

فهرست

۱.....	۱- تعاریف
۱.....	۱-۱- تعریف جامعه (Population)
۱.....	۲-۱- تعریف پارامتر (Parameter)
۱.....	۳-۱- تعریف نمونه (Sample)
۲.....	۴-۱- اندازه نمونه (Sample Size)
۲.....	۵-۱- تعریف نمونه تصادفی (Random Sample)
۲.....	۶-۱- توزیع نمونه
۲.....	۷-۱- تعریف آماره
۳.....	۸-۱- میانگین نمونه
۳.....	۹-۱- گشتاور نمونه
۳.....	۱۰-۱- گشتاور ۲ ام حول میانگین
۴.....	۱۱-۱- واریانس نمونه
۴.....	۱۲-۱- انحراف معیار نمونه
۵.....	۱۳-۱- قضیه حد مرکزی
۵.....	۱۴-۱- توزیع Gamma
۶.....	۲- توزیع مرربع کای (Chi Square)
۱۰.....	۳- توزیع t_student
۱۳.....	۴- توزیع تفاوت بین دو میانگین نمونه
۱۶.....	۵- توزیع F
۱۹.....	۶- توزیع میانگین جامعه های متناهی
۲۲.....	۷- آماره های ترتیبی

۱- تعاریف

۱-۱- 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.

۱-۲- تعریف پارامتر (Parameter)

مقدار ثابتی مربوط به جامعه آماری است که معمولاً مورد علاقه بررسی کننده واقع می‌شود. مثلاً در مورد لامپ‌ها، متوسط طول عمر لامپ‌ها.

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

X	.	۱	۲	
P	$\frac{\theta}{2}$	$1-\theta$	$\frac{\theta}{2}$	مدل آمار

X	.	۱	۲	مدل
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

۱-۳- تعریف نمونه (Sample)

در حالاتی که به جامعه دسترسی نداریم و یا به علت محدودیت در زمان و هزینه، بنای مطالعه و بررسی بر روی بخشی از جامعه به عنوان نمونه انجام می‌شود. 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: معمولاً 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.

۱-۴- اندازه نمونه (Sample Size)

تعداد افراد نمونه را حجم نمونه یا اندازه نمونه می‌نامند.

۱-۵- تعریف نمونه تصادفی (Random Sample)

متغیرهای تصادفی x_1, x_2, \dots, x_n تشکیل یک نمونه تصادفی به اندازه n می‌دهند اگر:
۱- هم توزیع باشند.
۲- از یکدیگر مستقل باشند.

۱-۶- 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 x_1, x_2, \dots, x_n جامعه $(f_x(x))$ باشد در این صورت توزیع نمونه تصادفی x_1, x_2, \dots, x_n عبارتست از:

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

۱-۷- تعریف آماره

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

بستگی ندارد) مثلاً در یک نمونه گیری 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.

۱- میانگین نمونه

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

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

۲- 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 میں 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.

۳- 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 ۱: فرض کنید x_1, x_2, \dots, x_n یک نمونه تصادفی از جامعه با تابع چگالی $f_x(x)$ باشد و

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

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

اثبات:

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n\mu = \mu \\ \text{Var}[\bar{X}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

۱۱-۱- واریانس نمونه

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

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

واریانس نمونه نامیده می شود.

۱۲-۱- انحراف معیار نمونه

انحراف معیار n مشاهده x_1, x_2, \dots, x_n را به عنوان ریشه دوم واریانسها یشان تعریف می کنیم:

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

قضیه ۲: اگر تمام نمونه های تصادفی به حجم n از یک جامعه نرمال با میانگین μ و واریانس σ^2

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

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

قضیه ۳: اگر تمام نمونه های 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.

۱۳-۱- 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.

۱۴-۱- توزیع Gamma

اگر متغیر تصادفی گاما دارای تابع چگالی ذیل است:

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

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

۲- توزیع مربع کای (Chi Square)

توزیع مربع کای حالت خاصی از توزیع گاماست که در آن $\lambda = \frac{v}{2}$ است و بر این اساس تابع

چگالی احتمال متغیر تصادفی مربع کای بشرح ذیل است.

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

که Γ معروف تابع گاما و معروف درجه آزادی است و

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

قضیه ۴: اگر متغیر تصادفی Z_1 دارای توزیع نرمال استاندارد باشد، آنگاه متغیر تصادفی Z_1^2 دارای

توزیع مربع کای با ۱ درجه آزادی خواهد بود. این قضیه در واقع بیانگر اهمیت توزیع مربع کای است.

قضیه ۵: اگر $\chi^2_k, \chi^2_1, \chi^2_2, \dots, \chi^2_n$ متغیرهای تصادفی مستقل هر یک دارای توزیع مربع کای با درجات

آزادی به ترتیب v_1, v_2, \dots, v_n باشند، آنگاه متغیر تصادفی $\chi^2 = \chi^2_1 + \chi^2_2 + \dots + \chi^2_n$ 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 ۶: اگر متغیرهای تصادفی مستقل Z_1, Z_2, \dots, Z_v 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_i ‌ها یک توزیع مربع کای ندارد.

قضیه ۷: اگر متغیرهای تصادفی x_1, x_2, \dots, x_v دارای توزیع نرمال مستقل با میانگین‌های

$\mu_1, \mu_2, \dots, \mu_v$ و واریانس‌های $\sigma_1^2, \sigma_2^2, \dots, \sigma_v^2$ باشند در این صورت متغیر تصادفی $Z = \frac{x_i - \mu_i}{\sigma_i}$

برای $i = 1, 2, \dots, v$ دارای توزیع نرمال استاندارد می‌باشد 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 ۸: اگر Z_1, Z_2, \dots, Z_n نمونه‌های تصادفی حاصل از توزیع نرمال استاندارد $N(0, 1)$ باشند در

این صورت:

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

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

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

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

$$4-\chi^2 \text{ دارای توزیع مربع کای با } n-1 \text{ درجه آزادی است.} \quad \sum_{i=1}^n \left(\frac{x_i - \bar{X}}{\sigma} \right)^2$$

قضیه ۹: اگر x_1, x_2, \dots, x_n 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 می‌پذیریم که هرگاه x_1, x_2, \dots, x_n یک نمونه تصادفی n تایی از یک جامعه

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

نمونه است از یکدیگر مستقل هستند. حال داریم:

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

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

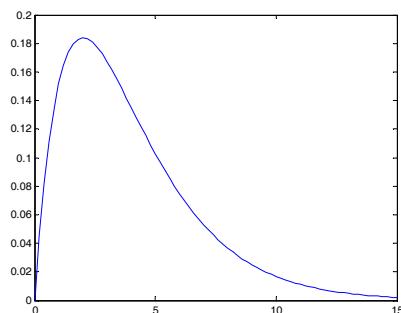
$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} - \frac{2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu)}{\sigma^2}$$

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 - \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 = \chi_{(n-1)}^2$$

توزیع مرربع کای با درجه آزادی ۱
یک درجه آزادی

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

مثال توزیع مرربع کای با چهار درجه آزادی.



قضیه ۱۰: اگر x_1, x_2, \dots, x_n یک نمونه تصادفی از جامعه‌ای نامشخص $f_x(x)$ باشد آنگاه:

اثبات:

$$\begin{aligned}
 E(S^r) &= \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^r \right] = \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu) - (\bar{X} - \mu) \right]^r = \\
 &= \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu)^r - (\bar{X} - \mu)^r - r \sum_{i=1}^n (X_i - \mu)(\bar{X} - \mu) \right] = \\
 &= \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu)^r + n(\bar{X} - \mu)^r - r(\bar{X} - \mu) \underbrace{\sum_{i=1}^n (X_i - \mu)}_{n(\bar{X} - \mu)} \right] = \\
 &= \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu)^r + n(\bar{X} - \mu)^r - rn(X_i - \mu)^r \right] = \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu)^r - n(\bar{X} - \mu)^r \right] = \\
 &= \frac{1}{n-1} \left[\sum_{i=1}^n E[(X_i - \mu)^r] - nE[(\bar{X} - \mu)^r] \right] = \frac{1}{n-1} \left[\sum_{i=1}^n \sigma^r - n\text{Var}(\bar{X}) \right] = \\
 &= \frac{1}{n-1} \left(n\sigma^r - n \frac{\sigma^r}{n} \right) = \sigma^r
 \end{aligned}$$

قضیه ۱۱: اگر x_1, x_2, \dots, x_n یک نمونه تصادفی از یک جامعه نرمال $N(\mu, \sigma^2)$ باشد آنگاه:

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

اثبات:

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

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

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

۳- 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:

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

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.

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

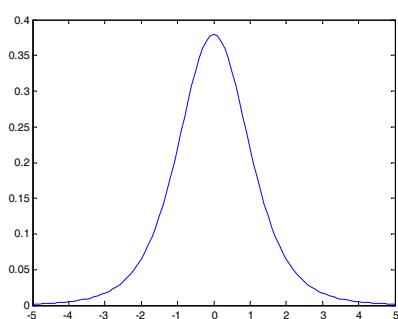
$; -\infty < x < +\infty$

$$E[x] = 0$$

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

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 تراز توزیع نرمال است. در توزیع t با افزایش درجه آزادی، واریانس کمتر می شود.

مثال: توزیع t با ۵ 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 .

$$\begin{aligned} P(T \geq t_{\alpha,v}) &= \alpha \\ P(T_{(1)} \geq 1.812) &= P(T \geq t_{(1),0.05}) = 0.05 \\ P(T_{(5)} \geq 2.015) &= P(T \geq t_{(5),0.05}) = 0.05 \end{aligned}$$

قضیه ۱۲: در توزیع t هر چه v بزرگتر باشد، توزیع t به سمت توزیع نرمال استاندارد میل خواهد

کرد و وقتیکه $v = +\infty$ 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 مجھول است لذا برای تصمیم‌گیری لازم است توزیع متغیر تصادفی

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

قضیه ۱۳: اگر x_1, x_2, \dots, x_n نمونه تصادفی حاصل از جامعه $N(\mu, \sigma^2)$ و \bar{X} و S بترتیب

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

توزیع t با $n-1$ درجه آزادی است.

$$\begin{aligned} t_{v=(n-1)} &= \frac{Z}{\sqrt{\frac{\chi_{(n-1)}}{n-1}}} \\ t_{v=(n-1)} &= \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \end{aligned}$$

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

نکته: اگر از یک جامعه نرمال $X \sim N(\mu, \sigma^2)$ 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:

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

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

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

نکته: اگر یک نمونه n_1 از جامعه $X \sim N(\mu_1, \sigma_1^2)$ گرفته و \bar{X} را از روی آن تعریف کنیم و سپس یک نمونه n_2 از جامعه $Y \sim N(\mu_2, \sigma_2^2)$ 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 همچنانی این توزیع دارای واریانس نیست.

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

۴- توزیع تفاوت بین دو میانگین نمونه

فرض کنید X_1, X_2, \dots, X_{n_x} یک نمونه تصادفی متشكل 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

متشكل 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:

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

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 مجہول باشد:

$$\begin{aligned}\frac{(n_x - 1)S_x^2}{\sigma^2} &\sim \chi^2_{(n_x - 1)} \\ \frac{(n_y - 1)S_y^2}{\sigma^2} &\sim \chi^2_{(n_y - 1)}\end{aligned}$$

آنگاه با استفاده از خاصیت جمع در توزیع مربع کای، متغیر تصادفی $\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{\sigma^2}$ دارای توزیع مربع کای با ۲ درجه آزادی است.

از طرفی توزیع آن از $\bar{X} - \bar{Y}$ مستقل است. بنا

به تعریف متغیر تصادفی t داریم:

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

$$\text{لذا متغیر تصادفی } \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \text{ دارای توزیع } t \text{ با } n_x + n_y - 2 \text{ درجه آزادی است}$$

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 $\sigma_1^2 = 200$ و $\sigma_2^2 = 150$ اختیار شده‌اند. احتمال اینکه میانگین

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

$$\begin{aligned} & \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim Z \\ P(\bar{X}_1 - \bar{X}_2 \geq 4.8) &= P\left[\frac{(\bar{X}_1 - \bar{X}_2) - (78 - 75)}{\sqrt{\frac{150}{30} + \frac{200}{50}}} \geq \frac{4.8 - (78 - 75)}{\sqrt{\frac{150}{30} + \frac{200}{50}}}\right] = P(Z \geq 0.53) \end{aligned}$$

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 تفاصل 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 می‌پذیرد صادق باشد، مشروط بر آنکه اندازه جمعیت‌ها یعنی N_1 و N_2 به ترتیب در مقابل اندازه نمونه یعنی n_1 و n_2 بزرگ باشد. به حال اگر جمعیت‌ها 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:

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

F-توزیع

اگر متغیرهای تصادفی مستقل χ_1^2 و χ_2^2 دارای توزیع های مریع 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 اساستابع چگالی متغیر تصادفی F به صورت ذیل

است:

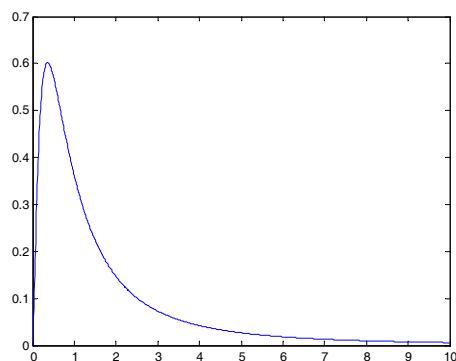
$$f_{F_{v_1,v_2}(x)} = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} (v_1)^{\frac{v_1}{2}} (v_2)^{\frac{v_2}{2}} \frac{x^{\left(\frac{v_1-1}{2}\right)}}{(v_2 + v_1 x)^{\frac{v_1+v_2}{2}}} ; \quad x > 0$$

که در v_1 و v_2 بترتیب درجه آزادی صورت و مخرج است. شکل تابع چگالی متغیر تصادفی F مانند توزیع مریع 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:

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

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

مثال : توزیع F با ۵ درجه آزادی در صورت و ۳ 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

$$P(F > F_{\alpha; v_1, v_2}) = \alpha$$

قضیه ۱۴: اگر X_1, X_2, \dots, X_m یک نمونه تصادفی به حجم $m+1$ از یک توزیع نرمال $N(\mu_x, \sigma_x^2)$ باشد، همچنین اگر Y_1, Y_2, \dots, Y_n 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:

$$\frac{\frac{1}{\sigma_x^2} \sum_{i=1}^{m+1} (X_i - \bar{X})^2}{\frac{1}{\sigma_y^2} \sum_{i=1}^{n+1} (Y_i - \bar{Y})^2} \sim \chi_{(m)}^2$$

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

قضیه ۱۵: (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:

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

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

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:

$$F_{v_1, v_2} = \frac{\chi_{(v_1)} / v_1}{\chi_{(v_2)} / v_2}$$

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

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

چند نکته مهم

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

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

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

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

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

پخش $F_{2,2}$ است. توزيع F با درجات آزادی ۲ و ۲ به صورت زیر است:

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

۶- 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, \dots, x_n اولین n 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:

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

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

میانگین و واریانس آنرا میانگین و واریانس جامعه متناهی می‌نامیم. یعنی میانگین و واریانس

جامعه متناهی $\{c_1, \dots, c_N\}$ عبارتست از :

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

توزیع حاشیه‌ای توام هر دوتا 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:

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

قضیه ۱۶: اگر تمام نمونه‌های 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 نمونه‌ای \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 ۳۶ تایی به روش جایگزین انتخاب شده است. نمونه تصادفی ۷ عدد و عوّعو ۵ و ۴ و ۳ داده شده است. احتمال اینکه میانگین نمونه بیشتر 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 ۱ میل

می‌کند.

قضیه ۱۷: اگر نمونه‌ای تصادفی به اندازه n از جامعه‌ای متناهی که متشکل از N عدد صحیح مثبت

است انتخاب شود آنگاه:

$$E(\bar{x}) = \frac{N-1}{\gamma}$$

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

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

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

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

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

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

۷- آماره‌های ترتیبی

فرض کنید از یک جامعه نامتناهی پیوسته یک نمونه تصادفی n تایی بشرح X_1, \dots, X_n انتخاب کردہ‌ایم. 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 ۱۸: 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:

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

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

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

با استفاده از قضیه میانگین در حسابان داریم:

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

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

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

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

الف - توزیع اولین آماره ترتیبی

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

ب - توزیع آخرین آماره ترتیبی

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

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, X_2, \dots, X_n یک نمونه تصادفی از یک جامعه‌ای نمایی با $\lambda = 1$ باشد.تابع چگالی اولین و آخرین آماره ترتیبی را بدست آورید.

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

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