توزیع‌های نمونه‌ای

دانشگاه مازندران

گروه آمار

توزیع t $(t - student)$

اگر $Z \sim N(0,1)$ و $Y \sim \chi^2(n)$ و Z و Y مستقل باشند در اینصورت متغیر تصادفی T که به صورت زیر تعریف می شود دارای توزیع t با n درجه آزادی است.

$$T = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

یعنی

$$\begin{cases} Z \sim N(0,1), \\ Y \sim \chi^2(n) \\ X, Y \ \text{مستقل} \end{cases} \Rightarrow \frac{Z}{\sqrt{\frac{Y}{n}}} \sim t(n)$$

تابع چگالی توزیع $t(n)$: اگر $T \sim t(n)$ می توان نشان داد که تابع چگالی T به صورت زیر می باشد:

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{\sqrt{n\pi}(1+\frac{t^2}{n})^{\frac{n+1}{2}}}, \qquad -\infty < t < \infty$$

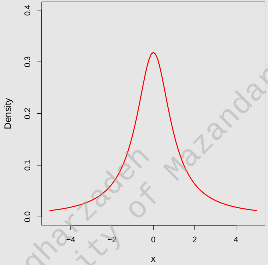
میانگین و واریانس توزیع t

اگر $T \sim t(n)$ می توان نشان داد که

$$\mathbb{E}(T) = ۰, \quad n > ۱$$

$$Var(T) = \frac{n}{n-۲}, \quad n > ۲$$

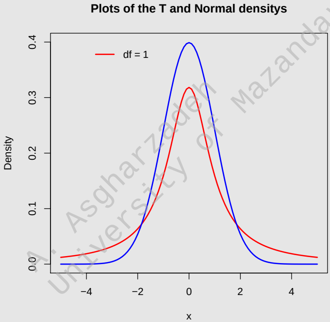
منحنی تابع چگالی توزیع t با درجه آزادی $n = ۱$:



برنامه R

```
curve(dt(x,1), -5, 5, type="l", col="red",ylim=c(0, 0.4))
```

مقایسه منحنی تابع چگالی توزیع‌های t و نرمال استاندارد: منحنی چگالی توزیع t شبیه منحنی تابع چگالی توزیع نرمال استاندارد متقارن و زنگوله ای شکل می باشد با این تفاوت که در دو انتهای منحنی چگالی t احتمال بیشتری نهفته است در حالیکه در مرکز یعنی حول صفر احتمال بیشتری برای نرمال استاندارد وجود دارد.



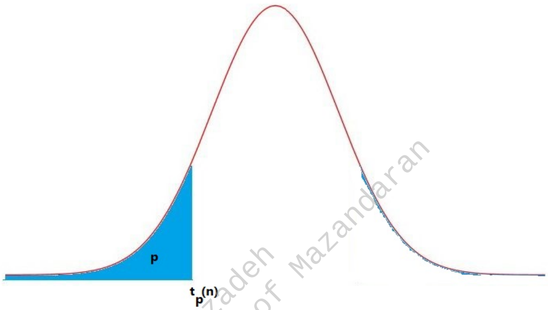
شکل: مقایسه تابع چگالی توزیع t (قرمز) و نرمال استاندارد (آبی)

برنامه R رسم نمودار تابع چگالی توزیع t در مقایسه با توزیع نرمال

```
curve(dt(x,1), -5, 5, type="l", col="red",ylim=c(0, 0.4), ylab="Density")
curve(dnorm(x), -5, 5, type="l", col="blue",lwd=2,add=T)
```

dt(x, 1): تابع چگالی توزیع t با درجه آزادی ۱ در بازه $(-۵, ۵)$

dnorm(x): تابع چگالی توزیع نرمال استاندارد



قرارداد. مطابق شکل $t_p(n)$ نقطه‌ای است روی محور طول‌ها که سطح زیر منحنی در سمت چپ این نقطه p می‌شود یعنی

$$P(T(n) \ < \ t_p(n)\ ) \ = \ p$$

$t_p(n)$ را چندک مرتبه p-ام توزیع t با درجه آزادی n می‌نامند.

انجام محاسبات براساس توزیع‌های آماری در R

رابطه (فرمول)	محاسبه	تابع
$f_X(x)$	تابع احتمال یا چگالی احتمال (density)	$d +$ نام توزیع در R
$P(X \leq q)$	تابع توزیع احتمال تجمعی Probability distribution function	$p +$ نام توزیع در R
$F^{-1}(p)$	چندک‌های توزیع احتمال (quantiles)	$q +$ نام توزیع در R
	تولید عدد تصادفی از توزیع دلخواه random variable	$r +$ نام توزیع در R

مثال. محاسبه چندک و تابع توزیع در توزیع‌های t و نرمال استاندارد در R:

- در توزیع t با ۵ درجه آزادی، مقدار چندک برای $p = ۰/۰۵$ به صورت زیر است:
qt(0.05, 5)=-2.015

- در توزیع t با ۱۰ درجه آزادی، مقدار چندک برای $p = ۰/۰۱$ به صورت زیر است:
qt(0.01, 10)=-2.764

- در توزیع نرمال استاندارد، $P(X < ۲)$ به صورت زیر محاسبه می‌شود:
pnorm(2)=0.977

مثال.

$$i)\ P(T(۱۰) > ۱/۳۷۲) = ۱ - P\Big(T(۱۰) < ۱/۳۷۲\Big)$$
$$= ۱ - P\Big(T(۱۰) < t_{۰/۹}(۱۰)\Big) = ۱ - ۰/۹ = ۰/۱$$

$$ii)\ P(T(۹) < ۳/۲۵) = P(T(۹) < ۳/۲۵)$$
$$= P\Big(T(۹) < t_{۰/۹۹۵}(۹)\Big) = ۰/۹۹۵$$

روش دوم: با استفاده از برنامه R

i) $P(T(10) > 1.372) = 1 - P(T(10) < 1.372) = 1 - F_T(1.372, 10)$

```
> 1-pt(1.372, 10)
[1] 0.10
```

ii) $P(T(9) < 3.25) = F_T(3.25, 9)$

```
> pt(3.25, 9)
[1] 0.995
```

تذکر. اگر n یعنی درجه آزادی بزرگ باشد می توان توزیع t را به وسیله توزیع نرمال استاندارد تقریب زد.
یعنی $(n \geq ۳۰)$

$$\lim_{n\to\infty} f(t) = \frac{۱}{\sqrt{۲\pi}} e^{-\frac{t^۲}{۲}}, \qquad -\infty < t < \infty$$

مثال.

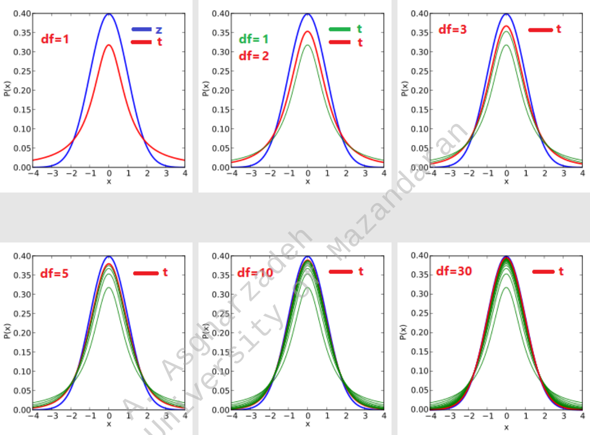
$$P(T(۵۰) < ۱/۵) \approx P(Z < ۱/۵) = ۰/۹۳۳۲$$

برنامه R

$$P(T(۵۰) < ۱/۵) = F_T(۱/۵,\ ۵۰)$$
$$P(Z < ۱/۵) = \Phi(۱/۵)$$

```
> pt(1.5, 50)
  [1] 0.930

> pnorm(1.5)
  [1] 0.933
```



شکل: منحنی تابع چگالی توزیع t در مقایسه با منحنی تابع چگالی توزیع نرمال استاندارد برای های n مختلف

توزیع فیشر یا F

تعریف: اگر $U \sim \chi^{۲}(m)$ و $V \sim \chi^{۲}(n)$ و U و V مستقل باشند آنگاه متغیر تصادفی F که به صورت زیر تعریف می شود

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

دارای توزیع فیشر با درجات آزادی m و n خواهد بود که می‌نویسیم $F \sim f(m, n)$ یعنی

$$\begin{cases} U \sim \chi^{۲}(m), \\ V \sim \chi^{۲}(n), \\ U,\ V \text{ مستقل} \end{cases} \Rightarrow \frac{U/m}{V/n} \sim f(m,n)$$

تابع چگالی

اگر $X \sim f(m,\ n)$ می‌توان نشان داد که تابع چگالی آن به صورت زیر می‌باشد:

$$f_X(x) = \frac{T(\frac{m+n}{۲})}{\Gamma(\frac{m}{۲})\,\Gamma(\frac{n}{۲})} \left(\frac{m}{n}\right)^{\frac{m}{۲}} \frac{x^{\frac{m-۲}{۲}}}{[۱ + \frac{m}{n}x]^{\frac{m+n}{۲}}}, \qquad x > ۰$$

حالت خاص از توزیع فیشر:

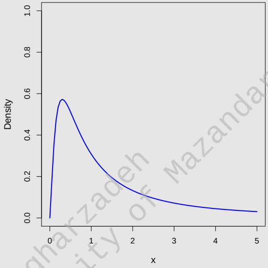
$$X \sim f(۲,\ ۲) \ \Rightarrow f(x) = \frac{۱}{(۱+x)^۲}, \quad x > ۰$$

میانگین و واریانس. اگر $X \sim f(m\,,n)$ می‌توان نشان داد که

$$\mathbb{E}(X) = \frac{n}{n-۲}, \qquad n > ۲$$

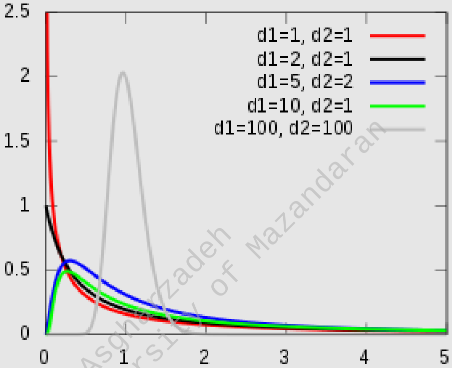
$$Var(X) = \frac{۲n^۲(m+n-۲)}{m(n-۲)^۲\,(n-۴)}, \qquad n > ۴$$

منحنی تابع چگالی توزیع فیشر با درجات آزادی $m = ۵, n = ۲$:



برنامه R

```
curve(df(x,5, 2), 0, 5, type="l", col="blue",ylim=c(0, 1))
```



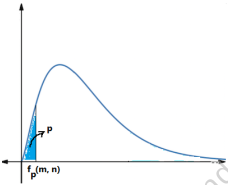
شکل: نمودار تابع چگالی احتمال توزیع فیشر به ازای مقادیر مختلف m و n

در آن $d۱$ و $d۲$ به ترتیب مقادیر m و n می‌باشند.

برنامه R منحنی اسلاید قبل

```
curve(df(x,2, 1),0,5, type="l", col="red", ylim=c(0,2), ylab="Density")
curve(df(x, 5, 2), 0, 5, type="l", col="blue", lwd=2, add=T)
curve(df(x,10, 1), 0, 5, type="l", col="green", lwd=2, add=T)
curve(df(x,100, 100), 0, 5, type="l", lwd=2, add=T)
```

منحنی توزیع F. منحنی تابع چگالی احتمال توزیع فیشر برای $n, m > ۱$ به صورت زیر خواهد بود.



قرارداد. مطابق شکل $f_p(m, n)$ نقطه‌ای است روی محور طول‌ها که سطح زیر منحنی در سمت چپ این نقطه p می‌شود یعنی

$$P(F_{(m,n)} < f_p(m, n)) = p$$

مقادیر $f_p(m, n)$ را می‌توان برای m و n های مختلف از روی جدول توزیع F پیدا کرد.
شبیه به منحنی توزیع t متقارن و زنگوله‌ای شکل نیست و یک منحنی نامتقارن و چوله به راست است.

مثال. مقادیر زیر را از روی جدول بیابید.

$$i)\ f_{۰/۰۵}(۶,\ ۷)\ =\ ۰/۲۳۷ \qquad ii)\ f_{۰/۰۱}(۱۰,\ ۱۵)\ =\ ۰/۲۱۹$$

برنامه R

```
qf(0.05, 6, 7)
[1] 0.2377
qf(0.01, 10, 15)
[1] 0.2193
```


$$iii)\ P(F(۵,۱۰)\ <\ ۵/۶۴)\ =\ P(F(۵,۱۰)\ <\ f_{۰/۰۱}(۵,۱۰))=۰/۹۹$$

برنامه R

```
pf(5.64, 5, 10)
[1]  0.99
```

چند نکته در رابطه با توزیع فیشر و توزیع t .

$$i)\quad X \sim t(n) \;\Rightarrow\; X^{۲} \sim f(۱, n)$$

$$ii)\quad X \sim f(m, n) \Rightarrow \frac{۱}{X} \sim f(n, m)$$

$$iii)\quad f_p(m, n) = \frac{۱}{f_{۱-p}(n, m)}$$

نمونه تصادفی: متغیرهای تصادفی $X_1, X_2, \ldots, X_n$ را یک نمونه تصادفی از جامعه ای نامتناهی با تابع چگالی $f(x)$ یا $(f(x,\theta))$ گویند هرگاه این متغیرها مستقل و هم توزیع باشند. θ پارامتر جامعه است (پارامتر موجود در تابع چگالی جامعه می باشد)

۱) شرط استقلال

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1).f_{X_2}(x_2)\ldots f_{X_n}(x_n)$$

۲) شرط هم توزیع

$$f_{X_1}(t) = f_{X_2}(t) = \ldots = f_{X_n}(t)$$

نمونه تصادفی را با i.i.d نمایش می دهند که

$$i.i.d = \text{''}\, independent\ indentical\ distributed \text{''}$$

به عنوان مثال اگر $X_1,\ldots,X_n$ یک نمونه تصادفی از توزیع $N(\mu,\sigma^2)$ باشد می نویسیم

$$X_1,\ldots,X_n \overset{i.i.d}{\sim} N(\mu,\sigma^2)$$

یافته های نمونه تصادفی: اگر برای نمونه تصادفی $X_1, X_2, \ldots, X_n$ در عمل مقادیر $x_1, x_2, \ldots, x_n$ را مشاهده کنیم در آن صورت $x_1, x_2, \ldots, x_n$ را یافته های نمونه تصادفی گویند.

مثال. اگر سکه ای را ۵ بار پرتاب کنیم و X_i را نتیجه پرتاب i ام بگیریم، در اینصورت

$$X_1,\ X_2,\ X_3,\ X_4,\ X_5 \overset{i.i.d}{\sim} f_X(x) \qquad H \leftrightarrow 1,\ T \leftrightarrow 0$$

$$f_X(x) = P(X = x) = \frac{1}{2}, \qquad x = 0,\ 1$$

حال اگر در ۵ پرتاب سکه نتایج ۱، ۰، ۱، ۰، ۰ مشاهده شود در اینصورت یافته های نمونه تصادفی $X_1,\ X_2,\ X_3,\ X_4,\ X_5$ می شود

$$x_1 = 1, \quad x_2 = 0, \quad x_3 = 1, \quad x_4 = 0, \quad x_5 = 0$$

آماره: هر تابعی از نمونه تصادفی است مانند $U = g(X_1, \ldots, X_n)$ که در ساختار آن پارامتر مجهول θ نباشد که آن را معمولا با حروف بزرگ نمایش می دهند.

مثال. با توجه به $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} f(x;\theta)$ که در آن θ پارامتر مجهول می باشد، کدامیک از موارد زیر آماره هست؟

۱) $X_n - X_1$ $\checkmark$

۲) $\sum_{i=1}^{n} X_i$ $\checkmark$

۳) $\max(X_1, X_2, \ldots, X_n)$ $\checkmark$

۴) $\min(X_1 - \theta, X_2 - \theta, \ldots, X_n - \theta)$ $\times$

۵) $\sum_{i=1}^{n} X_i^2 - \theta^2$ $\times$

* به $U = \sum_{i=1}^{n} X_i$ آماره و $v = \sum_{i=1}^{n} x_i$ یافته آماره گفته می شود.

تعریف

اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی باشند آنگاه میانگین و واریانس نمونه ای به صورت زیر تعریف می شوند:

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

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

قضیه

اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از جامعه ای دلخواه با میانگین μ و واریانس σ^2 باشد آنگاه

$$E(\bar{X}) = \mu$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

قضیه

اگر $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ و $\bar{X}$ و S^2 به ترتیب میانگین و واریانس نمونه ای باشند آنگاه

۱. $\bar{X}$ و S^2 مستقل اند.

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

قضیه

اگر $\bar{X}$ و $S^۲$ به ترتیب میانگین و واریانس نمونه ای به حجم n از جامعه نرمال با میانگین μ و واریانس $\sigma^۲$ باشند آنگاه

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-۱)$$

قضیه

اگر $\bar{X}$ و $S^۲$ به ترتیب میانگین و واریانس نمونه ای به حجم n از جامعه نرمال با میانگین μ و واریانس $\sigma^۲$ باشند آنگاه

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-۱)$$

قضیه

اگر $S_۱^۲$ و $S_۲^۲$ به ترتیب واریانس های نمونه تصادفی مستقل به حجم های m و n از جوامع نرمال با واریانس های $\sigma_۱^۲$ و $\sigma_۲^۲$ باشند. در آن صورت

$$\frac{S_۱^۲}{S_۲^۲} . \frac{\sigma_۲^۲}{\sigma_۱^۲} \sim f(m-۱, n-۱)$$

تمرین ۱. اگر $N(۰,\ ۱) \overset{i.i.d}{\sim} Z_۱, Z_۲, \ldots, Z_{۱۶}$ باشد. پیدا کنید

$$a)\ P(\bar{Z} < \frac{۱}{۲}) \qquad b)\ P(Z_۱ - Z_۲ < ۲)$$

$$c)\ P(\sum_{i=۱}^{۱۶} Z_i^۲ < ۳۲) \qquad d)\ P(\sum_{i=۱}^{۱۶}(Z_i - \bar{Z})^۲ < ۲۵)$$

تمرین ۲. اگر $N(\mu, \sigma^۲) \overset{i.i.d}{\sim} X_۱, X_۲, \ldots, X_n$ و $Z_۱, Z_۲, \ldots, Z_k \overset{i.i.d}{\sim} N(۰, ۱)$ دو نمونه مستقل باشند. در هر مورد توزیع‌های زیر را پیدا کنید.

$$a)\ X_۱ - X_۲ \qquad b)\ \frac{Z_۱^۲}{Z_۲^۲}$$

$$c)\ \frac{\bar{X}}{\sigma^۲} + \frac{۱}{k}\sum_{i=۱}^{k} Z_i \qquad d)\ \frac{Z_۱}{\sqrt{Z_۲^۲}}$$

$e)\ k\,\bar{Z}^2$　　　$f)\ \dfrac{\bar{X}}{\bar{Z}}$

$g)\ \dfrac{X_1 - X_2}{\sigma\, S_Z \sqrt{2}}$　　　$h)\ \dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} + \sum_{i=1}^{k}(Z_i-\bar{Z})^2$

تمرین ۳. اگر $X_1, X_2, \ldots, X_{10} \overset{i.i.d}{\sim} N(1, 3)$ باشد پیدا کنید $P(\bar{X} > 3,\ S^2 < 1/39)$.