

فهرست

1	1- تعاریف
1	1-1- تعریف جامعه (Population)
1	2-1- تعریف پارامتر (Parameter)
1	3-1- تعریف نمونه (Sample)
2	4-1- اندازه نمونه (Sample Size)
2	5-1- تعریف نمونه تصادفی (Random Sample)
2	6-1- توزیع نمونه
2	7-1- تعریف آماره
3	8-1- میانگین نمونه
3	9-1- گشتاور نمونه
3	10-1- گشتاور r ام حول میانگین
4	11-1- واریانس نمونه
4	12-1- انحراف معیار نمونه
5	13-1- قضیه حد مرکزی
5	14-1- توزیع Gamma
6	2- توزیع مربع کای (Chi Square)
10	3- توزیع t_{student}
13	4- توزیع تفاوت بین دو میانگین نمونه
16	5- توزیع F
19	6- توزیع میانگین جامعه های متنهای
22	7- آماره های ترتیبی

1- تعاریف**1-1- تعریف جامعه (Population)**

جامعه مجموعه عناصر مورد مطالعه است که ممکن است تعداد آن محدود یا نامحدود باشد. مثلاً جامعه متولدین در یک سال و یا قیمت یک کالا در طول زمان معین.

1-2- تعریف پارامتر (Parameter)

مقدار ثابتی مربوط به جامعه آماری است که معمولاً مورد علاقه بررسی کننده واقع می شود. مثلاً در مورد لامپ ها، متوسط طول عمر لامپ ها.

نکته: موقعی روش های آماری مفید است که پارامتر مجهول داشته باشیم.

X	0	1	2	مدل آمار
P	$\theta/2$	$1-\theta$	$\theta/2$	

X	0	1	2	مدل
P	$1/4$	$1/2$	$1/4$	

1-3- تعریف نمونه (Sample)

در حالاتی که به جامعه دسترسی نداریم و یا به علت محدودیت در زمان و هزینه، بناچار مطالعه و بررسی بر روی بخشی از جامعه به عنوان نمونه انجام می شود. نمونه باید به گونه‌ای باشد که معرف جامعه اصلی بوده و نتایج حاصل از آن بتواند قابل تعمیم به جامعه برگرفته از آن باشد.

الف- نمونه تصادفی از جامعه متناهی بدون جاگذاری

در این حالت نمونه باید طوری انتخاب شود که احتمال انتخاب تمام نمونه‌های ممکن برابر باشد.

مثلا اگر جامعه دارای N عضو است انتخاب هر نمونه n تایی با احتمال $\frac{1}{\binom{N}{n}}$ رخ می‌دهد.

تذکر: معمولا اگر در جریان انتخاب هیچ عاملی یا محدودیت و شرطی دخالت نکند نمونه تصادفی خواهد بود.

ب- نمونه تصادفی از جامعه نامتناهی (متناهی با جاگذاری)

مقدار هر متغیر که در نمونه تصادفی ظاهر می‌شود مربوط به یک متغیر تصادفی است و این متغیر دارای یک توزیع است.

4-1- اندازه نمونه (Sample Size)

تعداد افراد نمونه را حجم نمونه یا اندازه نمونه می‌نامند.

5-1- تعریف نمونه تصادفی (Random Sample)

متغیرهای تصادفی x_1, x_2, \dots, x_n تشکیل یک نمونه تصادفی به اندازه n می‌دهند اگر:

1- هم توزیع باشند.

2- از یکدیگر مستقل باشند.

6-1- توزیع نمونه

اگر x_1, x_2, \dots, x_n نشان دهنده نمونه‌ای به حجم n باشد بنا به تعریف، تابع توزیع توام

x_1, x_2, \dots, x_n را تابع توزیع نمونه می‌نامند. لذا اگر x_1, x_2, \dots, x_n یک نمونه تصادفی به حجم n از

جامعه $f_x(x)$ باشد در این صورت توزیع نمونه تصادفی x_1, x_2, \dots, x_n عبارتست از:

$$f(x_1, x_2, \dots, x_n) = f_{x_1}(x_1) f_{x_2}(x_2) \dots f_{x_n}(x_n) = \prod_{i=1}^n f_{x_i}(x_i) = [f_x(x)]^n$$

7-1- تعریف آماره

تابعی است از نمونه تصادفی قابل مشاهده که شامل پارامتر مجهولی نیست (به پارامترهای نامعلوم

بستگی ندارد) مثلاً در یک نمونه‌گیری به اندازه n ، \bar{X} ، S^2 و S همه آماره‌اند. در این ارتباط توزیع آماره‌ها را توزیع نمونه‌ای می‌گویند و استنباطهای آماری معمولاً بر آماره‌ها متکی است.

8-1- میانگین نمونه

اگر x_1, x_2, \dots, x_n یک نمونه تصادفی از جامعه‌ای به حجم n باشد:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

9-1- گشتاور نمونه

فرض می‌کنیم x_1, x_2, \dots, x_n یک نمونه تصادفی از جامعه با تابع چگالی احتمال $f_x(x)$ باشند. در این صورت r مین گشتاور نمونه حول مبدا برابر است با:

$$M'_r = \frac{1}{n} \sum_{i=1}^n X_i^r$$

که در حالت خاص $r=1$ ، میانگین نمونه بدست می‌آید.

10-1- گشتاور r ام حول میانگین

$$M_r = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^r \quad \text{گشتاور } r \text{ ام حول میانگین برابر است با:}$$

قضیه 1: فرض کنید x_1, x_2, \dots, x_n یک نمونه تصادفی از جامعه با تابع چگالی $f_x(x)$ باشد و

$$E[\bar{X}] = \mu \quad , \quad \text{Var}[\bar{X}] = \frac{\sigma^2}{n} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{میانگین نمونه باشد در این صورت:}$$

که در آن μ و σ^2 به ترتیب میانگین و واریانس جامعه هستند.

اثبات:

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} E\left[\sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n\mu = \mu$$

$$\text{Var}[\bar{X}] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n x_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

11-1- واریانس نمونه

فرض کنید x_1, x_2, \dots, x_n یک نمونه تصادفی از جامعه $f_x(x)$ باشد در این صورت:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 ; n > 1$$

واریانس نمونه نامیده می‌شود.

12-1- انحراف معیار نمونه

انحراف معیار n مشاهده x_1, x_2, \dots, x_n را به عنوان ریشه دوم واریانسهایشان تعریف می‌کنیم:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

قضیه 2: اگر تمام نمونه‌های تصادفی به حجم n از یک جامعه نرمال با میانگین μ و واریانس σ^2

باشد در این صورت \bar{X} دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

$$M_{\bar{X}} = M_X[(\sum x_i / n)t] = (M_X(t/n))^n = (e^{\mu(t/n) + 1/2(t/n)^2 \sigma^2})^n = e^{\mu t + \frac{1}{2}t^2(\sigma^2/n)}$$

قضیه 3: اگر تمام نمونه‌های تصادفی به حجم n از یک جامعه با میانگین μ و واریانس σ^2 با

جایگذاری انتخاب شود، توزیع نمونه \bar{X} تقریباً دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$

است.

نکته:

اگر $n \geq 30$ باشد، بدون توجه به نوع توزیع جامعه، توزیع \bar{X} نرمال است.

اگر $n < 30$ باشد، توزیع \bar{X} تقریباً نرمال است اگر توزیع جامعه تقریباً نرمال باشد.

13-1- قضیه حد مرکزی

اگر x_1, x_2, \dots, x_n نمونه‌ای تصادفی از یک جامعه نامتناهی با میانگین μ و واریانس σ^2 باشد

در این صورت توزیع حدی متغیر

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

وقتی $n \rightarrow \infty$ ، توزیع نرمال استاندارد است.

14-1- توزیع Gamma

اگر متغیر تصادفی گاما دارای تابع چگالی ذیل است:

$$f_x(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{s-1}}{\Gamma(s)} \quad ; \quad x > 0$$

$$\phi(t) = \left(\frac{\lambda}{\lambda - t}\right)^s; E[x] = \frac{s}{\lambda}; \text{Var}(x) = \frac{s}{\lambda^2}$$

2- توزیع مربع کای (Chi Square)

توزیع مربع کای حالت خاصی از توزیع گاما است که در آن $\lambda = \frac{1}{2}; s = \frac{v}{2}$ است و بر این اساس تابع چگالی احتمال متغیر تصادفی مربع کای بشرح ذیل است.

$$f_x(x) = \frac{1}{2^{v/2} \Gamma\left(\frac{v}{2}\right)} x^{\left(\frac{v-2}{2}\right)} e^{-\frac{x}{2}} ; x > 0$$

که Γ معرف تابع گاما و معرف درجه آزادی است و

$$E(x^2) = v, \text{Var}(x^2) = 2v, M_x(t) = \phi(t) = (1-2t)^{-\frac{v}{2}}$$

قضیه 4: اگر متغیر تصادفی Z_1 دارای توزیع نرمال استاندارد باشد، آنگاه متغیر تصادفی Z_1^2 دارای توزیع مربع کای با 1 درجه آزادی خواهد بود. این قضیه در واقع بیانگر اهمیت توزیع مربع کای است.

قضیه 5: اگر $\chi_1^2, \chi_2^2, \dots, \chi_k^2$ متغیرهای تصادفی مستقل هر یک دارای توزیع مربع کای با درجات آزادی به ترتیب v_1, v_2, \dots, v_k باشند، آنگاه متغیر تصادفی $\chi^2 = \chi_1^2 + \chi_2^2 + \dots + \chi_k^2$ نیز یک توزیع مربع کای با $v = \sum_{i=1}^k v_i$ درجه آزادی دارد.

اثبات:

$$M_{x_i}(t) = (1-2t)^{-v_i/2}; M_{\sum x_i}(t) = \prod_{i=1}^n (1-2t)^{-v_i/2} = (1-2t)^{-\sum v_i/2}$$

قضیه 6: اگر متغیرهای تصادفی مستقل Z_v, \dots, Z_2, Z_1 دارای توزیع نرمال استاندارد باشند، آنگاه متغیر تصادفی $\chi^2 = Z_1^2 + Z_2^2 + \dots + Z_v^2$ یک توزیع مربع کای با v درجه آزادی خواهد بود.

تذکر: اگر x_1, x_2, \dots, x_v متغیرهای تصادفی مستقل نرمال به ترتیب با میانگین‌های $\mu_1, \mu_2, \dots, \mu_v$ و واریانسهای $\sigma_1^2, \sigma_2^2, \dots, \sigma_v^2$ باشند، جمع مربعات X ها یک توزیع مربع کای ندارد.

قضیه 7: اگر متغیرهای تصادفی x_1, x_2, \dots, x_v دارای توزیع نرمال مستقل با میانگین‌های

$$Z = \frac{x_i - \mu_i}{\sigma_i}$$

و واریانس‌های $\sigma_1^2, \sigma_2^2, \dots, \sigma_v^2$ باشند در این صورت متغیر تصادفی

برای $i=1,2,\dots,v$ دارای توزیع نرمال استاندارد می باشد و متغیر تصادفی:

$$\chi^2 = \sum_{i=1}^v \left(\frac{X_i - \mu_i}{\delta_i} \right)^2$$

دارای توزیع مربع کای با v درجه آزادی است.

قضیه 8: اگر Z_n, \dots, Z_2, Z_1 نمونه های تصادفی حاصل از توزیع نرمال استاندارد $N(0,1)$ باشند در

این صورت:

1- \bar{Z} دارای توزیع نرمال $N\left(0, \frac{1}{n}\right)$ است.

2- \bar{Z} و $\sum_{i=1}^n (Z_i - \bar{Z})^2$ از هم مستقلند.

3- $\sum_{i=1}^n (Z_i - \bar{Z})^2$ دارای توزیع Chi Square با $n-1$ درجه آزادی است.

نتیجه: اگر x_n, \dots, x_2, x_1 نمونه تصادفی حاصل از جامعه $N(\mu, \sigma^2)$ باشد در این صورت متغیر

تصادفی $\chi^2 = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$ دارای توزیع مربع کای با $n-1$ درجه آزادی است.

قضیه 9: اگر x_n, \dots, x_2, x_1 یک نمونه تصادفی به اندازه n از یک جامعه نرمال با میانگین μ

و واریانس σ^2 باشد در این صورت $\frac{(n-1)S^2}{\sigma^2}$ یک توزیع مربع کای با $n-1$ درجه آزادی است.

اثبات: در ابتدا بدون اثبات می پذیریم که هرگاه x_n, \dots, x_1 یک نمونه تصادفی n تایی از یک جامعه

نرمال با میانگین μ و واریانس σ^2 باشند آنگاه توزیع های S^2, \bar{X} که به ترتیب میانگین و واریانس

نمونه است از یکدیگر مستقل هستند. حال داریم:

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

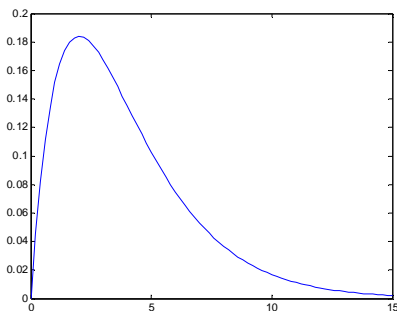
$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} + \frac{n(\bar{x} - \mu)^2}{\sigma^2} - \frac{2(\bar{x} - \mu) \sum_{i=1}^n (x_i - \mu)}{\sigma^2}$$

$$\frac{(n-1)s^2}{\sigma^2} = \underbrace{\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2}_{\text{توزیع مربع کای با } n \text{ درجه آزادی}} - \underbrace{\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2}_{\text{توزیع مربع کای با یک درجه آزادی}} = \chi_{(n-1)}^2$$

تذکر: در یک نمونه تصادفی، احتمال اینکه مقدار عددی یک متغیر تصادفی مربع کای بزرگتر از عدد مشخصی شود عبارتست از سطح زیر چگالی مربع کای دم سمت راست آن عدد مشخص و این عدد معمولاً توسط χ_{α}^2 نمایش داده می‌شود بطوری که سطح زیر چگالی توزیع مربع کای واقع در سمت راست آن عبارتست از α .

مثال توزیع مربع کای با چهار درجه آزادی.



قضیه 10: اگر x_1, \dots, x_n یک نمونه تصادفی از جامعه‌ای نامشخص $f_x(x)$ باشد آنگاه: $E[S^2] = \sigma^2$

اثبات:

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2\right] = \frac{1}{n-1} E\left[\sum_{i=1}^n (x_i - \mu) - (\bar{x} - \mu)\right]^2 = \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2 - 2 \sum_{i=1}^n (x_i - \mu)(\bar{x} - \mu)\right] = \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n (x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2(\bar{x} - \mu) \sum_{i=1}^n (x_i - \mu)\right] = \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n (x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2n(\bar{x} - \mu)^2\right] = \frac{1}{n-1} E\left[\sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2\right] = \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n E(x_i - \mu)^2\right] - nE[(\bar{x} - \mu)^2] = \frac{1}{n-1} \left[\sum_{i=1}^n \sigma^2 - n\text{Var}(\bar{x})\right] = \\ &= \frac{1}{n-1} \left(n\sigma^2 - n \frac{\sigma^2}{n}\right) = \sigma^2 \end{aligned}$$

قضیه 11: اگر x_1, \dots, x_n یک نمونه تصادفی از یک جامعه نرمال $N(\mu, \sigma^2)$ باشد آنگاه:

$$\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

اثبات:

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

$$\text{Var}\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$$

$$\frac{(n-1)^2}{\sigma^4} \text{Var}(S^2) = 2(n-1) \rightarrow \text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

3- توزیع t_student

توزیع t حاصل از تقسیم دو متغیر تصادفی است که صورت کسر متغیر تصادفی نرمال استاندارد و مخرج کسر جذر یک متغیر تصادفی دارای توزیع مربع کای تقسیم بر درجه آزادی آن است یعنی:

$$t_{(v)} = \frac{Z}{\sqrt{\frac{\chi^2_{(v)}}{v}}}$$

با استفاده از تکنیک تبدیل متغیر در تابع توزیع تابعی از یک متغیر تصادفی، تابع چگالی احتمال متغیر تصادفی t-student بشرح ذیل خواهد بود.

$$f_{t(v)} = \frac{1}{\sqrt{\pi v}} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \quad ; \quad -\infty < x < +\infty$$

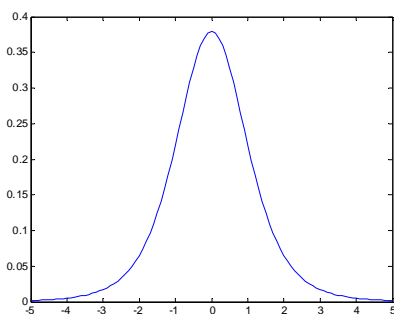
$$E[x] = 0$$

$$\text{Var}(x) = \frac{v}{v-2}; v > 2$$

شکل تابع چگالی متغیر تصادفی t، مانند توزیع نرمال با میانگین صفر حول نقطه صفر متقارن است.

تفاوت توزیع t با توزیع نرمال در آنست که توزیع t کمی پهن تر از توزیع نرمال است. در توزیع t با افزایش درجه آزادی، واریانس کمتر می شود.

مثال: توزیع t با 5 درجه آزادی



نکته: در یک نمونه تصادفی، احتمال اینکه مقدار عددی یک متغیر تصادفی t بزرگتر از عدد مشخصی باشد عبارتست از سطح زیر چگالی t در سمت راست آن عدد مشخص و این عدد معمولاً

با t_α نمایش داده می‌شود بطوری که سطح زیر چگالی t واقع در سمت راست t_α عبارتست از α .

$$P(T \geq t_{\alpha, v}) = \alpha$$

$$P(T_{(10)} \geq 1.812) = P(T \geq t_{(10), \%5}) = \%5$$

$$P(T_{(5)} \geq 2.015) = P(T \geq t_{(5), \%5}) = \%5$$

قضیه 12: در توزیع t هر چه v بزرگتر باشد، توزیع t به سمت توزیع نرمال استاندارد میل خواهد

کرد و وقتی که $v = +\infty$ باشد توزیع t همان توزیع نرمال استاندارد است. یعنی:

$$\lim_{v \rightarrow \infty} t_{(v)} = Z$$

اهمیت توزیع t

در حالتیکه σ معلوم باشد بر اساس قضیه حد مرکزی، متغیر تصادفی $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ دارای توزیع نرمال

استاندارد است اما در عمل σ مجهول است لذا برای تصمیم‌گیری لازم است توزیع متغیر تصادفی

را داشته باشیم که این مهم بر اساس توزیع t و قضیه ذیل برآورده می‌شود.

قضیه 13: اگر x_n, \dots, x_2, x_1 نمونه تصادفی حاصل از جامعه $N(\mu, \sigma^2)$ و \bar{X} و S بترتیب

میانگین و انحراف معیار نمونه مزبور باشد در این صورت متغیر تصادفی $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ دارای

توزیع t با $n-1$ درجه آزادی است.

$$t_{v=(n-1)} = \frac{Z}{\sqrt{\frac{\chi_{(n-1)}^2}{n-1}}}$$

$$t_{v=(n-1)} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

لذا متغیر تصادفی $\frac{\bar{X} - \mu}{\delta/\sqrt{n}}$ دارای توزیع t با $(n-1)$ درجه آزادی است.

نکته: اگر از یک جامعه نرمال $X \sim N(\mu, \sigma)$ یک نمونه n_1 تایی بگیریم و \bar{X} را بر اساس این نمونه

n_1 تایی تعریف کنیم و پس یک نمونه n_2 تایی مستقل از n_1 گرفته و S^2 را بر اساس آن تعریف

کنیم داریم آنگاه:

$$\frac{\bar{X} - \mu}{S/\sqrt{n_1}} \sim t_{(n_2-1)}$$

$$t_{v=(n-1)} = \frac{Z}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1}}}$$

$$t_{v=(n_1-1)} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n_1}}}{\sqrt{\frac{(n_2-1)\sigma^2}{\sigma^2}/(n_2-1)}} = \frac{\bar{X} - \mu}{\frac{S}{\sigma}} = \frac{\bar{X} - \mu}{S/\sqrt{n_1}}$$

نکته: اگر یک نمونه n_1 از جامعه $X \sim N(\mu_1, \sigma_1)$ گرفته و \bar{X} را از روی آن تعریف کنیم و سپس یک نمونه n_2 از جامعه $Y \sim N(\mu_2, \sigma_2)$ مستقل از X گرفته و S_y^2 را بر اساس آن تعریف کنیم داریم:

$$\frac{\sigma_2}{\sigma_1} \cdot \frac{\bar{X} - \mu_1}{S_y/\sqrt{n_1}} \sim t_{(n_2-1)}$$

نکته: توزیع t با یک درجه آزادی یک توزیع کوشی است. توزیع کوشی دارای میانگین نیست همچنین این توزیع دارای واریانس نیست.

$$f_{t(0)} = \frac{1}{\pi(1+x^2)} \quad ; \quad -\infty < x < +\infty$$

4- توزیع تفاوت بین دو میانگین نمونه

فرض کنید X_1, X_2, \dots, X_{n_x} یک نمونه تصادفی متشکل از n_x متغیر تصادفی مستقل با توزیع نرمال هر یک با میانگین μ_x و واریانس σ^2 باشد و فرض کنید Y_1, Y_2, \dots, Y_{n_y} یک نمونه تصادفی متشکل از n_y متغیر تصادفی مستقل با توزیع نرمال هر یک با میانگین μ_y و واریانس σ^2 باشد. همچنین فرض کنید که تمام X ها، Y ها مستقل باشند در این صورت:

الف- اگر σ معلوم باشد:

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma^2}{n_x}\right)$$

$$\bar{Y} \sim N\left(\mu_y, \frac{\sigma^2}{n_y}\right)$$

$$\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y, \frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}\right)$$

آنگاه متغیر تصادفی $Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$ دارای توزیع نرمال استاندارد است

ب- اگر σ مجهول باشد:

$$\frac{(n_x - 1)S_x^2}{\sigma^2} \sim \chi^2_{(n_x - 1)}$$

$$\frac{(n_y - 1)S_y^2}{\sigma^2} \sim \chi^2_{(n_y - 1)}$$

آنگاه با استفاده از خاصیت جمع در توزیع مربع کای، متغیر تصادفی $\frac{(n_x - 1)S_x^2}{\sigma^2} + \frac{(n_y - 1)S_y^2}{\sigma^2}$ دارای

توزیع مربع کای با $n_x + n_y - 2$ درجه آزادی است. از طرفی توزیع آن از $\bar{X} - \bar{Y}$ مستقل است. بنا

به تعریف متغیر تصادفی t داریم:

$$\frac{\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}}{\sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{\sigma^2} / (n_x + n_y - 2)}} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

لذا متغیر تصادفی $\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$ دارای توزیع t با $n_x + n_y - 2$ درجه

آزادی است

مثال: نمونه‌های تصادفی مستقل به اندازه $n_1 = 30$ و $n_2 = 50$ از دو جامعه نرمال با میانگین‌های

$\mu_1 = 78$ و $\mu_2 = 75$ و واریانس‌های $\sigma_1^2 = 150$ و $\sigma_2^2 = 200$ اختیار شده‌اند. احتمال اینکه میانگین

نمونه اول از میانگین نمونه دوم حداقل 4.8 بیشتر باشد چقدر است؟

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim Z$$

$$P(\bar{X}_1 - \bar{X}_2 \geq 4.8) = P\left[\frac{(\bar{X}_1 - \bar{X}_2) - (78 - 75)}{\sqrt{\frac{150}{30} + \frac{200}{50}}} \geq \frac{4.8 - (78 - 75)}{\sqrt{\frac{150}{30} + \frac{200}{50}}} \right] = P(Z \geq 0.53)$$

مثال: میانگین نمرات تست هوش دانشجویان سال اول یک دانشکده 540 و انحراف معیار آن 50

است. دو نمونه تصادفی با حجم $n_1 = 32$ و $n_2 = 50$ انتخاب می‌کنیم احتمال اینکه تفاضل

میانگین نمرات این دو نمونه:

الف- بیش از 20 باشد چقدر است؟

$$P(\bar{X}_1 - \bar{X}_2 > 20) = P\left[\frac{\bar{X}_1 - \bar{X}_2}{50\sqrt{\frac{1}{32} + \frac{1}{50}}} > \frac{20}{50\sqrt{\frac{1}{32} + \frac{1}{50}}} \right] =$$

$$P(Z > 1.76) = 1 - \phi(1.76)$$

ب- بین 5 و 10 باشد چقدر است؟

$$P(5 < \bar{X}_1 - \bar{X}_2 < 10) = P\left[\frac{5}{50\sqrt{\frac{1}{32} + \frac{1}{50}}} < Z < \frac{10}{50\sqrt{\frac{1}{32} + \frac{1}{50}}} \right] = P(0.44 < Z < 0.88)$$

نکته: نتایج بدست آمده از توزیع نمونه $\bar{X}_1 - \bar{X}_2$ ، نیز برای جمعیت‌های محدود وقتیکه نمونه گیری به روش بدون جایگذاری انجام می پذیرد صادق باشد، مشروط بر آنکه اندازه جمعیت ها یعنی N_1 و N_2 به ترتیب در مقابل اندازه نمونه یعنی n_1 و n_2 بزرگ باشد. بهر حال اگر جمعیت ها کوچک باشند و نمونه گیری به روش بدون جایگذاری انجام پذیرد در آنصورت ما باید $\sigma_{\bar{x}_1}$ و $\sigma_{\bar{x}_2}$ را بتوسط رابطه زیر محاسبه نمائیم:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

5- توزیع F

اگر متغیرهای تصادفی مستقل χ_1^2 و χ_2^2 دارای توزیع های مربع کای با درجات آزادی به ترتیب v_1 و v_2 باشند متغیر تصادفی:

$$F = \frac{\chi_1^2 / v_1}{\chi_2^2 / v_2}$$

توزیع F با v_1 و v_2 درجه آزادی دارد. بر این اساس تابع چگالی متغیر تصادفی F به صورت ذیل است:

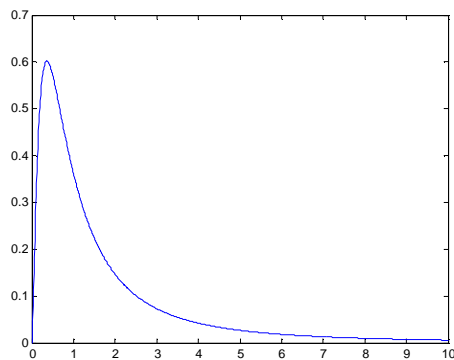
$$f_{F_{v_1, v_2}}(x) = \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} (v_1)^{\frac{v_1}{2}} (v_2)^{\frac{v_2}{2}} \frac{x^{\left(\frac{v_1-1}{2}\right)}}{(v_2 + v_1 x)^{\frac{v_1+v_2}{2}}} ; x > 0$$

که در v_1 و v_2 بترتیب درجه آزادی صورت و مخرج است. شکل تابع چگالی متغیر تصادفی F مانند توزیع مربع کای است. اهمیت توزیع F در آنست که برای مقایسه واریانسهای دو جامعه بکار می‌رود. میانگین و واریانس توزیع F با درجات آزادی v_1 و v_2 به ترتیب برابر است با:

$$E[x] = \frac{v_2}{v_2 - 2} ; v_2 > 2$$

$$\text{Var}[x] = \frac{v_2^2(2v_2 + 2v_1 - 4)}{v_1(v_2 - 2)^2(v_2 - 4)} ; v_2 > 4$$

مثال : توزیع F با 5 درجه آزادی در صورت و 3 درجه آزادی در مخرج



نکته: در یک نمونه تصادفی، احتمال اینکه مقدار عددی یک متغیر تصادفی F بزرگتر از عدد مشخصی باشد عبارتست از سطح زیر چگالی F در سمت راست آن عدد مشخص و این عدد معمولاً با $F_{v_1, v_2, \alpha}$ نمایش داده می‌شود بطوری که سطح زیر چگالی F واقع در سمت راست آن عبارتست از α یعنی $P(F > F_{\alpha; v_1, v_2}) = \alpha$

قضیه 14: اگر X_1, \dots, X_{m+1} یک نمونه تصادفی به حجم $m+1$ از یک توزیع نرمال $N(\mu_x, \sigma_x^2)$ باشد، همچنین اگر Y_1, \dots, Y_{n+1} یک نمونه تصادفی به حجم $n+1$ از یک توزیع نرمال $N(\mu_y, \sigma_y^2)$ باشد و نمونه‌های تصادفی ناهمبسته باشند در این صورت:

$$\frac{1}{\sigma_x^2} \sum_{i=1}^{m+1} (X_i - \bar{X})^2 \sim \chi^2(m)$$

$$\frac{1}{\sigma_y^2} \sum_{i=1}^{n+1} (Y_i - \bar{Y})^2 \sim \chi^2(n)$$

$$\frac{\sum_{i=1}^{m+1} (X_i - \bar{X})}{\sum_{i=1}^{n+1} (Y_i - \bar{Y})} \sim F_{m, n}$$

قضیه 15: (توزیع نسبت دو واریانس) فرض کنید X_1, X_2, \dots, X_{n_x} یک نمونه تصادفی متشکل از n_x متغیر تصادفی با توزیع نرمال هر یک با میانگین μ_x و واریانس σ_x^2 باشد. همچنین فرض کنید Y_1, Y_2, \dots, Y_{n_y} یک نمونه تصادفی متشکل از n_y متغیر تصادفی با توزیع نرمال هر یک با میانگین μ_y و واریانس σ_y^2 باشد. همچنین فرض کنید تمام X ها و Y ها مستقل باشند لذا:

$$\frac{(n_x - 1)S_x^2}{\sigma_x^2} \sim \chi^2_{(n_x - 1)}$$

$$\frac{(n_y - 1)S_y^2}{\sigma_y^2} \sim \chi^2_{(n_y - 1)}$$

چون X ها و Y ها مستقل اند پس این دو متغیر تصادفی مربع کای مستقل از هم هستند.

اثبات: طبق تعریف متغیر تصادفی F داریم:

$$F_{v_1, v_2} = \frac{\chi^2_{(v_1)}/v_1}{\chi^2_{(v_2)}/v_2}$$

$$\frac{(n_x - 1)S_x^2 / (n_x - 1)}{\sigma_x^2} = F_{n_x - 1, n_y - 1}$$

$$\frac{(n_y - 1)S_y^2 / (n_y - 1)}{\sigma_y^2}$$

لذا متغیر تصادفی $\frac{S_x^2}{S_y^2} \cdot \frac{\sigma_y^2}{\sigma_x^2}$ دارای توزیع F با درجات آزادی $n_x - 1$ و $n_y - 1$ است.

چند نکته مهم

1- اگر x دارای توزیع F_{v_1, v_2} باشد آنگاه متغیر $Y = \frac{1}{X}$ نیز دارای توزیع F_{v_2, v_1} است.

2- مربع توزیع $t_{(v)}$ دارای توزیع F با درجات آزادی $v, 1$ است.

$$t_{(v)} = \frac{Z}{\sqrt{\frac{\chi^2_{(v)}}{v}}} \rightarrow t_{(v)}^2 = \frac{Z^2 / 1}{\chi^2_{(v)} / v}$$

$$F_{1-\alpha; v_2, v_1} \frac{1}{F_{\alpha; v_1, v_2}} \text{ یا } F_{\alpha; v_1, v_2} = \frac{1}{F_{1-\alpha; v_2, v_1}} \quad -3$$

4- اگر X_1 و X_2 یک نمونه تصادفی از پخش $f(x) = e^{-x}$; $x > 0$ باشد آنگاه $Z = \frac{X_1}{X_2}$ دارای

پخش $F_{2,2}$ است. توزیع F با درجات آزادی 2 و 2 به صورت زیر است:

$$f_z(z) = \frac{1}{(1+z)^2}$$

6- توزیع میانگین جامعه‌های متناهی

اگر آزمایش متشکل از انتخاب یک مقدار یا بیشتر از مجموعه‌ای متناهی از اعداد $\{c_1, \dots, c_n\}$ باشد این مجموعه را جامعه‌ای متناهی با اندازه N می‌نامند. اگر انتخاب بدون جایگذاری باشد و x_1 اولین و x_n امین عددی باشد که استخراج می‌شوند، این متغیرهای تصادفی، نمونه‌ای تصادفی به اندازه n این جامعه متناهی را تشکیل می‌دهند مشروط بر آنکه توزیع احتمال توأم آنها به ازای هر n تایی مرتب مقادیر انتخاب شده از مجموعه $\{c_1, \dots, c_n\}$ به صورت زیر باشد:

$$f(x_1, \dots, x_n) = \frac{1}{N(N-1)\dots(N-n+1)}$$

$$f(x_i) = \frac{1}{N}, x_i = c_1, \dots, c_N, i = 1, \dots, n$$

میانگین و واریانس آنرا میانگین و واریانس جامعه متناهی می‌نامیم. یعنی میانگین و واریانس جامعه متناهی $\{c_1, \dots, c_N\}$ عبارتست از:

$$\mu = \frac{\sum_{i=1}^N c_i}{N}, \sigma^2 = \frac{\sum_{i=1}^N (c_i - \mu)^2}{N}$$

توزیع حاشیه‌ای توأم هر دو تا از متغیرهای تصادفی x_1, \dots, x_n برای هر زوج مرتب از مقادیر جامعه متناهی است و برابر است با:

$$g(x_i, x_j) = \frac{1}{N(N-1)}$$

و کوواریانس آنها عبارتست از:

$$\text{Cov}(x_i, x_j) = \frac{\sigma^2}{N-1}$$

اگر \bar{X} میانگین یک نمونه تصادفی به اندازه n از جامعه‌ای متناهی به اندازه N با میانگین μ و واریانس σ^2 باشد آنگاه:

$$E(\bar{x}) = \mu, \text{Var}(\bar{x}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

قضیه 16: اگر تمام نمونه های تصادفی n تایی ممکنه با روش بدون جایگذاری از یک جمعیت محدود N تایی با حد متوسط μ و انحراف معیار σ بیرون کشیده شود در آنصورت توزیع نمونه ای \bar{X} به طور تقریبی یک توزیع نرمال با حد متوسط و انحراف معیار زیر خواهد بود.

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

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

مثال:

جمعیت 1 و 1 و 3 و 4 و 5 و 6 و 6 و 6 و 7 داده شده است. نمونه تصادفی 36 تایی به روش جایگزین انتخاب شده است. احتمال اینکه میانگین نمونه بیشتر از $3/8$ و کوچکتر از $4/5$ باشد چقدر است؟

توزیع احتمال جمعیت مربوط به صورت زیر است:

x	1	3	4	5	6	7
P(x)	0.3	0.1	0.1	0.1	0.3	0.1

$$E(x) = \mu = 4$$

$$\text{Var}(x) = \sigma^2 = 5$$

$$\bar{x} \sim N\left(4, \frac{5}{36}\right)$$

$$P(3.8 < \bar{x} < 4.5) = P(-0.405 < z < 1.216) = 0.5453$$

نکته: برای N هایی که در مقابل تعداد نمونه n عدد بزرگی هستند، ضریب $\frac{N-n}{N-1}$ به سمت 1 میل می کند.

قضیه 17: اگر نمونه ای تصادفی به اندازه n از جامعه ای متناهی که متشکل از N عدد صحیح مثبت است انتخاب شود آنگاه:

$$E(\bar{x}) = \frac{N+1}{2}$$

$$\text{Var}(\bar{x}) = \frac{(N+1)(N-n)}{12n}$$

$$E(y) = \frac{n(N+1)}{2}$$

$$\text{Var}(y) = \frac{n(N+1)(N-n)}{12}$$

$$\mu = E(\bar{x}) = \sum_{i=1}^N x_i \cdot f(x_i) = \sum_{i=1}^N x_i \cdot \frac{1}{N} = \frac{1}{N} (x_1 + x_2 + \dots + x_N) = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \frac{N+1}{2}$$

$$E(x^2) - (E(x))^2 = \sum_{i=1}^N x_i^2 \cdot f(x_i) - \frac{(N+1)^2}{4} = \frac{1}{N} \cdot \frac{N(2N+1)(N+1)}{6} - \frac{(N+1)^2}{4} = \frac{N^2-1}{12}$$

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right) = \frac{N^2-1}{12n} \cdot \frac{N-n}{N-1} = \frac{(N+1)(N-n)}{12n}$$

7- آماره‌های ترتیبی

فرض کنید از یک جامعه نامتناهی پیوسته یک نمونه تصادفی n تایی بشرح X_1, \dots, X_n انتخاب کرده‌ایم. اگر کوچکترین مقدار x ها را Y_1 ، بزرگترین مقدار بعد از آن را Y_2 و به همین ترتیب بزرگترین آنها را Y_n بنامیم در این صورت $Y_1 < Y_2 < \dots < Y_n$ را آماره ترتیبی می‌نامیم. قضیه 18: برای نمونه تصادفی به اندازه n از جامعه‌ای نامتناهی که دارای تابع چگالی $f(x)$ است.

چگالی r امین آماره ترتیبی y_r ، به ازای $-\infty < y_r < +\infty$ عبارتست از:

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!} \left[\int_{-\infty}^{y_r} f(x) dx \right]^{r-1} f(y_r) \left[\int_{y_r}^{\infty} f(x) dx \right]^{n-1}$$

اثبات: فرض کنید که محور اعداد حقیقی را به سه بازه تقسیم کرده‌ایم یکی از $-\infty$ تا y_r ، دومی از y_r تا y_r+h (که در آن h ثابت است) و سومی از y_r+h تا $+\infty$. در اینصورت اگر چگالی جامعه‌ای که از آن نمونه برگرفته شده است $f(x)$ باشد، احتمال اینکه $r-1$ تا از نمونه‌ها در اولین باز قرار بگیرد، یکی در بازه دوم و $n-r$ تا در بازه سوم قرار بگیرند طبق فرمول توزیع چند جمله‌ای عبارتست از

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!} \left[\int_{-\infty}^{y_r} f(x) dx \right]^{r-1} \left[\int_{y_r}^{y_r+h} f(x) dx \right]^{r-1} \left[\int_{y_r}^{\infty} f(x) dx \right]^{n-1}$$

با استفاده از قضیه میانگین در حسابان داریم:

$$\left[\int_{y_r}^{y_r+h} f(x) dx \right] = f(\xi) \cdot h \quad y_r \leq \xi \leq y_r + h$$

و در زمانی که $h \rightarrow 0$ خواهیم داشت

$$\lim_{h \rightarrow 0} \int_{y_r}^{y_r+h} f(x) dx = f(y_r)$$

و سرانجام

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!} \left[\int_{-\infty}^{y_r} f(x) dx \right]^{r-1} f(y_r) \left[\int_{y_r}^{\infty} f(x) dx \right]^{n-1}$$

بر این اساس توزیع اولین، آخرین و میانه آماره ترتیبی عبارتست از

الف - توزیع اولین آماره ترتیبی

$$g_1(y_1) = n f(y_1) \left[\int_{y_1}^{\infty} f(x) dx \right]^{n-1} \quad -\infty \leq y_1 \leq +\infty$$

ب- توزیع آخرین آماره ترتیبی

$$g_n(y_n) = n f(y_n) \left[\int_{-\infty}^{y_n} f(x) dx \right]^{n-1} \quad -\infty \leq y_n \leq +\infty$$

ج - در نمونه‌ای تصادفی به اندازه $n=2m+1$ توزیع $\tilde{x} = y_{m+1}$ میانه آماره‌های ترتیبی عبارتست از

$$h(\tilde{x}) = \frac{(2m+1)!}{m!m!} \left[\int_{-\infty}^{\tilde{x}} f(x) dx \right]^m f(\tilde{x}) \left[\int_{\tilde{x}}^{\infty} f(x) dx \right]^m \quad -\infty \leq \tilde{x} \leq +\infty$$

د - در نمونه‌ای تصادفی به اندازه $n=2m$ توزیع \tilde{x} میانه آماره‌های ترتیبی عبارتست از

$$h(\tilde{x}) = \frac{g_m(y_m) + g_{m+1}(y_{m+1})}{2}$$

ه- چگالی توام $g_{r,j}(y_1, y_j)$ عبارتست از

$$g_{r,j}(y_r, y_j) = \frac{n!}{(r-1)!(j-r-1)!(n-j)!} \left[\int_{-\infty}^{y_r} f(x) dx \right]^{r-1} f(y_r) \left[\int_{y_r}^{y_j} f(x) dx \right]^{j-1} f(y_j) \left[\int_{y_j}^{\infty} f(x) dx \right]^{n-j}$$

مثال: فرض کنید X_1, \dots, X_n یک نمونه تصادفی n تایی از یک جامعه‌ای نمایی با $\lambda = 1$ باشد. تابع چگالی اولین و آخرین آماره ترتیبی را بدست آورید.

$$g_1(y_1) = n e^{-y_1} \left[\int_{y_1}^{\infty} e^{-x} dx \right]^{n-1} = n e^{-y_1} (e^{-y_1})^{n-1} = n e^{-ny_1}$$

$$g_n(y_n) = n e^{-y_n} \left[\int_{-\infty}^{y_n} e^{-x} dx \right]^{n-1} = n e^{-y_n} (1 - e^{-y_n})^{n-1}$$