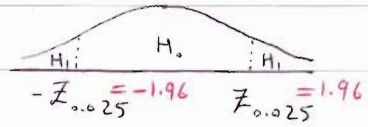
$H_0: P_1 = P_2$ 
$H_1: P_1 > P_2$

$H_0: P_1 = P_2$ 
$H_1: P_1 < P_2$

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

مثال: مدیر استاندارد ایران اعلام نموده، نسبت خودروهای پراید که عدم کیفیت دارند، 5% بیشتر از پژو 405 می‌باشد. برای این منظور از 150 پراید 75 مورد عدم کیفیت خورده و از 100 پژو 40 مورد عدم کیفیت خورده است. در سطح 5% = $\alpha$، ارزیابی شما از ادعای استاندارد چیست؟

آزمون دوطرفه است.




$$H_0: P_1 = P_2 + 0.05$$

$$H_1: P_1 \neq P_2 + 0.05$$

$$P_1 = \frac{75}{150} = 0.5 = 50\%$$

$$P_2 = \frac{40}{100} = 0.4 = 40\%$$

$$Z = \frac{0.5-(0.4+0.05)}{\sqrt{\frac{0.5\times0.5}{150}+\frac{0.4\times0.6}{100}}} = 0.78$$

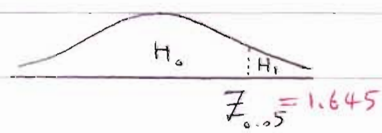
ادعا تأیید شد.

یعنی فرض $H_0$ تأیید می‌شود.

یعنی 5% تأیید می‌شود.

مثال: برای اینکه نسبت خانوارهای زیر خط فقر را در دو شهر تبریز و تهران مقایسه کنیم، در تهران از 1000 خانوار، 400 خانوار و در تبریز از 500 خانوار، 150 خانوار زیر خط فقر بوده‌اند. آیا می‌توان ادعا نمود این نسبت در تهران بیشتر از تبریز است؟

آزمون یک دامنه - وسطی




$$H_0: P_1 = P_2$$

$$H_1: P_1 > P_2$$

$$P_1 = \frac{400}{1000} = 0.4$$

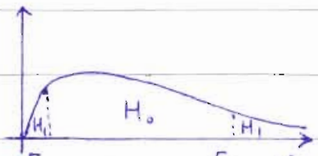
$$P_2 = \frac{150}{500} = 0.3$$

$$Z = \frac{0.4-0.3}{\sqrt{\frac{0.4\times0.6}{1000}+\frac{0.3\times0.7}{500}}} = 3.89$$

ادعا تأیید می‌شود.

فرض $H_1$ تأیید می‌شود.

آزمون واریانس دو جامعه: هرگاه بخواهیم مقایسه‌ای بین ریسک، تغییرپذیری دو جامعه به عمل بیاوریم، از این آزمون استفاده می‌کنیم. ابتدا نمونه‌های تصادفی انتخاب کرده و پس از بدست آوردن واریانس نمونه‌ها، ضمن رعایت مراحل آزمون‌ها، از آماره F استفاده نموده، ارزیابی صحیح به عمل می‌آوریم. نکته قابل ذکر این است که جدول توزیع فیشر یا F، تنها براساس $\frac{\alpha}{2}$ یا $\alpha$ تنظیم شده و اگر $1-\frac{\alpha}{2}$ یا $1-\alpha$ را بخواهیم، می‌بایستی به جای آنها از $\frac{\alpha}{2}$ یا $\alpha$ استفاده کرده و با جابجایی درجات آزادی، به جدول مراجعه کنیم و هرچه که بدست آید، معکوس نمائیم.

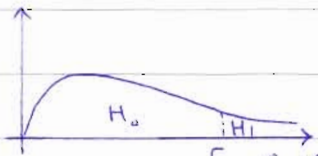



$F_{1-\frac{\alpha}{2}, df_1, df_2}$ $\quad F_{\frac{\alpha}{2}, df_1, df_2}$

$\dfrac{1}{F_{\frac{\alpha}{2}, df_2, df_1}}$

$H_0: \sigma_1^2 = \sigma_2^2$

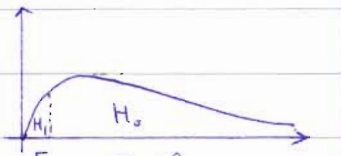
$H_1: \sigma_1^2 \neq \sigma_2^2$




$F_{\alpha, df_1, df_2}$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 > \sigma_2^2$




$F_{1-\alpha, df_1, df_2}$

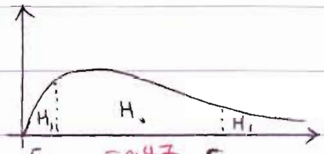
$\dfrac{1}{F_{\alpha, df_2, df_1}}$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 < \sigma_2^2$

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

مثال: در نظر داریم میزان تغییرپذیری عملکرد دو نوع دستگاه را مورد مقایسه قرار دهیم. در 30 روز کاری، از دستگاه اول میانگین تولید روزانه 120 با پراکندگی 8 و در 25 روز کاری از دستگاه دوم انتخاب می‌کنیم. میانگین 110 با پراکندگی 5 شده است. آیا می‌توان ادعای برابری تغییرپذیری عملکرد این دو دستگاه را پذیرفت؟




$F_{0.975, 29, 24} = 0.47 \quad F_{0.025, 29, 24} = 2.27$

$\dfrac{1}{F_{0.025, 24, 29} = 2.15}$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

$F = \dfrac{8^2}{5^2} = 2.56$

ادعای برابری تأیید نمی‌شود.

فرض $H_1$ تأیید می‌شود (چون پراکندگی دستگاه اول بیشتر است)

---

جلسه هشتم $\quad$ 92,2,31

تحلیل واریانس ANOVA (آزمون میانگین چند جامعه): هرگاه بخواهیم میانگین چند جامعه را مقایسه نماییم، از این آزمون استفاده می‌کنیم که هم به صورت پارامتریک و هم ناپارامتریک، هم به صورت یک عامله و هم به صورت دو عامله، مطرح می‌باشد.

تحلیل واریانس یک عامله - پارامتریک:

الف) با حجم نمونه‌های مساوی: پس از اخذ نمونه‌های تصادفی، ابتدا انحرافات کل داده‌ها (مجذور اختلاف هر داده از میانگین) را بدست می‌آوریم $(SST = \sum (x_{ij} - \bar{x})^2)$. پس از آن انحرافات گروه‌ها یا تیمارها یا سطری را از رابطه زیر بدست آورده، پس از آن انحرافات خطا یا پسماند انحرافات (SSE) را بدست می‌آوریم (از کل انحرافات مقداری توسط انحرافات تیمارها یا گروه‌ها بیان شده است. آن مقداری که توسط انحرافات تیمارها پوشش داده نمی‌شود را انحرافات خطا می‌نامیم). پس از این مرحله میانگین انحرافات تیمارها و خطاها محاسبه کرده، آماره فیشر یا F را بدست می‌آوریم.

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{k.n} T^2$$

$$SSr = \frac{1}{n}\sum_{i=1}^{k} T_i^2 - \frac{1}{k.n} T^2$$

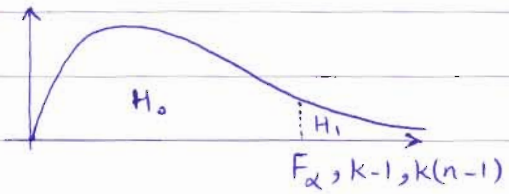
$$SSE = SST - SSr$$

$$MSr = \frac{SSr}{k-1}$$

$$MSE = \frac{SSE}{k(n-1)} \qquad F = \frac{MSr}{MSE}$$

$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots$

$H_1$: حداقل میانگین دوتا از تیمارها با همدیگر برابر نیستند




مثال: می خواهیم بررسی نمائیم که آیا میزان فروش 4 شعبه شهروند یکسان می باشد یا خیر. 6 روز کاری به تصادف انتخاب نمودیم و فروش آنها ثبت شده است. با توجه به داده های جدول، آیا برابری فروش را می توان ادعا نمود؟

| شهروند / نمونه | 1 | 2 | 3 | 4 | 5 | 6 | $T_i$ | $\bar{x}$ |
|---|---|---|---|---|---|---|---|---|
| A | 4.2 | 3.8 | 6 | 5.3 | 5.7 | 6 | 31 | 5.17 |
| B | 5.5 | 6 | 6.4 | 4.9 | 7 | 6.5 | 36.3 | 6.05 |
| C | 3.8 | 4.6 | 5.3 | 7 | 6.1 | 5.1 | 31.9 | 5.3 |
| D | 4 | 4.2 | 4 | 4.3 | 4.8 | 5 | 26.3 | 4.38 |
| | | | | | | | 125.5 | |

($x_1, \dots, x_6$)

$$SST = (4.2^2 + 3.8^2 + 6^2 + \cdots + 4.8^2 + 5^2) - \frac{1}{4 \times 6} \times 125.5^2 = 679.21 - 656.26 = 22.95$$

$$SSr = \frac{1}{6}(31^2 + 36.3^2 + 31.9^2 + 26.3^2) - \frac{1}{4 \times 6} \times 125.5^2 = 664.67 - 656.26 = 8.41$$

$$SSE = 22.95 - 8.41 = 14.54$$

$$MSr = \frac{8.41}{4-1} = 2.8$$

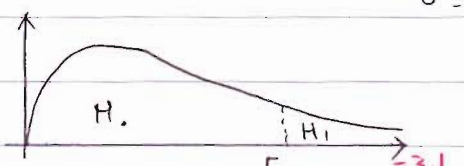
$$MSE = \frac{14.54}{4(6-1)} = 0.727$$

$$F = \frac{2.8}{0.727} = 3.85$$

برابری میانگین فروش این 4 شعبه را نمی توان ادعا کرد.
فرض $H_1$ تأیید می شود.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1$: حداقل میانگین فروش دو شعبه شهروند با همدیگر برابر نیستند




ب) با حجم نمونه های متفاوت: اگر تعداد نمونه ها از هر گروه یکسان نباشد، می بایستی تغییرات کوچکی در رابطه های فوق قائل شویم.

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{N} T^2$$

$$SSr = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{1}{N} T^2$$

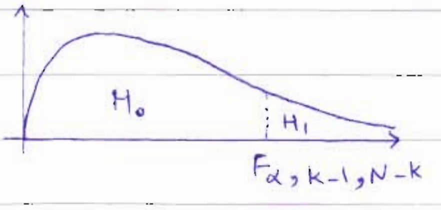
$$SSE = SST - SSr$$

$$MSr = \frac{SST}{k-1}$$

$$MSE = \frac{SSE}{N-k} \qquad F = \frac{MSr}{MSE}$$

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots$

$H_1$: حداقل میانگین دوتا از تیمارها با همدیگر برابر نیستند.




مثال: در نظر داریم میانگین هزینه تفریحات چهار گروه مرفه در یزد، کاشان، اصفهان و تبریز را مقایسه کنیم. نمونه‌های تصادفی از هر یک از این شهرها انتخاب نمودیم. هزینه تفریحات هفتگی آنها در جدول آمده است. آیا بین میانگین هزینه تفریح این چهار شهر اختلاف وجود دارد؟

×1000

| $\bar{x}$ | $T_i$ | 6 | 5 | 4 | 3 | 2 | 1 | نمونه / شهر |
|---|---|---|---|---|---|---|---|---|
| 11.25 | 45 | — | — | 8 | 10 | 15 | 12 | یزد |
| 4.72 | 23.6 | — | 5 | 3.5 | 0 | 8 | 7.1 | کاشان |
| 13 | 78 | 15 | 16 | 14 | 10 | 10 | 13 | اصفهان |
| 17.67 | 106 | 16 | 10 | 25 | 17 | 18 | 20 | تبریز |
| | 252.6 | | | | | | | |

$$SST = (12^2 + 15^2 + 10^2 + \dots + 10^2 + 16^2) - \frac{1}{21} \times (252.6)^2 = 3724.66 - 3038.42 = 686.24$$

$$SSr = \left(\frac{45^2}{4} + \frac{23.6^2}{5} + \frac{78^2}{6} + \frac{106^2}{6}\right) - \frac{1}{21} \times (252.6)^2 = 3504.31 - 3038.42 = 465.89$$

$$SSE = 686.24 - 465.89 = 220.35$$

$$MSr = \frac{465.89}{4-1} = 155.29$$

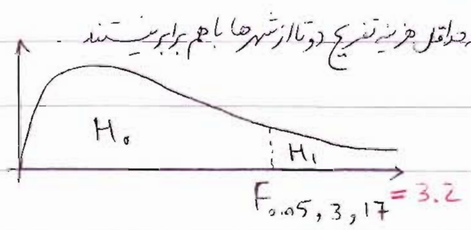
$$MSE = \frac{220.35}{21-4} = 12.96 \qquad F = \frac{155.29}{12.96} = 11.98$$

فرض $H_1$ تأیید می‌شود

می‌توانیم بگوئیم که حداقل هزینه تفریح دوتا از شهرها با هم برابر نیستند

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1$: حداقل هزینه تفریح دوتا از شهرها با هم برابر نیستند.




آزمون اولویت بندی گروه‌ها (جوامع): پس از اینکه مشخص گردید فرض $H_1$ تأیید می‌گردد، می‌بایستی بررسی نمائیم کدامین دو میانگین با همدیگر یکسان یا متفاوت می‌باشند. برای این منظور چند روش وجود دارد.

۱) روش توکی Tukeys: برای این منظور مقدار HSD را بدست آورده، سپس اختلاف هر دو میانگین را تعیین می‌کنیم. اگر اختلافها از HSD کمتر باشد، اختلاف معنی‌دار نبوده اما اگر بیشتر باشد، معنی‌دار خواهد بود. اگر تعداد نمونه‌های مساوی باشد، در مخرج کسر n و اگر مساوی نباشد، میانگین تعداد نمونه‌ها را قرار می‌دهیم.

$$HSD = q_{\alpha, k, N-k} \sqrt{\frac{MSE}{n_{\text{یا}} \bar{n}}}$$

مثال: HSD مثال ص 36 را محاسبه کنید و اولویت بندی نمائید.

$$HSD = q_{0.05, 4, 17} \sqrt{\frac{12.96}{21/4}} \Rightarrow HSD = 4.02 \times 1.57 = 6.3$$

$|\bar{x}_1 - \bar{x}_2| = |11.25 - 4.72| = 6.53$ ✓ — اولویت اول: 3 و 4 (اصفهان و تبریز) ⇐ چون 3 و 4 رابطه معنی داری ندارند، پس در یک اولویت قرار می گیرند

$|\bar{x}_1 - \bar{x}_3| = |11.25 - 13| = 1.75$ — اولویت دوم: 1 (یزد)

$|\bar{x}_1 - \bar{x}_4| = |11.25 - 17.67| = 6.42$ ✓ — اولویت سوم: 2 (کاشان)

$|\bar{x}_2 - \bar{x}_3| = |4.72 - 13| = 8.28$ ✓

$|\bar{x}_2 - \bar{x}_4| = |4.72 - 17.67| = 12.95$ ✓

$|\bar{x}_3 - \bar{x}_4| = |13 - 17.67| = 4.67$

2) آزمون کمترین اختلاف معنی داری یا فیشر یا LSD: برای استفاده از این آزمون ابتدا مقدار LSD را بدست آورده، بیشترین میانگین را منهای این عدد نموده، داده هایی که در این فاصله قرار خواهند گرفت، در اولویت اول قرار می گیرند. سپس نتیجه بالا را منهای LSD نموده، میانگین هایی که در این فاصله باشند، اولویت دوم می باشند و الی آخر.

$$LSD = \sqrt{\frac{2MSE}{n_{\text{یا}} \bar{n}}} \times t_{N-k, \alpha}$$

مثال: مثال ص 36 را از روش LSD اولویت بندی نمائید.

$$LSD = \sqrt{\frac{2 \times 12.96}{21/4}} \times t_{17, 0.05} = 2.22 \times 1.74 = 3.86$$

$\bar{x}_{max} - LSD = 17.67 - 3.86 = 13.81$ — تبریز — $13.81 < 17.67 \leq 17.67$

$13.81 - 3.86 = 9.95$ — یزد و اصفهان — $9.95 < 13 < 13.81$, $9.95 < 11.25 < 13.81$

$9.95 - 3.86 = 6.09$ — ―

$6.09 - 3.86 = 2.23$ — کاشان — $2.23 < 4.72 < 6.09$

تحلیل واریانس یک عامله ناپارامتریک: کروسکال - والیس Kruskal-Wallis

معمولاً هنگامیکه داده ها نرمال نباشند، از این روش استفاده می گردد. ابتدا رتبه هر یک از داده ها را در مجموع بدست آورده، مجموع رتبه های هرکدام را تعیین، پس از آن آماره H را محاسبه کرده و مقدار آن را با مقدار بحرانی که از جدول $\chi^2$ بدست می آید $\chi^2_{k-1,\alpha}$ مقایسه می کنیم. نحوه نوشتن $H_0$ و $H_1$ عیناً مانند روش پارامتریک می باشد.

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)$$

مثال: در نظر داریم مقایسه ای بین عمر چهار نوع لامپ به عمل آوریم. چهار شرکت انتخاب می کنیم. نمونه های تصادفی از هر کدام در نظر گرفتیم و جدول زیر حاصل شده است. در سطح $\alpha = 5\%$ آیا بین میانگین عمر آنها اختلاف معناداری وجود دارد یا خیر؟ (با پیش فرض نرمال) و (بدون پیش فرض نرمال)

| $R_i$ | $\bar{x}$ | $T_i$ | 5 | 4 | 3 | 2 | 1 | نمونه / شرکت |
|---|---|---|---|---|---|---|---|---|
| 40 | 25.8 | 129 | (3.5) 24 | (14) 29 | (18) 32 | (3.5) 24 | (1) 20 | A |
| 66 | 29.2 | 146 | (9.5) 27 | (19) 34 | (16) 30 | (16) 30 | (5.5) 25 | B |
| 57 | 28.6 | 143 | (7.5) 26 | (20) 36 | (12) 28 | (5.5) 25 | (12) 28 | C |
| 47 | 26.8 | 134 | (12) 28 | (16) 30 | (7.5) 26 | (9.5) 27 | (2) 23 | D |
| | | 552 | | | | | | |

با پیش فرض نرمال :

$$SST=(20^2+24^2+32^2+\dots+30^2+28^2)-\frac{1}{4\times5}\times552^2=15510-15235.2=274.8$$

$$SSr=\frac{1}{5}(129^2+146^2+143^2+134^2)-\frac{1}{4\times5}\times552^2=15272.4-15232.2=40.2$$

$$SSE=274.8-40.2=234.6$$

$$MSr=\frac{40.2}{4-1}=13.4$$

$$MSE=\frac{234.6}{4(5-1)}=14.66 \qquad F=\frac{13.4}{14.66}=0.91$$

$$0.91<3.24$$

فرض $H_0$ تایید می‌شود.

$$H_0: \mu_1=\mu_2=\mu_3=\mu_4$$

$H_1$: حداقل میانگین هر دو شرکت با یکدیگر برابر نیستند.

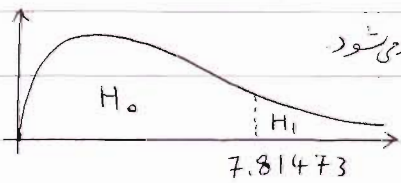
$$F_{0.05,3,16}=3.24$$

بدون پیش فرض نرمال :

$$H=\frac{12}{20(20+1)}\left(\frac{40^2}{5}+\frac{66^2}{5}+\frac{57^2}{5}+\frac{47^2}{5}\right)-3(20+1)=2.22$$

$$\chi^2_{3,0.05}=7.81473$$

فرض $H_0$ تایید می‌شود.




---

92,3,7

جلسه نهم

فصل 7

مثال 1. 30٪ ایرانیانی که دارای سهام می‌باشند، حداقل 1,000,000 تومان ارزش سهام آنهاست. اگر 1000 سهامدار به تصادف انتخاب کنیم، احتمال اینکه بیشتر از 25٪ آنها ارزش سهامشان حداقل 1,000,000 تومان باشد، چقدر است؟

$(Z_{-3.45}=0.0003)$

$P=30\%$ $\quad n=1000$

$$p(\hat{P}>0.25)\simeq p(x>0.25\times1000)=p\left(Z>\frac{0.25-0.3}{\sqrt{\frac{0.3\times0.7}{1000}}}\right)\simeq p\left(Z>\frac{\overbrace{250-1000\times0.3}^{300}}{\underbrace{\sqrt{1000\times0.3\times0.7}}_{210}}\right)$$

$$Z=\frac{\hat{P}-P}{\sqrt{\frac{pq}{n}}} \qquad Z=\frac{x-np}{\sqrt{npq}}$$

$$=p(Z>-3.45)=1-p(Z<-3.45)=1-0.0003=0.9997$$

فصل 7

مثال 2 میانگین وزن هر ترب برابر ۰.5 کیلوگرم با انحراف معیار ۰.1 می باشد. اگر 100 مورد نمونه گیری شود، احتمال اینکه میانگین وزن آنها بیشتر از ۰.7 کیلوگرم باشد، چقدر است؟

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$p(\bar{x} > 0.7) = p\left(Z > \frac{0.7 - 0.5}{\frac{0.1}{\sqrt{100}}}\right) = p(Z > 20) = 1 - p(Z < 20) = 1 - 1 \simeq 0$$

برآوردها

مثال 3 از یک خط تولید، 35 روز کاری به تصادف انتخاب می کنیم. میانگین و انحراف معیار، 150 و 20 گردیده است. اولاً) پراکندگی و میانگین را برآورد نمائید. ثانیاً) با دقت 5 واحد، آیا حجم نمونه کافی بوده است؟

چون حجم نمونه بالای 30 می باشد، از رابطهٔ (2) استفاده می کنیم.

$$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow Z_{0.025} = 1.96$$

$$\bar{x} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\frac{S}{\sqrt{n}} \Rightarrow 150 - 1.96 \times \frac{20}{\sqrt{35}} < \mu < 150 + 1.96 \times \frac{20}{\sqrt{35}} \Rightarrow 143.38 < \mu < 156.62$$

$$df = n - 1 = 35 - 1 = 34$$

$$\frac{(n-1)S^2}{\chi^2_{df,\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{df,(1-\alpha/2)}} \Rightarrow \frac{34 \times 20^2}{\chi^2_{34,0.025} = 46.979} < \sigma^2 < \frac{34 \times 20^2}{\chi^2_{34,0.975} = 16.791} \Rightarrow 289.5 < \sigma^2 < 809.96$$

چون $\chi^2_{30,0.975}$ نزدیکتر است، آن را در نظر می گیریم

$$17 < \sigma < 28.46$$

$$n = \left(\frac{Z_{\alpha/2} \times \sigma}{e}\right)^2 = \left(\frac{1.96 \times 20}{5}\right)^2 = 62$$

$$62 - 35 = 27$$

پس 27 نمونه کمتر نمونه گرفتیم.

برآوردها

مثال 4 برای اینکه نسبت خانواده هایی که حداقل 2 خودرو دارند را برآورد کنیم، 500 مورد انتخاب، 100 مورد حداقل 2 خودرو داشتند. در سطح $\alpha = 1\%$ این نسبت را برآورد کنید؟ آیا حجم نمونه کافی بوده است؟

$$p - Z_{\alpha/2}\sqrt{\frac{pq}{n}} < P < p + Z_{\alpha/2}\sqrt{\frac{pq}{n}} \Rightarrow 0.2 - 2.576 \times \sqrt{\frac{0.2 \times 0.8}{500}} < P < 0.2 + 2.576 \times \sqrt{\frac{0.2 \times 0.8}{500}}$$

$$0.154 < P < 0.246$$

$$p = \frac{100}{500} = 0.2$$

$$\alpha = \%1 \Rightarrow \frac{\alpha}{2} = 0.005 \Rightarrow Z_{0.005} = 2.576$$

$$e = \alpha = 0.01$$

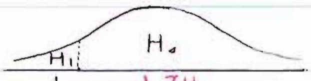
$$n = \left(\frac{Z_{\alpha/2}}{e}\right)^2 pq = \left(\frac{2.576}{0.01}\right)^2 \times 0.2 \times 0.8 \simeq 10618$$

$$10618 - 500 = 10118$$

پس 10118 نمونه کمتر انتخاب کردیم

آزمون یک جامعه

مثال 5 ادعا شده میانگین مدتی را که مراجعین به ایستگاه اتوبوس منتظر می مانند، کمتر از 10 دقیقه می باشد. یک نمونه به حجم 25 نفر انتخاب، میانگین و انحراف معیار زمان توقف آنها به ترتیب 9 و 5.0 دقیقه شده است. با فرض اینکه از این تعداد، 2 نفر دقیقاً 10 دقیقه، 13 نفر کمتر از 10 دقیقه و مابقی بیشتر از 10 دقیقه منتظر شده باشند. الف) با فرض نرمال، ب) بدون پیش فرض، ارزیابی شما چیست؟

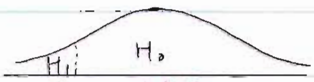



$-t_{24,0.05} = -1.711$

$H_0: \mu = 10$

$H_1: \mu < 10$

$t = \dfrac{\bar{x} - \mu_0}{S/\sqrt{n}} = \dfrac{9 - 10}{\frac{0.5}{\sqrt{25}}} = -10$ ادعا تأیید می‌شود




$-Z_{0.05} = 1.645$

$x = 10 \qquad n = 23$

$Z = \dfrac{x - np}{\sqrt{npq}} = \dfrac{10 - 23 \times \frac{1}{2}}{\sqrt{23 \times \frac{1}{2} \times \frac{1}{2}}} = -0.63$

ادعا تأیید می‌شود.

آزمون یک جامعه

مثال 6 از تعدادی که به تصادف انتخاب شده، سؤال برآمده که در هفته چند کیلو نان بربری مصرف می‌کنند. نتایج در جدول آمده است. اولاً) با فرض نرمال، آیا می‌توان پراکندگی مصرف هر شهروند را در هفته 0.5 کیلوگرم ادعا نمود؟ ثانیاً) بدون پیش فرض ادعا مبنی بر اینکه هر ایرانی در هفته حداقل 2.5 کیلو نان بربری مصرف می‌کند را تأیید می‌کنید؟

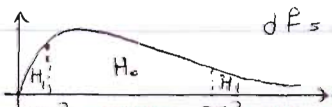
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3.5 | 0 | 4.5 | 3.2 | 2.3 | 2.8 | 3.2 | 3.8 |
| 6 | 1.2 | 3 | 1.5 | 1.8 | 3.1 | 4 | 1.2 |
| 0.2 | 2.5 | 1.7 | 3.6 | 3.5 | 3.2 | 3 | 3.4 |

الف) $n = 24 \qquad \bar{x} = \dfrac{66.2}{24} = 2.76$

$\sum (x - \bar{x})^2 = 0.5476 + 7.6176 + 3.0276 + 0.1936 + 0.2116 + 0.0016 + 0.1936 + 1.0816 + 10.4976$
$+ 2.4336 + 0.0576 + 1.5876 + 0.9216 + 0.1156 + 1.5376 + 2.4336 + 6.5536 + 0.0676 + 1.1236$
$+ 0.7056 + 0.5476 + 0.1936 + 0.0576 + 0.4096 = 42.1184$

$S = \sqrt{\dfrac{42.1184}{24 - 1}} = 1.35$

$df = n - 1 = 24 - 1 = 23$




$\chi^2_{df,0.975} = 11.6885 \qquad \chi^2_{df,0.025} = 38.0757$

$\begin{cases} H_0: \sigma^2 = 0.5^2 \\ H_1: \sigma^2 \neq 0.5^2 \end{cases}$

$\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} = \dfrac{(24-1)(1.35)^2}{0.5^2} = 167.67$

فرض $H_1$ تأیید می‌شود.

ب)


$-Z_{0.05} = 1.645$

داده‌های بیشتر از 2.5

$x = 15 \qquad n = 23$

$H_0: \tilde{M} \geq 2.5$

$H_1: \tilde{M} < 2.5$

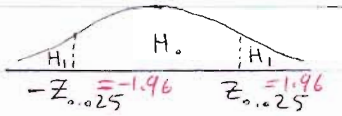
چون گفته حداقل 2.5 کیلو و مساوی هم شامل آن می‌شود، ادعا را در فرض $H_0$ قرار می‌دهیم.

$Z = \dfrac{x - np}{\sqrt{npq}} = \dfrac{15 - 23 \times \frac{1}{2}}{\sqrt{23 \times \frac{1}{2} \times \frac{1}{2}}} = 1.46$ فرض $H_0$ تأیید می‌شود

آزمون یک جامعه

مثال 7 برای اینکه ادعای سازمان ثبت احوال مبنی بر اینکه 20% میت‌های بهشت زهرا براساس تصادف بوده است را بررسی نمائیم 1000 میت به تصادف انتخاب می‌کنیم. 150 میت بر اثر تصادف فوت نموده‌اند. ارزیابی شما از این ادعا چیست؟

* در آزمون نسبت، ناپارامتریک نداریم. فقط در مورد میانگین ها آزمون ناپارامتریک داریم.




$$P=\frac{150}{1000}=0.15$$

$H_0: P=0.2$

$H_1: P\neq 0.2$

$$Z=\frac{P-P_0}{\sqrt{\frac{P_0 q_0}{n}}}=\frac{0.15-0.2}{\sqrt{\frac{0.2\times 0.8}{1000}}}=-3.95$$

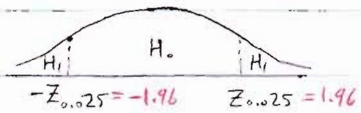
فرض $H_1$ تایید می شود

ادعای محقق تایید می شود.

آزمون دو جامعه

مثال 8 برای اینکه مقایسه ای بین قیمت دو نوع سهام داشته باشیم، 25 روز از سهام آلفا انتخاب، میانگین و انحراف معیار، 1800 و 75 تومان و از سهام بتا 30 روز کاری انتخاب، اعداد 1500 و 150 شده است. الف) در سطح $\alpha=\%5$، آیا بین میانگین قیمت این دو سهام، اختلاف وجود دارد یا خیر؟ ب) اختلاف آنها را برآورد کنید.

الف)




$H_0: \mu_1=\mu_2$

$H_1: \mu_1\neq\mu_2$

$$Z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}=\frac{1800-1500}{\sqrt{\frac{75^2}{25}+\frac{150^2}{30}}}=9.61$$

فرض $H_1$ تایید می شود

اولی 100٪ میانگین قیمتش بیشتر است.

ب)
$$(\bar{x}_1-\bar{x}_2)-Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$$(1800-1500)-1.96\sqrt{\frac{75^2}{25}+\frac{150^2}{30}}<\mu_1-\mu_2<(1800-1500)+1.96\sqrt{\frac{75^2}{25}+\frac{150^2}{30}}\Rightarrow 238.8<\mu_1-\mu_2<361.2$$

چون هر دو سمت هم علامتند، پس اختلاف معنی دار است.

آزمون دو جامعه

مثال 9 برای مقایسه عمر دو نوع لامپ تصویر تلویزیون، از هر کدام نمونه های تصادفی انتخاب و داده های جدول بدست آمده است. بدون پیش فرض نرمال، اولاً) اختلاف میانگین ها را برآورد نمائید. ثانیاً) ادعا مبنی بر اینکه میانگین طول عمر لامپ تصویر نوع دوم بیشتر است را چگونه ارزیابی می کنید.

| نوع اول | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | جمع |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | 22 | 24 | 31 | 35 | 20 | 24 | 23 | 27 | 20 | 25 | — | — | |
| رتبه | 4 | 6.5 | 19.5 | 22 | 2 | 6.5 | 5 | 12 | 2 | 8.5 | | | $R_1=88$ |
| نوع دوم | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | جمع |
| $x_2$ | 20 | 27 | 26 | 28 | 32 | 30 | 25 | 28 | 31 | 29 | 27 | 28 | |
| رتبه | 2 | 12 | 10 | 15 | 21 | 18 | 8.5 | 15 | 19.5 | 17 | 12 | 15 | $R_2=165$ |

الف)
$$\bar{x}_1=\frac{251}{10}=25.1 \qquad \bar{x}_2=\frac{331}{12}=27.58$$

$$\Sigma(x_i-\bar{x}_1)^2=9.61+1.21+34.81+98.01+26.01+1.21+4.41+3.61+26.01+0.01=204.9$$

$$S_1=\sqrt{\frac{204.9}{10-1}}=4.77$$

$$\sum(x_2-\bar{x}_2)^2=57.4564+0.3364+2.4964+0.1764+19.5364+5.8564+6.6564+0.1764+11.6964+2.0164+0.3364+0.1764=106.9168$$

$$S_2=\sqrt{\frac{106.9168}{12-1}}=3.12$$

$$(\bar{x}_1-\bar{x}_2)-\sqrt{\frac{1}{\alpha}}\times S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+\sqrt{\frac{1}{\alpha}}\times S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$S_p=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}=\sqrt{\frac{9\times 4.77^2+11\times 3.12^2}{10+12-2}}=3.95$$

$$(25.1-27.58)-\sqrt{\frac{1}{0.05}}\times 3.95\times\sqrt{\frac{1}{10}+\frac{1}{12}}<\mu_1-\mu_2<(25.1-27.58)+\sqrt{\frac{1}{0.05}}\times 3.95\times\sqrt{\frac{1}{10}+\frac{1}{12}}$$

$-10.04<\mu_1-\mu_2<5.08$ چون هر دو سمت هم علامت نیستند، اختلاف معنی‌دار نیست

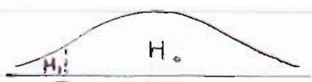
ب) $S_n=\frac{n}{2}(a_1+a_n)=\frac{22}{2}(1+22)=253$

$$U_1=n_1n_2+\frac{n_1(n_1+1)}{2}-\sum R_1=10\times 12+\frac{10(10+1)}{2}-88=87$$

$$U_2=n_1n_2+\frac{n_2(n_2+1)}{2}-\sum R_2=10\times 12+\frac{12(12+1)}{2}-165=33$$

$$E(U_1)=E(U_2)=\frac{n_1n_2}{2}=\frac{10\times 12}{2}=60$$

$$\sigma(U_1)=\sigma(U_2)=\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}=\sqrt{\frac{10\times 12(10+12+1)}{12}}=15.16$$




$-Z_{0.05}=-1.645$

چون ادعا روی μ اول است از $Z_2$ استفاده می‌شود

$H_0: \mu_1 \geqslant \mu_2$

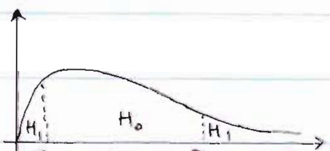
$H_1: \mu_1 < \mu_2$

$$Z_2=\frac{U_2-E(U_2)}{\sigma(U_2)}=\frac{33-60}{15.16}=-1.78$$ فرض $H_1$ تأیید می‌شود

آزمون دو طرفه

مثال 10 الف) در مثال قبل با فرض توزیع نرمال، آیا برابر واریانس‌ها را می‌توان ادعا نمود؟ ب) نسبت واریانس‌ها را برآورد بازه‌ای به عمل آورید.

الف)




$F_{9,11,0.975}=0.26$ $F_{9,11,0.025}=3.59$

$\frac{1}{F_{11,9,0.025}=3.87}$

$H_0: \sigma_1^2=\sigma_2^2$

$H_1: \sigma_1^2\neq\sigma_2^2$

$$F=\frac{S_1^2}{S_2^2}=\frac{4.77^2}{3.12^2}=2.34$$ ادعای برابری تأیید می‌شود.

ب)

$$\frac{S_1^2}{S_2^2}\times F_{df_2,df_1,\frac{\alpha}{2}}$$

$$\frac{\frac{S_1^2}{S_2^2}}{F_{df_1,df_2,\frac{\alpha}{2}}}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{\frac{S_1^2}{S_2^2}}{F_{df_1,df_2,1-\frac{\alpha}{2}}}\Rightarrow\frac{\frac{4.77^2}{3.12^2}}{3.59}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{4.77^2}{3.12^2}\times 3.87$$

$$\Rightarrow 0.65<\frac{\sigma_1^2}{\sigma_2^2}<9.05\Rightarrow 0.81<\frac{\sigma_1}{\sigma_2}<3$$

92/3/21

جلسه دهم

## آزمون تحلیل واریانس دو عامله بدون تأثیر متقابل:

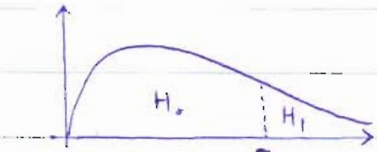
هرگاه بخواهیم دو معیار را در چند جامعه مقایسه نمائیم، چنانچه داده‌ها نرمال باشند، از این آزمون استفاده می‌کنیم. در این آزمون نیز از انحرافات کل، انحرافات تیمارها، انحرافات گروه‌ها را نیز محاسبه نموده، یک بار آماره F را برای تیمارها و یک بار آماره F را برای بلوک‌ها محاسبه می‌کنیم تا ارزیابی به عمل آید.

$$SST = \sum\sum x_{ij}^2 - \frac{1}{N}T^2$$

$$SSr = \frac{1}{n_i}\sum T_i^2 - \frac{1}{N}T^2 \qquad MSr = \frac{SSr}{k-1}$$

$$SS_B = \frac{1}{n_j}\sum T_j^2 - \frac{1}{N}T^2 \qquad MS_B = \frac{SS_B}{n-1}$$

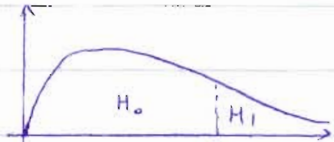
$$SSE = SST - SSr - SS_B \qquad MSE = \frac{SSE}{(k-1)(n-1)}$$




$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \cdots$

$H_1:$ حداقل میانگین دو تا از تیمارها با هم برابر نیستند.

$$F = \frac{MSr}{MSE}$$




$H'_0: \beta_1 = \beta_2 = \cdots$

$H'_1:$ حداقل میانگین دو تا از بلوکها با همدیگر برابر نیستند.

$$F = \frac{MS_B}{MSE}$$

$$L_{SD} = \sqrt{\frac{2MSE}{k}} \times t_{N-k, \alpha}$$

مثال: آیا قابلیت جناح‌های مختلف از نظر تحمل سیاسی یکسان است؟ آیا این تحمل در شهرهای مختلف متفاوت است؟ نمونه‌های تصادفی از تهران، تبریز، مشهد، یزد و سمنان به تصادف انتخاب می‌کنیم. نتایج در جدول آمده است. اولاً) آیا ظرفیت چهار جناح یکسان می‌باشند؟ ثانیاً) آیا ظرفیت پنج شهر یکسان می‌باشد؟ (با فرض توزیع نرمال)

| جناح / نمونه | تهران | تبریز | مشهد | یزد | سمنان | $T_i$ | $\bar{x}_i$ |
|---|---|---|---|---|---|---|---|
| ا.ت | 70 | 50 | 72 | 85 | 80 | 357 | 71.4 |
| م.ا | 60 | 70 | 60 | 90 | 75 | 355 | 71 |
| ط.ا | 65 | 60 | 80 | 95 | 60 | 360 | 72 |
| ی.ج | 70 | 60 | 65 | 80 | 70 | 345 | 69 |
| $T_j$ | 265 | 240 | 277 | 350 | 285 | 1417 | |
| $\bar{x}_j$ | 66.25 | 60 | 69.25 | 87.5 | 71.25 | | |

* همه SSها باید مثبت باشند.

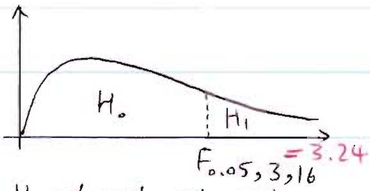
$$SST = (70^2+50^2+72^2+\dots+80^2+70^2) - \frac{1}{20}\times 1417^2 = 102{,}909 - 100{,}394.45 = 2{,}514.55$$

$$SSr = \frac{1}{5}(357^2+355^2+360^2+345^2) - \frac{1}{20}\times 1417^2 = 100{,}419.8 - 100{,}394.45 = 25.35$$

$$SS_B = \frac{1}{4}(265^2+240^2+277^2+350^2+285^2) - \frac{1}{20}\times 1417^2 = 102{,}069.75 - 100{,}394.45 = 1{,}675.3$$

$$SSE = 2{,}514.55 - 25.35 - 1{,}675.3 = 813.9$$

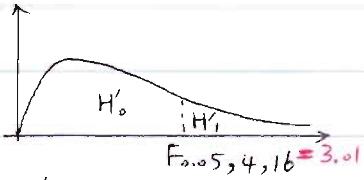
$$MSr = \frac{25.35}{4-1} = 8.45 \quad \| \quad MS_B = \frac{1{,}675.3}{5-1} = 418.825 \quad \| \quad MSE = \frac{813.9}{(4-1)(5-1)} = 67.825$$




$H_0: \alpha_1=\alpha_2=\alpha_3=\alpha_4$

$H_1:$ حداقل میانگین ظرفیت دو جناح با یکدیگر برابر نیستند

$F = \frac{8.45}{67.825} = 0.12$ فرض $H_0$ تأیید می‌شود




$H'_0: \beta_1=\beta_2=\beta_3=\beta_4=\beta_5$

$H'_1:$ حداقل میانگین ظرفیت دو شهر با یکدیگر برابر نیستند

$F' = \frac{418.825}{67.825} = 6.175$ فرض $H'_1$ تأیید می‌شود

$$L_{SD} = \sqrt{\frac{2\times 67.825}{4}} \times \overset{1.746}{t_{16,0.05}} = 10.17$$

$\bar{x}_{max} - L_{SD} = 87.5 - 10.17 = 77.33$ یزد

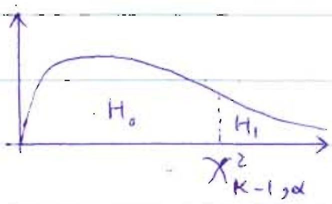
$77.33 - 10.17 = 67.16$ مشهد و سمنان

$67.16 - 10.17 = 56.99$ تهران و تبریز

## آزمون فریدمن (غیرنرمال):

این آزمون که به صورت رتبه‌ای می‌باشد، ابتدا رتبه داده‌ها را در سطرها مشخص، سپس جمع رتبه‌های هر ستون را بدست آورده و آماره آزمون F را محاسبه کرده تا بتوانیم ارزیابی صحیح به عمل آوریم.

$$F = \frac{12}{kn(n+1)}\sum R^2 - 3k(n+1)$$




مثال:

| | دری | فارسیجانی | کسایی | قلیچ لی | مؤمنی |
|---|---|---|---|---|---|
| احمد | 8(3.5) | 7(2) | 8(3.5) | 5(1) | 9(5) |
| نیره | 9(4) | 10(5) | 4(2) | 3(1) | 5(3) |
| غضنفر | 6(3) | 7(4) | 5(2) | 3(1) | 8(5) |
| سوسن | 8(4) | 8(4) | 6(2) | 5(1) | 8(4) |
| اصغر | 6(4) | 4(1) | 5(2.5) | 5(2.5) | 8(5) |
| | $R_1=18.5$ | $R_2=16$ | $R_3=12$ | $R_4=6.5$ | $R_5=22$ |

$\chi^2_{4,0.05} = 9.48773$

$H_0:$ رتبه تمام اساتید یکسان است

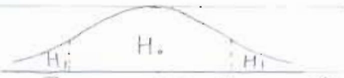
$H_1:$ حداقل رتبه دو تا از استادها با هم دیگر برابر نیستند

$$F = \frac{12}{5\times5(5+1)}(18.5^2+16^2+12^2+6.5^2+22^2) - 3\times5\times(5+1) = 101.48 - 90 = 11.48$$

فرض $H_1$ تأیید می‌شود

مثال 1. ۱۰٪ خانمها ارزیاب خوبی نیستند. از ۲۰۰ خانم نظرسنجی به عمل آمد (در مورد آقای x). 15 مورد اشتباه ارزیابی کردند. اولاً) ارزیابی شما از ادعای او چیست؟ ثانیاً) حجم نمونه انتخاب شده کافی بوده است؟

$P_0 = \%10 \qquad p = \frac{15}{200} = \%7.5$




$-Z_{0.025} = -1.96 \qquad Z_{0.025} = 1.96$

$Z = \frac{p - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.075 - 0.1}{\sqrt{\frac{0.1 \times 0.9}{200}}} = -1.18$ فرض $H_0$ تأیید می‌شود

$H_0 : P = \%10$

$H_1 : P \neq \%10$

$n = \left(\frac{Z_{\alpha/2}}{e}\right)^2 pq = \left(\frac{1.96}{0.05}\right)^2 \times 0.075 \times 0.925 \simeq 107 \qquad 200 - 107 = 93$ 93 نمونه بیشتر انتخاب شده است

مثال 2. برای بررسی میانگین تحمل خانمها در رابطه با صحبت نکردن، بررسی صورت گرفته است. میانگین تحمل آنها 25 دقیقه با پراکندگی 5 دقیقه شده است. اولاً) یک خانم به تصادف انتخاب می‌کنیم. احتمال اینکه دقیقاً 20 دقیقه تحمل کند، چقدر است؟ ثانیاً) احتمال اینکه بیشتر از 20 دقیقه تحمل کند چقدر است؟ ثالثاً) یک نمونه به حجم 25 خانم به تصادف انتخاب می‌کنیم. احتمال میانگین ظرفیت آنها بین 20 تا 30 دقیقه باشد، چقدر است؟

$\mu = 25 \qquad \sigma = 5$

$p(x = 20) = 0$

$p(x > 20) = p\left(Z > \frac{20 - 25}{5}\right) = p(Z > -1) = 1 - p(Z < -1) = 1 - 0.1587 = 0.8413$

$p(20 < \bar{x} < 30) = p\left(\frac{20 - 25}{\frac{5}{\sqrt{25}}} < Z < \frac{30 - 25}{\frac{5}{\sqrt{25}}}\right) = p(-5 < Z < 5) = p(Z < 5) - p(Z < -5) = 1 - 0 \simeq 1$

مثال 3. برای اینکه مقایسه‌ای بین ریسک خرید 2 نوع ماشین آلات به عمل آوریم، از دستگاه اول 25 ماشین انتخاب، میانگین و انحراف معیار زمان خواب دستگاه‌ها در هفته به ترتیب 50 و 10 دقیقه و از دستگاه دوم 20 ماشین انتخاب، و مقادیر فوق به ترتیب 40 و 6 دقیقه شده است. الف) آیا می‌توان ادعا نمود میانگین خواب دستگاه اول 5 دقیقه بیشتر از دستگاه دوم می‌باشد؟ ثانیاً) از طریق برآورد، پراکندگی زمان خواب دستگاه‌ها را ارزیابی کنید؟ (با فرض توزیع نرمال)

$n_1 = 25 \qquad \bar{x}_1 = 50 \qquad S_1 = 10 \qquad n_2 = 20 \qquad \bar{x}_2 = 40 \qquad S_2 = 6$


$-Z_{0.025} = -1.96 \qquad Z_{0.025} = 1.96$

$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{50 - (40 + 5)}{\sqrt{\frac{10^2}{25} + \frac{6^2}{20}}} = 2.08$ فرض $H_1$ تأیید می‌شود

$H_0 : \mu_1 = \mu_2 + 5$

$H_1 : \mu_1 \neq \mu_2 + 5$

$$\frac{\frac{S_1^2}{S_2^2}}{F_{df_1, df_2, \alpha/2}} < \sigma^2 < \frac{\frac{S_1^2}{S_2^2}}{F_{df_1, df_2, 1-\alpha/2}} \Rightarrow \frac{\frac{S_1^2}{S_2^2}}{F_{df_1, df_2, \alpha/2}} < \sigma^2 < \frac{S_1^2}{S_2^2} \times F_{df_2, df_1, \alpha/2}$$

$$\Rightarrow \frac{\frac{10^2}{6^2}}{F_{24, 19, 0.025}} < \sigma^2 < \frac{10^2}{6^2} \times F_{19, 24, 0.025} \Rightarrow \frac{\frac{100}{36}}{2.45} < \sigma^2 < \frac{100}{36} \times 2.33 \Rightarrow 1.13 < \sigma^2 < 6.47$$

مثال 4 می خواهیم بررسی نمائیم آیا تغییرات مدیریتی سبب کاهش ضایعات گردیده است یا خیر. پنج روز قبل از اعمال تغییرات و پنج روز بعد از اعمال تغییرات، ضایعات 12 دستگاه را طی 3 روز ثبت نمودیم و نتایج جدول حاصل شده است. اولاً) با فرض توزیع نرمال، ثانیاً) بدون پیش فرض آیا تغییرات مدیریتی سبب کاهش ضایعات گردیده است یا خیر؟

| دستگاه | قبل از تغییرات $X_1$ | بعد از تغییرات $X_2$ | $d$ | $(d-\bar{d})^2$ | علامت |
|---|---|---|---|---|---|
| 1 | 0.05 | 0.04 | -0.01 | 0 | — |
| 2 | 0.02 | 0.02 | 0 | 0.00008 | × |
| 3 | 0.07 | 0.06 | -0.01 | 0 | — |
| 4 | 0.04 | 0.05 | 0.01 | 0.0004 | + |
| 5 | 0.06 | 0.07 | 0.01 | 0.0004 | + |
| 6 | 0.03 | 0.02 | -0.01 | 0 | — |
| 7 | 0.06 | 0.04 | -0.02 | 0.0001 | — |
| 8 | 0.05 | 0.03 | -0.02 | 0.0001 | — |
| 9 | 0.04 | 0.02 | -0.02 | 0.0001 | — |
| 10 | 0.12 | 0.10 | -0.02 | 0.0001 | — |
| 11 | 0.08 | 0.06 | -0.02 | 0.0001 | — |
| 12 | 0.07 | 0.07 | 0 | 0.00008 | × |
| | | | -0.11 | 0.00146 | |


$-t_{11,0.05} = -1.796$

$H_0 : \mu_d = 0$

$H_1 : \mu_d < 0$

$\bar{d} = \frac{-0.11}{12} = -0.0092$

$S_d = \sqrt{\frac{\sum (d-\bar{d})^2}{n-1}} = 0.012$

$t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{-0.0092}{\frac{0.012}{\sqrt{12}}} = -2.66$ فرض $H_1$ تأیید می شود


$-Z_{0.05} = -1.645$

$H_0 : \tilde{M}_1 = \tilde{M}_2$

$H_1 : \tilde{M}_1 < \tilde{M}_2$

$x = 2 \qquad n = 10$

$Z = \frac{x - np}{\sqrt{npq}} = \frac{2 - 10 \times 1/2}{\sqrt{10 \times \frac{1}{2} \times 1/2}} = -1.9$ فرض $H_1$ تأیید می شود

مثال 5 در نظر است مقایسه ای بین میانگین عمر لامپهای دو برند A و B به عمل بیاوریم. از هر کدام نمونه های تصادفی انتخاب، نتایج جدول حاصل شده است. در سطح $\alpha = 5\%$ (بدون پیش فرض نرمال) آیا برابری میانگین عمر آنها را می توان ادعا نمود؟

* همیشه دومی را از اولی کم می کنیم. ← اگر یکی داشته بود.

* اگر نگفت با پیش فرض نرمال، نرمال نیست

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | جمع |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (ماه) $X_1$ | 31 | 28 | 45 | 15 | 22 | 27 | 26 | 34 | 30 | 30 | 25 | 22 | |
| رتبه | 18 | 15 | 22 | 1 | 7 | 13.5 | 12 | 20 | 16.5 | 16.5 | 10.5 | 7 | $\sum R_1 = 159$ |
| (ماه) $X_2$ | 18 | 25 | 23 | 36 | 32 | 27 | 22 | 20 | 18 | 21 | — | — | |
| رتبه | 2.5 | 10.5 | 9 | 21 | 19 | 13.5 | 7 | 4 | 2.5 | 5 | | | $\sum R_2 = 94$ |

$S_n = \frac{n}{2}(a_1 + a_n) = \frac{22}{2}(1+22) = 253$


$-Z_{0.025} = -1.96 \qquad Z_{0.025} = 1.96$

$H_0 : \tilde{M}_1 = \tilde{M}_2$

$H_1 : \tilde{M}_1 \neq \tilde{M}_2$

$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1 = 12 \times 10 + \frac{12(12+1)}{2} - 159 = 39$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2 = 12 \times 10 + \frac{10(10+1)}{2} - 94 = 81$

$E(U_1) = E(U_2) = \frac{n_1 n_2}{2} = \frac{12 \times 10}{2} = 60$

$\sigma(U_1) = \sigma(U_2) = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{12 \times 10 (12 + 10 + 1)}{12}} = 15.17$

$Z_1 = \frac{U_1 - E(U_1)}{\sigma(U_1)} = \frac{39 - 60}{15.17} = -1.4$ فرض $H_0$ تأیید می شود

$Z_2 = \frac{U_2 - E(U_2)}{\sigma(U_2)} = \frac{81 - 60}{15.17} = 1.4$ فرض $H_0$ تأیید می شود

## کاربردهای آزمون های $\chi^2$ :

از این آزمون در مورد استقلال و وابستگی ، آزمون نسبت چند جامعه (آزمون همگونی) و آزمون نیکویی برازش (آزمون نوع توزیع داده ها) مانند نرمال بودن ، نمایی بودن و ... استفاده می گردد:

### آزمون همگونی (کاربرد اول) :

اگر بخواهیم بررسی در مورد برابری یا عدم برابری نسبت های چندجامعه ، عمل بیاوریم ، ابتدا جدول توافقی را تشکیل می دهیم. در این جدول ، در سطر ، گروهها و در ستون ، گزینه ها که به صورت دو پاسخی می باشد (برنولی) ، قرار گرفته و اعداد داخل جدول ، فراوانی پاسخها می باشد که به آن فراوانی مشاهده شده می گوییم ($F_o$) . براساس فراوانی مشاهده شده ، فراوانی مورد انتظار ($F_e$) را از رابطه زیر بدست می آوریم:

$$F_e = \frac{\overbrace{\sum r}^{\text{row جمع سطرها}} \cdot \overbrace{\sum c}^{\text{column جمع ستونها}}}{\underbrace{\sum F}_{\text{Frequency فراوانی}}}$$

پس از بدست آوردن فراوانی مورد انتظار ، آماره $\chi^2$ را بدست می آوریم:

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

فرض $H_0$ و $H_1$ آن به صورت ذیل می باشد:

$$\begin{cases} H_0 : P_1 = P_2 = \cdots \\ H_1 : \text{حداقل نسبت دوتا از گروهها برابر نیستند} \end{cases}$$

---

92/3/28

### جلسه یازدهم

مثال: می خواهیم بررسی نمائیم آیا نسبت رأی دهندگان به رئیس جمهور منتخب در نزد چهار گروه تحصیلی یکسان بوده است. برای این منظور نمونه های تصادفی انتخاب و نتایج در جدول آمده است. در سطح $\alpha = 5\%$ آیا برابری نظرات را می توان پذیرفت؟

| | مثبت | منفی | |
|---|---|---|---|
| دیپلم و کمتر | 80 / 67.2 | 40 / 52.8 | 120 |
| کمتر از فوق دیپلم | 80 / 72.8 | 50 / 57.2 | 130 |
| کمتر از لیسانس | 70 / 84 | 80 / 66 | 150 |
| کمتر از دکتری | 50 / 56 | 50 / 44 | 100 |
| | 280 | 220 | 500 |

* در آزمون همگونی فقط دو ستون وجود دارد.
* داده های داخلی جدول $F_o$ هستند.

$H_0 : P_1 = P_2 = P_3 = P_4$

$H_1$: حداقل نسبت نظرات دوتا از گروههای تحصیلی با هم دیگر متفاوت است

$$F_{11} = \frac{120 \times 280}{500} = 67.2 \qquad F_{12} = \frac{120 \times 220}{500} = 52.8$$

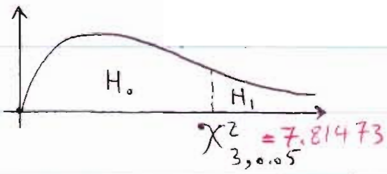
$$F_{21} = \frac{130 \times 280}{500} = 72.8 \qquad F_{22} = \frac{130 \times 220}{500} = 57.2$$

$$F_{31} = \frac{150 \times 280}{500} = 84 \qquad F_{32} = \frac{150 \times 220}{500} = 66$$

$$F_{41} = \frac{100 \times 280}{500} = 56 \qquad F_{42} = \frac{100 \times 220}{500} = 44$$

$$\chi^2 = \frac{(80-67.2)^2}{67.2} + \frac{(40-52.8)^2}{52.8} + \frac{(80-72.8)^2}{72.8} + \frac{(50-57.2)^2}{57.2} + \frac{(70-84)^2}{84} + \frac{(80-66)^2}{66} + \frac{(50-56)^2}{56}$$

$$+ \frac{(50-44)^2}{44} = 13.92$$




$$df = (r-1)(c-1) = (4-1)(2-1) = 3$$

$$\chi^2_{3,0.05} = 7.81473$$ ...... فرض $H_1$ تأیید می‌شود.

آزمون استقلال و وابستگی (کاربرد دوم):

هرگاه بخواهیم رابطه بین دو متغیر یا معیار را بسنجیم ، اگر حداقل یکی از معیارها کیفی باشد ، از این آزمون استفاده می‌کنیم. (اگر هر دو معیار کمی باشد ، از آزمون ضریب همبستگی استفاده می‌شود) . فرآیند این آزمون عیناً مانند آزمون همگونی است ، تنها در نوشتن فرض $H_0$ و $H_1$ تفاوت وجود دارد.

مثال: می‌خواهیم بررسی نمائیم بین عملکرد و میزان رضایتمندی رابطه وجود دارد یا خیر. دو پرسشنامه توزیع ، عملکرد و رضایتمندی را اندازه گرفتیم. با توجه به نتایج در سطح $\alpha = \%5$ ، ارزیابی شما چیست ؟

| رضایتمندی / عملکرد | کم | متوسط | زیاد | |
|---|---|---|---|---|
| ضعیف | 40 / 29.3 | 25 / 25.3 | 15 / 25.3 | 80 |
| متوسط | 50 / 44 | 40 / 38 | 30 / 38 | 120 |
| قوی | 20 / 36.7 | 30 / 31.7 | 50 / 31.7 | 100 |
| | 110 | 95 | 95 | 300 |

* فرق این آزمون با آزمون همگونی این است که در آزمون همگونی دو ستون داریم ولی در آزمون استقلال و وابستگی ، سه ستون یا بیشتر وجود دارد. رابطه بین رضایتمندی و عملکرد است.

$H_0$: بین میزان عملکرد و میزان رضایتمندی رابطه وجود ندارد

$H_1$: بین میزان عملکرد و میزان رضایتمندی رابطه وجود دارد.

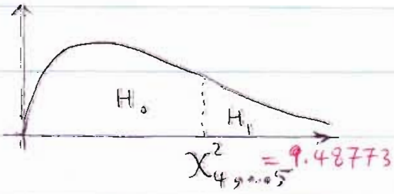
$$F_{11} = \frac{80 \times 110}{300} = 29.3 \qquad F_{21} = \frac{120 \times 110}{300} = 44 \qquad F_{31} = \frac{100 \times 110}{300} = 36.7$$

$$F_{12} = \frac{80 \times 95}{300} = 25.3 \qquad F_{22} = \frac{120 \times 95}{300} = 38 \qquad F_{32} = \frac{100 \times 95}{300} = 31.7$$

$$F_{13} = \frac{80 \times 95}{300} = 25.3 \qquad F_{23} = \frac{120 \times 95}{300} = 38 \qquad F_{33} = \frac{100 \times 95}{300} = 31.7$$

$$\chi^2 = \frac{(40-29.3)^2}{29.3} + \frac{(25-25.3)^2}{25.3} + \frac{(15-25.3)^2}{25.3} + \frac{(50-44)^2}{44} + \frac{(40-38)^2}{38} + \frac{(30-38)^2}{38} + \frac{(20-36.7)^2}{36.7}$$

$$+ \frac{(30-31.7)^2}{31.7} + \frac{(50-31.7)^2}{31.7} = 27.7$$




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

$$\chi^2_{4,0.05} = 9.48773$$ فرض $H_1$ تأیید می‌شود

## آزمون نیکویی برازش (آزمون نوع تابع توزیع) (کاربرد سوم):

برای اینکه تشخیص دهیم داده‌ای نرمال یا پواسون و یا دوجمله‌ای می‌باشد، از این آزمون و آزمون کالموگراف - اسمیرنف (KS : Kalmograph - Smirinov) استفاده می‌شود. برای آزمون $\chi^2$ ابتدا اگر میانگین و انحراف معیار در اختیار نباشد آنها را محاسبه می‌کنیم، سپس برای هر طبقه $Z$ را تعیین می‌کنیم. سپس به جدول توزیع نرمال مراجعه می‌کنیم تا احتمال هر مقدار را بدست بیاوریم. با توجه به اینکه این احتمالات به صورت تجمعی می‌باشد، آنها را در مجموعه فراوانی‌ها ضرب می‌کنیم تا فراوانی تجمعی مورد انتظار حاصل شود. برای اینکه فراوانی مورد انتظار هر طبقه را بدست بیاوریم، فراوانی تجمعی هر طبقه را منهای فراوانی تجمعی طبقه قبل نموده تا فراوانی مورد انتظار هر طبقه بدست آید. پس از آن آماره $\chi^2$ را محاسبه می‌کنیم و برای اینکه مقدار بحرانی بدست بیاوریم، با درجه آزادی $K-3$ و $\alpha$ ($\chi^2_{(K-3)(\alpha)}$) به جدول مربوطه مراجعه می‌کنیم که در اینجا فرض $H_0$ فرض نرمال بودن می‌باشد.
اگر احیاناً فراوانی مورد انتظار طبقه‌ای کمتر از 5 باشد، آن را با طبقه مجاور خود تجمیع می‌نمائیم.

اما برای آزمون KS، پس از اینکه احتمال هر $Z$ را از جدول بدست آوردیم، بر اساس فراوانی مشاهده شده، فراوانی نسبی را محاسبه کرده، تجمعی آن بدست می‌آوریم تا بتوانیم آماره $D$ را بدست آورده، مقدار آن را با مقدار بحرانی جدول که از جدول KS بدست می‌آید، مقایسه کنیم.

$$D = \max |P_{C_o} - P_{C_e}| \qquad D_{N, \alpha}$$

مثال: می‌خواهیم بررسی نمائیم که آیا EPS توزیع شده شرکتهای پذیرفته شده، از توزیع نرمال تبعیت می‌کنند یا خیر (میانگین 50 با پراکندگی 10). یک نمونه به حجم 100 شرکت به تصادف انتخاب می‌کنیم. توزیع EPS آنها مانند جدول زیر می‌باشد. در سطح $\alpha = 5\%$، آیا نرمال بودن توزیع را می‌توان پذیرفت؟

احتمال — $P_{Ce} \times N$ — $F_{ce}$ هر طبقه را از طبقه قبلی آن کم می‌کنیم — با توزیع $F_e$ عمل می‌کنیم — اگر فراوانی طبقه‌ای کمتر از 5 بود، با طبقه بعدی جمع می‌کنیم تا بیشتر از 5 شود

| $x$ | $F_o$ | $Z$ | $P_{Ce}$ | $F_{Ce}$ | $F_e$ | $F'_o$ | $F'_e$ | $P_o$ | $P_{Co}$ | $\lvert P_{Ce} - P_{Co} \rvert$ | $\frac{(F'_o - F'_e)^2}{F'_e}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| کمتر از 150 | 8 | -3.5 | 0.0002 | 0.02 | 0.02 | — | — | 0.08 | 0.08 | 0.0798 | — |
| 150-250 | 10 | -2.5 | 0.0062 | 0.62 | 0.6 | — | — | 0.1 | 0.18 | 0.1738 | — |
| 250-350 | 12 | -1.5 | 0.0668 | 6.68 | 6.06 | 30 | 6.68 | 0.12 | 0.3 | 0.2332 | $\frac{(30-6.68)^2}{6.68} = 81.41$ |
| 350-450 | 15 | -0.5 | 0.3085 | 30.85 | 24.17 | 15 | 24.17 | 0.15 | 0.45 | 0.1415 | $\frac{(15-24.17)^2}{24.17} = 3.48$ |
| 450-550 | 20 | 0.5 | 0.6915 | 69.15 | 38.3 | 20 | 38.3 | 0.2 | 0.65 | 0.0415 | $\frac{(20-38.3)^2}{38.3} = 8.74$ |
| 550-650 | 12 | 1.5 | 0.9332 | 93.32 | 24.17 | 12 | 24.17 | 0.12 | 0.77 | 0.1632 | $\frac{(12-24.17)^2}{24.17} = 6.13$ |
| 650-750 | 10 | 2.5 | 0.9938 | 99.38 | 6.06 | 23 | 6.68 | 0.1 | 0.87 | 0.1238 | $\frac{(23-6.68)^2}{6.68} = 39.87$ |
| 750-850 | 7 | 3.5 | 0.9998 | 99.98 | 0.6 | — | — | 0.07 | 0.94 | 0.0598 | — |
| 850-950 | 4 | 4.5 | 1 | 100 | 0.02 | — | — | 0.04 | 0.98 | 0.02 | — |
| بیشتر از 950 | 2 | 5.5 | 1 | 100 | 0 | — | — | 0.02 | 1 | 0 | — |
| | $N=100$ | | | | | | | | | | 139.63 |

* اگر حد پائین و بالا طبقه مشخص نباشد، میانه شاخص خوبی است.

* اگر میانگین را ندهد و حد بالا و پائین طبقه مشخص باشد، باید نماینده طبقه مشخص کنیم و میانگین بگیریم.

$H_0 : x \sim N(500, 100)$

$H_1 : x \nsim N(500, 100)$

$Z_1 = \frac{150-500}{100} = -3.5$ , $Z_2 = \frac{250-500}{100} = -2.5$ , $Z_3 = \frac{350-500}{100} = -1.5$ , $Z_4 = \frac{450-500}{100} = -0.5$

$Z_5 = \frac{550-500}{100} = 0.5$ , $Z_6 = \frac{650-500}{100} = 1.5$ , $Z_7 = \frac{750-500}{100} = 2.5$ , $Z_8 = \frac{850-500}{100} = 3.5$

$Z_9 = \frac{950-500}{100} = 4.5$ , $Z_{10} = \frac{1050-500}{100} = 5.5$

$df = K-3 = 5-3 = 2$ $\quad \chi^2_{2, 0.05} = 5.99147$ فرض $H_1$ تایید می‌شود

$D = \max|P_{Co} - P_{Ce}| = 0.2332 \Rightarrow$ از جدول صفحه قبل

$D = \frac{1.36}{\sqrt{100}} = 0.136 \Rightarrow$ از جدول کتاب

چون $N$ بزرگتر از 35 است، از روابط پائین جدول استفاده می‌کنیم

$0.2332 > 0.136$ پس نرمال نمی‌باشد

## رگرسیون و ضریب همبستگی:

از جمله مدلهای روندی می‌باشد و برای پیش‌بینی استفاده می‌گردد (اخیراً شبکه‌های عصبی، مدلهای داده‌کاوی و ... برای پیش‌بینی مطرح شده است، به خصوص اگر Data Base یا تعداد داده‌ها بسیار زیاد باشد). این نوع مدلها به صورت یک مستقل، یک وابسته؛ چند مستقل، یک وابسته؛ یک مستقل، چند وابسته؛ چند مستقل، چند وابسته؛ هم به صورت خطی و هم غیر خطی مطرح می‌گردد. هم به صورت توصیفی و هم استنباطی، داده‌ها مورد بررسی قرار می‌گیرد.

### مدل رگرسیون خطی یک متغیره (یا روش کمترین مجذورات):

هرگاه رابطه بین $x$ و $y$ خطی باشد، از این رگرسیون استفاده می‌کنیم و برای بدست آوردن دو پارامتر $a$ و $b$ از رابطه‌های زیر استفاده می‌کنیم:

$$y = a + bx$$

$$\begin{cases} \sum y = \overbrace{\sum a}^{na} + b\sum x \\ \sum xy = a\sum x + b\sum x^2 \end{cases} \qquad \begin{cases} b = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} \\ a = \bar{y} - b\bar{x} \end{cases} \qquad \begin{cases} b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \\ a = \bar{y} - b\bar{x} \end{cases}$$

همچنین برای محاسبه ضریب همبستگی از یکی از دو رابطه زیر استفاده می‌شود:

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{(\sum x^2 - n\bar{x}^2)(\sum y^2 - n\bar{y}^2)}} \qquad r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

$-1 < r < 1 \Rightarrow \begin{cases} \text{مثبت: رابطه مستقیم} \\ \text{منفی: رابطه معکوس} \end{cases}$

| | |
|---|---|
| $\lvert 0\rvert - \lvert 0.2\rvert$ | خیلی ضعیف |
| $\lvert 0.2\rvert - \lvert 0.4\rvert$ | ضعیف |
| $\lvert 0.4\rvert - \lvert 0.6\rvert$ | متوسط |
| $\lvert 0.6\rvert - \lvert 0.8\rvert$ | خوب |
| $\lvert 0.8\rvert - \lvert 1\rvert$ | خیلی خوب |