Date:.................. Subject:..................

توزیع‌های نمونه‌ای

به دلیل محدودیت‌هایی از قبیل محدودیت نیروی انسانی، مشکلات اقتصادی و زمان، در شناخت جامعه آماری دچار مشکل می‌شویم و به مطالعه بخشی از جامعه به جای کل جامعه بسنده می‌کنیم.

انتخاب قسمتی از جامعه را نمونه‌گیری و قسمت انتخاب شده را نمونه (sample) می‌نامند.

نمونه تصادفی

برای برآورد (Estimate) عمر مفید نوع معینی ترانزیستور، پژوهشگری ده تا از این ترانزیستورها را انتخاب می‌کند و برای مدتی آنها را مورد آزمایش قرار می‌دهد و زمان از کار افتادن آنها را یادداشت می‌کند.

زمان از کار افتادن آنها، متغیر تصادفی است با توزیع نمایی با پارامتر θ.

← ویژگی مورد نظر در جامعه را با یک متغیر تصادفی توصیف می‌کنیم. این متغیر تصادفی دارای یک تابع احتمال معین است، $f_X(x)$. توزیع این متغیر را توزیع احتمال جامعه می‌نامیم.

اگر جامعه ما، نوزادان یک بیمارستان باشد، X را وزن نوزاد در نظر بگیریم. $X \sim N(\mu, \sigma^2)$

از این جامعه، n نوزاد را به تصادف و مستقل از یکدیگر انتخاب می‌کنیم

X_i = وزن نوزاد i ام $\quad i = 1, \dots, n$

در نتیجه داریم $X_1, \dots, X_n$ مستقل از یکدیگر و $X_i \sim N(\mu, \sigma^2)$.

Date:.................... Subject:....................

تعریف - مجموعه متغیرهای تصادفی $(X_1, \dots, X_n)$ را یک نمونه تصادفی n تایی از X نامند

اگر و تنها اگر X_i ها مستقل از یکدیگر و هم توزیع باشند.

i.i.d ⟶ independent and identically distribution

پارامترها و آماره: (parameter & statistic)

هر ویژگی یک جامعه را یک پارامتر آن جامعه می نامند. برای مثال، در مثال قبل، میانگین

وزن نوزادان، μ یک پارامتر جامعه است.

حال اگر از جامعه نمونه گیری کنیم و $(X_1, \dots, X_n)$ یک نمونه تصادفی از جامعه

باشد،

آماره تابعی از نمونه تصادفی است که به پارامتر مجهول جامعه بستگی ندارد.

بعنوان مثال میانگین نمونه یعنی $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$، یک آماره است.

مثال- اگر X متغیر تصادفی با تابع احتمال $f_X(x, \theta)$ باشد که در آن θ

پارامتر مجهول جامعه است، در اینصورت $T = \frac{1}{\theta} \sum_{i=1}^{n} X_i$ آماره نیست

زیرا به θ بستگی دارد!

Date:.............................. Subject: ....

مثال- جعبه‌ای شامل چهار کارت با شماره های ۱، ۲، ۳ و ۴ است. یک نمونه تصادفی به اندازه $n=2$ از این کارت ها انتخاب می کنیم. متغیر تصادفی X شماره کارت انتخابی است. X دارای توزیع یکنواخت گسسته است. تعداد نمونه‌های تصادفی (X_1, X_2)، 16

(X_1, X_2)

و هر یک دارای شانس $\frac{1}{16}$ برای انتخاب شدن می باشند:

$$\mu = \mathbb{E}[X] = \frac{1+2+3+4}{4} = 2.5 \longrightarrow$$ میانگین جامعه

بوضوح μ، پارامتر جامعه همواره ثابت است.

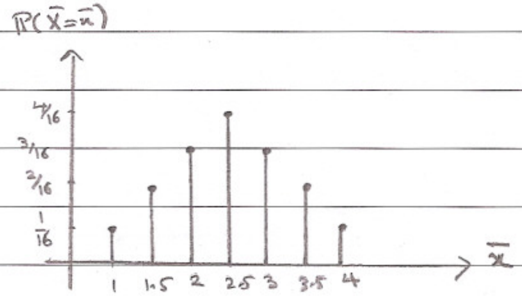
اما می توان از $\bar{X}$ برای برآورد کردن μ استفاده نمود. چون آماره ها، متغیرهای تصادفی هستند مقادیر آنها از یک نمونه به نمونه دیگر تغییر می کند. برای مثال

(X_1, X_2)	$\bar{x}$	$\mathbb{P}$
(1, 1)	1	1/16
(1, 2)	1.5	1/16
(1, 3)	2	1/16
⋮		

$\Rightarrow$

$\bar{x}$	1	1.5	2	2.5	3	3.5	4
$\mathbb{P}(\bar{X}=\bar{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

پس نمودار احتمال $\bar{X}$ بصورت زیر است:



اگرچه توزیع X یکنواخت گسسته بود ولی توزیع $\bar{X}$ به توزیع نرمال شبیه است.

← در این فصل هدف این است که درباره توزیع آماره های خاص کسب نماییم.

Date:........................ Subject:........................

۱- توزیع میانگین نمونه (با فرض معلوم بودن واریانس جامعه)

از این جامعه یک نمونه تصادفی $(X_1, \ldots, X_n)$ انتخاب می‌کنیم.

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

الف- اگر $X \sim N(\mu, \sigma^2)$:

در اینصورت $\bar{X}$ ترکیب خطی ای از متغیرهای تصادفی نرمال و مستقل از یکدیگر می‌باشد.

$$\bar{X} \sim N\left(\frac{1}{n} \sum_{i=1}^{n} \mu \,,\, \frac{1}{n^2} \sum \sigma^2\right)$$

$$\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

اگر قرار دهیم $Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ در اینصورت $Z \sim N(0,1)$.

ب- اگر جامعه دارای توزیع نرمال نباشد. در اینصورت توزیع $\bar{X}$ به توزیع جامعه بستگی دارد ولی با افزایش اندازه نمونه، $\bar{X}$ به توزیع نرمال نزدیک می‌شود.

قضیه حد مرکزی - فرض کنید X_1، X_2، ... دنباله‌ای از متغیرهای تصادفی هم توزیع با میانگین μ و واریانس (متناهی) σ^2 است بطوریکه هر تعداد متناهی از این متغیرهای تصادفی از یکدیگر مستقل هستند. در اینصورت $Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ زمانیکه $n \to \infty$ (تقریبا $n \geq 30$) دارای توزیع $N(0,1)$ می‌باشد.

Date:............................ Subject:...

مثال- یک دستگاه اتوماتیک فروشنده نوشابه لیوانی را طوری تنظیم کرده اند که مقدار نوشابه ای

که بعد از هر فشار دکمه آن خارج می شود، متغیر تصادفی با توزیع نرمال با میانگین 200

میلی لیتر و انحراف معیار 15 میلی لیتر است. مطلوب است احتمال اینکه میانگین

مقدار نوشابه ای که در یک نمونه تصادفی 36 تایی از آن خارج می شود، حداقل 204 میلی لیتر

باشد:

$$X \sim N(200, 15^2)$$

$$(X_1, \ldots, X_{36}) \Rightarrow Z = \frac{\bar{X} - 200}{\frac{15}{\sqrt{36}}} \sim N(0,1)$$

$$\Rightarrow \mathbb{P}(\bar{X} \geq 204) = \mathbb{P}\left(\frac{\bar{X} - 200}{2.5} \geq \frac{204 - 200}{2.5}\right) = \mathbb{P}(Z \geq 1.6)$$

$$= 1 - \mathbb{P}(Z < 1.6) = 1 - \Phi(1.6) = 1 - 0.9452 = 0.0548.$$

طبق جدول

مثال- در یک شهر نسبت واقعی خانواده هایی که مالک محل سکونت خود هستند، 0.7 است. اگر 84

خانواده در این شهر به تصادف مورد پرسش قرار بگیرند و جواب آنها به سؤال "آیا صاحب خانه هستید؟"

به عنوان مقادیر متغیر تصادفی مستقل با توزیع برنولی با پارامتر $\theta = 0.7$ تلقی شود، با چه احتمالی

می توان حکم کنیم که مقدار بدست آمده برای میانگین نمونه بین 0.64 و 0.76 قرار می گیرد؟

AZIN

Date:............................ Subject:............................

$X \sim B(1, 0.7) \longrightarrow$ می‌دانیم متغیر برنولی دارای میانگین $\mu = 0.7$ و واریانس $\sigma^2 = 0.7 \times 0.3$ می‌باشد.

$$n = 84 \geq 30 \checkmark \implies Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$P(0.64 \leq \bar{X} \leq 0.76) = P\left(\frac{0.64 - 0.7}{\sqrt{\frac{0.21}{84}}} \leq \frac{\bar{X} - 0.7}{\sqrt{\frac{0.21}{84}}} \leq \frac{0.76 - 0.7}{\sqrt{\frac{0.21}{84}}}\right)$$

$$= P(-1.2 \leq Z \leq 1.2) = 0.8849 - 0.1151 = 0.7698.$$

2_ توزیع واریانس نمونه

اگر از جامعه یک نمونه تصادفی $(X_1, \dots, X_n)$ انتخاب کنیم،

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

آماره ای برای برآورد σ^2 جامعه می‌باشد. هدف یافتن توزیع S^2 می‌باشد. بدین منظور توزیع مربع-کای (chi-square)، χ^2، را ابتدا معرفی می‌نماییم.

توزیع مربع-کای: اگر $Z \sim N(0,1)$ در اینصورت $Y = Z^2$ دارای توزیع مربع-کای با یک درجه آزادی است. اگر $(Z_1, \dots, Z_n)$ یک نمونه تصادفی از یک جامعه نرمال

استاندارد باشد، آنگاه $Y = Z_1^2 + \dots + Z_n^2$ دارای توزیع مربع-کای با درجه آزادی

n می‌باشد، اصطلاحاً آن را با نماد $Y \sim \chi^2_{(n)}$ نمایش می‌دهیم.

تابع چگالی احتمال $\chi^2_{(n)}$ به صورت زیر تعریف می‌گردد:

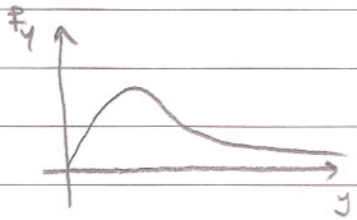
$$f_Y(y) = \frac{1}{2^{n/2}\,\Gamma(n/2)}\, y^{\frac{n}{2}-1}\, e^{-y/2} \qquad y>0$$

که در آن منظور از $\Gamma(t)=\int_0^\infty x^{t-1} e^{-x}\,dx$ است و می‌توان دید:

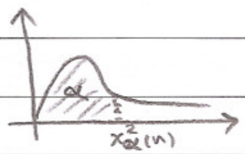
$\Gamma(t) = (t-1)\,\Gamma(t-1)$

$\Gamma(1/2) = \sqrt{\pi}$

$t \in \mathbb{N} \quad \Gamma(t) = (t-1)!$



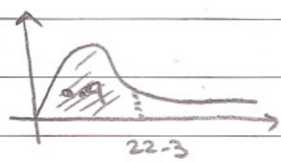
جدول مربع-کای:



اگر $Y \sim \chi^2_{(n)}$ در این صورت $P(Y \le \chi^2_\alpha(n)) = \alpha$

برای مثال:

$\chi^2_{0.9}(15) = 22.3 \quad \longleftrightarrow$



$\chi^2_{0.05}(28) = 16.9 \,,\quad \chi^2_{0.9}(24) = 33.2$

قضیه- اگر $(X_1, \dots, X_n)$ یک نمونه تصادفی از جامعه نرمال با میانگین μ و واریانس σ^2 باشد، آنگاه

الف- $\bar{X}$ و S^2 از یکدیگر مستقل هستند.

ب- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

Date: Subject:

مثال- یک نمونه تصادفی ۱۰ تایی از یک جامعه نرمال با واریانس 42.5 گرفته شده است، احتمال اینکه

انحراف معیار نمونه بین 3.14 تا 8.94 باشد را پیدا کنید.

$$P(3.14 < S < 8.94) = P(3.14^2 < S^2 < 8.94^2)$$

$$= P\left(\frac{9 \times 3.14^2}{42.5} < \frac{9\,S^2}{42.5} < \frac{9 \times 8.94^2}{42.5}\right)$$

$$= P(2.09 < Y < 16.9) = *$$

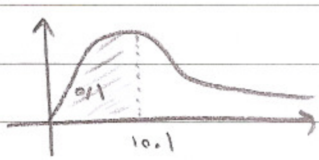
$$Y \sim \chi^2_{(9)}$$

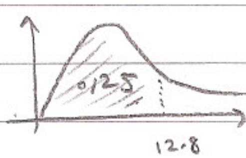
با استفاده از جدول

$$\Rightarrow * = P(Y < 16.9) - P(Y < 2.09) = 0.95 - 0.01 = 0.$$

مثال- مطلوبست محاسبه مقدار $\chi^2_{0.2}(17)$.

طبق جدول $\chi^2_{0.25}(17) = 12.8$ و $\chi^2_{0.1}(17) = 10.1$





با یک تناسب، بطور تقریبی مقدار $\chi^2_{0.2}(17)$ را محاسبه می‌کنیم.

$$0.25 - 0.1 \qquad 12.8 - 10.1$$

$$0.2 - 0.1 \qquad x \qquad \Rightarrow x = 1.8$$

$$\Downarrow$$

$$\chi^2_{0.2}(17) = \chi^2_{0.1}(17) + 1.8 = 10.1 + 1.8 = 11.$$

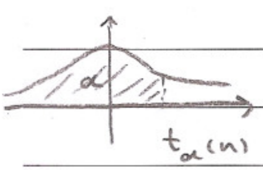
3 – توزیع میانگین نمونه (با فرض نامعلوم بودن واریانس جامعه)

فرض کنید در جامعه نرمال X، مقدار σ^2 مجهول باشد، به جای آن می‌توان از واریانس نمونه S^2 استفاده کرده. در ادامه ابتدا توزیع t را معرفی، سپس به توزیع $Z = \dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ می‌پردازیم.

توزیع t (t-student)

اگر متغیر تصادفی $Z \sim N(0,1)$ و $Y \sim \chi^2(n)$ ، Z و Y از یکدیگر مستقل باشند، در این صورت متغیر تصادفی $T = \dfrac{Z}{\sqrt{\frac{Y}{n}}}$ دارای توزیع t با درجه آزادی n است و آن را با نماد $T \sim t(n)$ نمایش می‌دهیم. تابع چگالی احتمال T عبارت است از

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad t \in \mathbb{R}$$



$$T \sim t(n) \Rightarrow P(T \le t_\alpha(n)) = \alpha$$

برای مثال - $t_{0.9}(15) = 1.34$, $t_{0.95}(20) = 1.72$

$t_{0.975}(30) = 2.04$

Date:........................ Subject:........................

قضیه- اگر $\bar{X}$ و S^2 میانگین و واریانس نمونه یک نمونه تصادفی n تایی از یک جامعه نرمال

با میانگین μ باشد، آنگاه

$$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

تذکر- اگر $n \to \infty$ (بعبارت دیگر $n \geq 30$) دراینصورت توزیع حدی T

یک توزیع نرمال استاندارد است.

مثال- نمرات دانشجویان یک کلاس دارای توزیع نرمال با میانگین 15 است. اگر از این

کلاس یک نمونه 20 تایی انتخاب کنیم و واریانس نمونه آن 4.28 باشد، احتمال

اینکه میانگین نمرات این نمونه از 17 بیشتر باشد، چقدر است؟

$$P(\bar{X} > 17) = P\left(\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} > \frac{17-15}{\frac{4.28}{\sqrt{20}}}\right) = P(T > 2.09)$$

$$= 1 - \underbrace{P(T \leq 2.09)}_{T \sim t(19)} = 1 - 0.975 = 0.025.$$

Date:............................ Subject:..

4- توزیع اختلاف میانگین ها

فرض کنید دو جامعه داریم که اولی دارای میانگین μ_1 و واریانس σ_1^2 و دومی دارای میانگین μ_2 و واریانس σ_2^2 است. یک نمونه n_1 تایی از جامعه اول با میانگین نمونه‌ای $\bar{X}_1$ و واریانس نمونه‌ای S_1^2 و یک نمونه n_2 تایی از جامعه دوم دارای میانگین نمونه‌ای $\bar{X}_2$ و واریانس نمونه‌ای S_2^2 انتخاب می‌کنیم. فرض کنید نمونه گیری از دو جامعه مستقل از یکدیگر است. هدف یافتن توزیع $\bar{X}_1 - \bar{X}_2$ است.

الف - واریانس دو جامعه σ_1^2 و σ_2^2 معلوم است.

(I) اگر دو جامعه نرمال باشند:

$$\bar{X}_1 \sim N(\mu_1 , \frac{\sigma_1^2}{n_1})$$

$$\bar{X}_2 \sim N(\mu_2 , \frac{\sigma_2^2}{n_2})$$

$\bar{X}_1$ و $\bar{X}_2$ از یکدیگر مستقل هستند، در نتیجه $\bar{X}_1 - \bar{X}_2$ نیز نرمال است.

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

بعبارت دیگر

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

AZIN

Date:.................... Subject:....................

(II) اگر دو جامعه نرمال نباشد ولی $n_1 \geq 30$ و $n_2 \geq 30$. در این حالت نیز $\bar{X}_1 - \bar{X}_2$

تقریباً نرمال است و توزیع بدست آمده در حالت قبل برقرار است.

مثال- یک نمونه تصادفی 25 تایی از یک جامعه نرمال با حد متوسط 80 و انحراف معیار 5 انتخاب شده است. یک نمونه تصادفی 36 تایی از یک جامعه نرمال دیگری با حد متوسط 75 و انحراف معیار 3 انتخاب شده است. احتمال اینکه میانگین نمونه 25 تایی حداقل 3.4 از میانگین نمونه 36 تایی بیشتر باشد را بیابید.

$$\bar{X}_1 \sim N\left(80, \frac{5^2}{25}\right) \qquad \bar{X}_2 \sim N\left(75, \frac{3^2}{36}\right)$$

$$P(\bar{X}_1 - \bar{X}_2 > 3.4) = P\left(\frac{\bar{X}_1 - \bar{X}_2 - (80-75)}{\sqrt{1 + \frac{1}{4}}} > \frac{3.4-5}{\sqrt{\frac{5}{4}}}\right)$$

$$= P(Z > -1.43) = P(Z \leq -1.43) = 0.0764$$

مثال- از دو جامعه نرمال با میانگین های مساوی و واریانس های 50 و 40 ، دو نمونه تصادفی 100 تایی بطور مستقل انتخاب کرده ایم. احتمال اینکه تفاضل میانگین ها بیش از 2 باشد.

$$\bar{X}_1 \sim N\left(\mu_1, \frac{500}{100}\right) \qquad \bar{X}_2 \sim N\left(\mu_2, \frac{40}{100}\right) \xRightarrow{\mu_1 = \mu_2} \bar{X}_1 - \bar{X}_2 \sim N(0, 0.9)$$

$$\Rightarrow P(|\bar{X}_1 - \bar{X}_2| > 2) = P\left(\left|\frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{0.9}}\right| > \frac{2-0}{\sqrt{0.9}}\right)$$

$$= P(|Z| > 2.11) = 1 - P(-2.11 \leq Z \leq 2.11) = 1 - 0.9826 + 0.0174 = 0.034$$

Date:............ Subject:............

ب - واریانس دو جامعه نامعلوم ولی مساوی می‌باشد. $\sigma_1^2 = \sigma_2^2 = \sigma^2$

در اینصورت σ^2 را با واریانس وزندار S_p^2 برآورد می‌کنیم:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

پس اگر دو جامعه نرمال باشند،

$$\frac{(n_1+n_2-2)\,S_p^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2)}$$

درنتیجه

$$T = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2).$$

5- توزیع نسبت واریانس ها

ابتدا توزیع فیشر را معرفی می‌نمائیم:

اگر $U \sim \chi^2_{(m)}$ و $V \sim \chi^2_{(n)}$ و U و V از یکدیگر مستقل باشند،

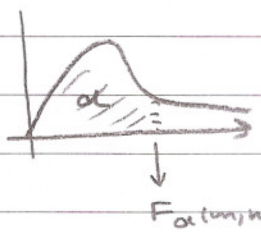
آنگاه

$$F = \frac{\frac{U}{m}}{\frac{V}{n}}$$

دارای توزیع فیشر با درجه آزادی m و n می‌نامند و به $F \sim F(m,n)$ نمایش می‌دهند.

تابع چگالی احتمال F عبارتست از $\quad \forall x>0$

$$f_F(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\,\Gamma(n/2)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}}$$



$$F \sim F(m,n) \Rightarrow \mathbb{P}(F < F_\alpha(m,n)) = \alpha$$

قضیه - اگر S_1^2 و S_2^2 به ترتیب واریانس های نمونه های تصادفی مستقل به اندازه n_1 و n_2 از جامعه های نرمال با واریانس های σ_1^2 و σ_2^2 باشند، آنگاه

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1-1, n_2-1).$$

مثال - از دو جامعه نرمال با واریانس های 20 و 30 به ترتیب نمونه های تصادفی ۸ و ۱۰ تایی انتخاب کرده ایم. احتمال اینکه واریانس نمونه اول بیش از 2 برابر واریانس نمونه دوم باشد را بیابید.

$$\mathbb{P}(S_1^2 > 2S_2^2) = \mathbb{P}\left(\frac{S_1^2}{S_2^2} > 2\right) = \mathbb{P}\left(\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} > \frac{\sigma_2^2}{\sigma_1^2}\,2\right)$$

$$= \mathbb{P}\left(F > \frac{30}{20}\times 2\right) = 1 - \mathbb{P}(F \leq 3)$$

$$\boxed{F \sim F(7,9)}$$

$$\approx 1 - 0/95 = 0/05.$$