تعاریف

آمار(Statistics): مجموعه ای از مفاهیم و روش ها است که در زمینه هـای پژوهـشی بـرای گـردآوری اطلاعات، تجزیه و تحلیل آنها و نتیجه گیری معتبر در مورد جامعه آماری هنگامی که عدم قطعیت وجـود دارد.

سئوال پژوهشی: آیا بین جنسیت و علاقه به رشته تحصیلی رابطه وجود دارد؟

جامعه آماری: مجموعه ای از افراد یا اشیاء که در یک زمان و مکان مشخص حداقل در یک ویژگـی یـا صفت مشترک باشند. برای مثال دانشجویان دانشگاه آزاد اسلامی آبـادان در نیمـسال دوم سـال تحـصیلی ۹۹-۱۳۹۸ تشکیل یک جامعه آماری می دهند. صفت هایی که در این جامعه آماری ممکن اسـت مطالعـه کنیم شامل سن، جنسیت، رشته تحصیلی، سال ورود، معدل و... است.

متغیر: صفت مورد مطالعه در جامعه آماری از فردی به فرد دیگر یا از شیئی بـه شـیء دیگـر قابـل تغییـر است بنابراین به آن متغیر گفته می شود.متغیرها به دو دسته کلی تقسیم می شوند:
الف) متغیرهای کیفی: که در این گونه متغیرهای جامعه آماری براساس صفت مورد مطالعـه طبقـه بنـدی می شود و اطلاعات در هر طبقه از طریق شمارش بدست می آید. مانند متغیرهای جنسیت، نـوع مـدرک دیپلم،گروه خونی و...

ب) متغیرهای کمی یا عددی: اطلاعات حاصل از مطالعه این متغیرها بـصورت اعـداد حقیقـی یـا ارزش ریاضی هستند. مانند سن، معدل، وزن و...

احتمال: در تعبیر احتمال به طریق فراوانی نسبی نسبت رخ دادن یک پیشامد را در تکرار زیاد آزمایش به عنوان احتمال رخ دادن آن پیشامد در نظر می گیریم

$$P(A)=\lim_{n\to\infty}\frac{n_A}{n}$$

که n تعداد تکرار های آزمـایش و n_A تعـداد رخ دادن پیـشامد A اسـت. همچنـین اگـر فـرض شـود برآمدهای فضای نمونه S دارای احتمال های مساوی بـرای رخ دادن هـستند و $n(S)$ تعـداد برآمـدهای فضای نمونه باشد و $n(A)$ تعداد برآمدهای پیشامد A باشد در تعبیر احتمال به طریق هم شانسی

$$P(A)=\frac{n(A)}{n(S)}$$

است. اگر A و B متعلق به فضای نمونه S باشند ۳ قانون اصلی زیر را برای احتمال خواهیم داشت

الف) $0 \leq P(A) \leq 1$

ب) $P(S) = 1$

ج)اگر A و B پیشامدهای ناسازگار باشند($A \cap B = \phi$) آنگاه

$$P(A \cup B) = P(A) + P(B)$$

همچنین قوانین فرعی دیگری را می توان داشت برای مثال قانون متمم احتمال

$$P(A) = 1 - P(A')$$

احتمال شرطی پیشامد A اگر بدانیم پیشامد B رخ داده است

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad ,P(B) \neq 0$$

بنابراین اگر بدانیم B رخ داده است احتمال وقوع هر دو پیشامد با هم

$$P(A \cap B) = P(B) \times P(A|B)$$

و پیشامدهای A و B مستقل هستند اگر احتمال شرطی آنها با احتمال ساده آنها برابر باشد

$$P(A|B) = P(A), \quad P(B|A) = P(B)$$

در نتیجه پیشامدهای A و B مستقل هستند اگر وتنها اگر

$$P(A \cap B) = P(A) \times P(B)$$

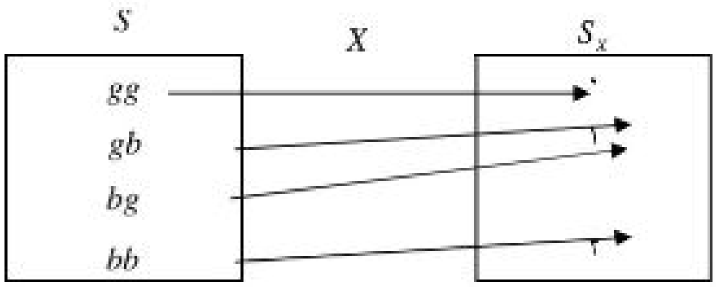
متغیر تصادفی

تابعی است که به هر مجموعه از فضای نمونه یک عدد حقیقی را می دهد

$$X : S \to R$$

در این درس متغیرهای تصادفی را با حروف بزرگ $X, Y, Z, ...$ ویافته های آن را در آزمایش تصادفی با حروف کوچک لاتین $x, y, z, ...$ نشان می دهیم. به مجموعه یافته های متغیر تصادفی X تکیه گاه گفته شده و آن را با S_X نشان می دهیم.

مثال ۱: یک خانواده دار ی ۲ فرزند است متغیر تصادفی X را تعداد فرزندان پسر خانواده در نظر می گیریم



برای مثال $X(\{bb\}) = ۲$ بنابراین علیرغم اینکه تابع بودن متغیر تصادفی سهواً به آن متغیر می گوئیم. متغیرهای تصادفی به ۲ دسته کلی تقسیم می شوند:

الف) متغیرهای تصادفی گسسته: در اینگونه متغیرها تکیه گاه مجموعه شمارش پذیر متناهی یا نامتناهی از اعداد حقیقی است مانند مثال ۱ که $S_X = \{۰,۱,۲\}$.

ب) متغیرهای تصادفی پیوسته: در اینگونه متغیرها تکیه بازه ای از اعداد حقیقی است بنابراین شمارش پذیر نیست. برای مثال اگر نقطه ای بصورت تصادفی در دایره ای به شعاع یک سانتی متر در نظر بگیریم و متغیر تصادفی X را فاصله نقطه تا مرکز دایره بگیریم $S_X = (۰,۱)$.

تابع احتمال

برای متغیر تصادفی X با تکیه گاه S_X تابع احتمال آن دارای دو خاصیت زیر است:

الف) $f(x) \geq ۰ \quad \forall x \in S_X$

ب) در حالت اینکه X پیوسته باشد $\int_{S_X} f(x)dx = ۱$ و اگر X گسسته باشد $\sum_{S_X} f(x) = ۱$.

مثال۲: در مثال ۱ تابع احتمال X را بنویسید؟

x	۰	۱	۲	جمع
$f(x) = P(X = x)$	$\frac{۱}{۴}$	$\frac{۱}{۲}$	$\frac{۱}{۴}$	۱

برای مثال $f(۰) = P(\{gg\}) = \frac{۱}{۴}$.

مثال۳: نشان دهید تابع زیر یک تابع احتمال است؟

$$f(x) = ۲x \qquad ۰ < x < ۱$$

حل: شرط اول برای هر مقدار متغیر در تکیه گاه $S_X = (۰,۱)$ برقرار است. برای شرط دوم چون متغیر تصادفی پیوسته است باید انتگرال روی تکیه گاه آن یک باشد

$$\int_۰^۱ ۲xdx = x^۲ \Big|_۰^۱ = ۱ - ۰ = ۱$$

تابع توزیع احتمال

برای متغیر تصادفی X تابع توزیع آن در نقطه x بصورت زیر تعریف می شود

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

و در صورتی که X گسسته باشد $F(x) = \sum_{t \le x} f(t)$. تابع توزیع دارای ۳ خاصیت زیر است:

الف) غیر نزولی است

ب)حداقل از سمت راست پیوسته است.

ج) $F(\infty) = ۱$ و $F(-\infty) = ۰$.

مثال۳: در مثال ۲ تابع احتمال تجمعی آن را محاسبه کنید؟

حل:

$$F(x) = \int_{۰}^{x} ۲t dt = t^{۲} |_{۰}^{x} = x^{۲} - ۰ = x^{۲}$$

بنابراین

$$F(x) = \begin{cases} ۰ & ۰ \le x \\ x^{۲} & ۰ < x < ۱ \\ ۱ & x \ge ۱ \end{cases}$$

مثال٤: در مثال ۲ تابع احتمال تجمعی آن را محاسبه کنید؟

حل:

$$F(۰) = P(X \le ۰) = f(۰) = \frac{۱}{٤}$$

$$F(۱) = P(X \le ۱) = f(۰) + f(۱) = \frac{۱}{٤} + \frac{۱}{۲} = \frac{۳}{٤}$$

$$F(۲) = P(X \le ۲) = f(۰) + f(۱) + f(۲) = \frac{۱}{٤} + \frac{۱}{۲} + \frac{۱}{٤} = ۱$$

بنابراین

$$F(x) = \begin{cases} ۰ & x < ۰ \\ \frac{۱}{٤} & ۰ \le x < ۱ \\ \frac{۳}{٤} & ۱ \le x < ۲ \\ ۱ & x \ge ۲ \end{cases}$$

نکته: از روی تابع احتمال تجمعی می توان تابع احتمال را به صورت زیر مشخص کرد در حالت پیوسته

$$\frac{dF(x)}{dx} = f(x)$$

و در حالتی که متغیر تصادفی X گسسته باشد $f(x) = F(x) - F(x^-)$ که x^- نقطه قبلی x در تکیه گـاه است. برای مثال ٤ احتمال در نقطه ١ بصورت زیر است

$$F(1) = f(1) - f(0) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

امید ریاضی و واریانس

برای متغیر تصادفی X امید ریاضی بیانگر مقداری است که انتظار می رود آن متغیر به طـور متوسـط در آزمایش تصادفی داشته باشد به همین دلیل به آن امید ریاضی گفته می شود که در حالت پیوسته بصورت زیر محاسبه می شود:

$$\mu \quad or \quad E(X) = \int_{S_X} x f(x) dx$$

و در حالت گسسته $E(X) = \sum_{S_X} x f(x)$. همچنین برای محاسبه پراکندگی متغیر تـصادفی واریـانس آن بصورت زیر محاسبه می شود:

$$\sigma^2 \quad or \quad V(X) = E(X - \mu)^2 = E(X^2) - \mu^2$$

که $E(X^2) = \int_{S_X} x^2 f(x) dx$.

مثال ٥: در مثال ٣ امید ریاضی و انحراف معیار X را محاسبه کنید؟

حل: مقدار مورد انتظار X برابر است با

$$E(X) = \int_0^1 x.2x dx = \int_0^1 2x^2 dx = \frac{2}{3} x^3 \Big|_0^1 = \frac{2}{3} - 0 = \frac{2}{3}$$

برای بدست آوردن واریانس ابتدا مقدار مورد انتظار X^2 محاسبه می شود

$$E(X^2) = \int_0^1 x^2.2x dx = \int_0^1 2x^3 dx = \frac{1}{2} x^4 \Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$$

بنابراین $V(X) = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$ در نتیجه انحراف معیار $SD(X) = \sqrt{\frac{1}{18}} = \frac{\sqrt{2}}{6}$.

نکته: اگر متغیر تصادفی X را در عدد ثابت $a \neq 0$ ضرب کرده و با عدد ثابت b ضرب کنیم آنگاه میانگین و انحراف معیار متغیر تصادفی $Y = aX + b$ بصورت زیر است:

$E(Y) = aE(X) + b$
$V(Y) = a^2 V(X)$
$SD(Y) = |a| SD(X)$

استاندارد کردن متغیر تصادفی

فرض کنید متغیر تصادفی X دارای میانگین $E(X)$ و انحراف معیار $SD(X)$ باشد است آنگاه استاندار شده متغیر تصادفی X بصورت زیر است که دارای میانگین صفر و انحراف معیار یک می باشد(مقدار $a = 1/\sigma$ و $b = -\mu/\sigma$ فرض شود):

$$Z = \frac{X - E(X)}{SD(X)} \quad or \quad Z = \frac{X - \mu}{\sigma}$$

مثال ٦: اگر متغیر تصادفی X دارای میانگین ٤ و انحراف معیار ٣ باشد میانگین و واریانس $Y = ۲X - ۱$ را محاسبه کنید؟

حل: با استفاده از نکته قبلی خواهیم داشت:

$E(Y) = ۲(٤) - ۱ = ۸ - ۱ = ۷$

و $SD(X) = ۲(۳) = ٦$ بنابراین $V(X) = (٦)^۲ = ۳٦$.

برای مطالعه یک یا چند صفت در جامعه آماری دو را وجود دارد:

الف) سرشماری: اگر همه اعضاء جامعه آماری را بر اساس صفت های مورد مطالعه بررسی کنیم سرشماری است. برای مثال اگر سن همه دانشجویان دانشگاه آزاد اسلامی آبادان بررسی شود سرشماری است. سرشماری معایبی دارد از جمله وقت گیر است، مشکلات اجرایی و اقتصادی دارد و گاهی اوقات ممکن است باعث از بین رفتن جامعه آماری شود.

ب)نمونه گیری: اگر تعداد محدودی از جامعه آماری را بصورت شانسی یا تصادفی انتخاب کرده و بجای کل جامعه آماری این تعداد محدود را بررسی کنیم شیوه نمونه گیری است. نمونه انتخاب شده از جامعه آماری باید بگونه ای باشد که ویژگی های کلی جامعه را در خود داشته باشد.

شیوه های مهم نمونه گیری

۱)نمونه گیری تصادفی ساده: اگر همه اعضای جامعه آماری شانس مساوی برای انتخاب شدن در نمونه را داشته باشند نمونه گیری را تصادفی ساده می گوئیم. به دو صورت ممکن است انجام شود اگر عضوانتخاب شده از جامعه آماری پس از برداشت اطلاعات به جامعه برگردانده شود نمونه گیری را

تصادفی ساده با جایگذاری می گوئیم. اگر عضو انتخاب به جامعه برگردانده نشود نمونه گیری را تصادفی ساده بدون جایگذاری می نامیم.

فرض کنید اندازه جامعه $N=۳۰$ باشد و بخواهیم نمونه ای به روش تصادفی ساده به اندازه $n=۲$ از این جامعه بگیریم. یک راه ساده استفاده از ماشین حساب است اگر ابتدا کلید *SHIFT* و سپس کلید *Ran* فشرده شود عددی تصادفی در بازه (۰،۱) تولید می گردد فرض کنید این عدد ۰/۲۴۹ باشد این عدد را در N ضرب کرده قسمت صحیح آن را با یک جمع می کنیم(چرا؟) که نمونه اول عضو شماره ۸ در جامعه آماری است. عضو دوم نمونه را به این شیوه انتخاب کنید. توجه شود که نرم افزارهای آماری نظیر *SPSS* به سادگی می توانند نمونه تصادفی تولید کنند.

۲) نمونه گیری سیستماتیک: هنگامی که جامعه آماری براساس یک لیست نظیر پلاک منازل، شماره های دانشجویی منظم شده باشد می توان از این روش استفاده کرد. در این روش ابتدا دوره نمونه گیری مشخص می شود($k=N/n$) سپس از بین k عضو اول جامعه یک عضو بصورت تصادفی به عنوان مبدأ نمونه گیری انتخاب می شود(r) و برای بدست آوردن بقیه اعضا به هر عضو نمونه مقدار k اضافه می گردد. در مثال قبلی $k=۳۰/۲=۱۵$ دوره نمونه گیری است و مبدأ نمونه گیری را با ماشین حساب بصورت زیر محاسبه می شود

$$SHIFT + Rnd \rightarrow ۰/۴۳۲ \xrightarrow{\times ۱۵} ۶/۴۸$$

بنابراین مبدأ نمونه گیری عضو شماره ۷ در جامعه و عضو دوم نمونه شماره ۲۲ است. علیرغم ساده بودن این روش بدلیل اینکه فقط انتخاب عضو اول تصادفی بوده و انتخاب بقیه اعضا تصادفی نیست شیوه مناسبی برای نمونه گیری نیست.

۳) نمونه گیری طبقه ای: هنگامی که جامعه آماری براساس صفت مورد مطالعه همگن نباشد برای افزایش دقت نمونه گیری از این روش استفاده می شود. ابتدا جامعه آماری به k طبقه که اعضای درون طبقات از لحاظ متغیر مورد نظر همگن باشد افراز می شود و سپس از هر طبقه یک نمونه تصادفی گرفته می شود. برای مشخص شدن اندازه نمونه در هر طبقه نسبت نمونه را متناسب با نسبت جامعه در هر طبقه می گیریم.

$$N = N_۱ + N_۲ + \dots + N_k$$
$$n = n_۱ + n_۲ + \dots + n_k$$

بنابراین اندازه نمونه در هر طبقه

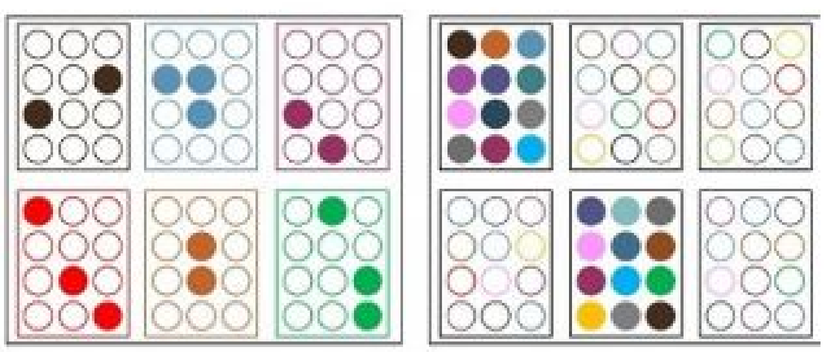
$$\frac{n_i}{n} = \frac{N_i}{N} \rightarrow n_i = n \times \frac{N_i}{N} \qquad i = ۱, \dots, k$$

مثال ٦: شرکت های موجود در یک بورس از سه صنعت می باشند می خواهیم یک نمونه ٦٠ تایی به روش طبقه ای از شرکت های این بورس انتخاب کنیم.

بنابراین بصورت تصادفی ١٨ نمونه از گروه اول، ١٥ نمونه تصادفی از صنعت دوم و ٢٧ نمونه بصورت تصادفی از گروه سوم گرفته می شود.

صنعت	تعداد هر صنعت N_i	نسبت هر طبقه در جامعه $\frac{N_i}{N}$	اندازه نمونه در هر طبقه $n_i = n \times \frac{N_i}{N}$
١	١٢٠	٠/٣٠	١٨
٢	١٠٠	٠/٢٥	١٥
٣	١٨٠	٠/٤٥	٢٧
جمع	$N = ٤٠٠$		$n = ٦٠$

٤)**نمونه گیری خوشه ای:** هنگامی که جامعه آماره وسیع یا گسترده باشد برای کاهش مشکلات اقتصادی و اجرایی از این شیوه نمونه گیری استفاده می شود. ابتدا اعضای مجاور را در یک خوشه قرار می دهیم سپس تعدادی از خوشه ها را بصورت تصادفی انتخاب کرده و سپس همه اعضاء درون این خوشه های بررسی می شوند. اگر از خوشه های انتخاب شده نمونه گیری شود نمونه گیری را دو مرحله ای می گوئیم.



شکل ١: مقایسه شیوه های نمونه گیری طبقه ای وخوشه ای

پارامتر

ویژگی مجهول جامعه آماری است که گاهی اوقات از طریق سرشماری بدست می آید. در این درس پارامترها با حروف یونانی مشخص می شود. برای مثال پارامترهای میانگین و واریانس که بترتیب نشان

دهنده متوسط و پراکندگی با حروف یونانی μ و σ^2 مشخص می شوند. همچنین فرض می کنیم پارامترهای جامعه اعداد ثابتی هستند.

آماره

تابعی از نمونه تصادفی است که دارای پارامتر مجهول نیست. در این درس آماره ها با حروف لاتین نمایش داده می شود. برای مثال آماره های میانگین و واریانس که بترتیب نشان دهنده متوسط و پراکندگی در نمونه هستند با حروف $\bar{X}$ و S^2 مشخص می شوند. چون آماره ها از طریق نمونه مشخص می شوند بنابراین متغیر بوده و دارای توزیع احتمال هستند که به آن توزیع نمونه گیری آماره گفته می شود.

برآوردگر نقطه ای

از آماره ها برای تخمین یا برآورد پارامترها جامعه استفاده می شود به همین دلیل به آنها برآوردگر نقطه ای گفته می شود. برای مثال از آماره $\bar{X}$ میانگین نمونه ای برای برآورد میانگین جامعه μ استفاده می شود. یافته برآوردگر $\bar{X}$ از نمونه را برآورد نقطه ای می گوئیم و بصورت $\bar{x}$ نمایش داده می شود.

در عمل برای برآورد پارامتر θ ممکن است برآوردگرهای $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \ldots$ موجود باشند که باید از میان آنها برآوردگر مناسب را انتخاب کنیم. به این منظور خواص زیر را در نظر می گیریم:

الف) نااریبی: برآوردگر $\hat{\theta}$ (بخوانید تتا هت) را برای برآورد پارامتر θ نااریب می گوئیم هرگاه انتظار داشته باشیم بطور متوسط مقداری که از نمونه می گیرد برابر پارامتر مجهول جامعه باشد. یعنی امید ریاضی توزیع نمونه گیری $\hat{\theta}$ برابر θ باشد

$$E(\hat{\theta}) = \theta$$

معمولاً برای برآورد یک پارامتر ممکن است چندین برآوردگر ناریب موجود باشد که باید از میان آنها یکی را انتخاب کنیم. برای مثال در نمونه گیری از توزیع نرمال میانگین نمونه ای $\bar{X}$ و میانه نمونه ای m هر دو برای برآورد میانگین جامعه μ نااریب هستند.

ب) کارآیی: فرض کنید برآوردگرهای $\hat{\theta}_1, \hat{\theta}_2$ برای برآورد پارامتر θ نااریب باشند آنگاه برآوردگر $\hat{\theta}_1$ را کارآتر از $\hat{\theta}_2$ در برآور پارامتر θ می گوئیم هرگاه $V(\hat{\theta}_1) \leq V(\hat{\theta}_2)$ باشد(یعنی واریانس کمتری داشته باشد).

برای مثال در نمونه گیری از توزیع نرمال با میانگین μ و واریانس σ^2 واریانس برآوردگر نااریب $\bar{X}$ برابر σ^2/n و برآوردگر نااریب میانه نمونه ای دارای واریانس $\pi\sigma^2/2n$ است بنابراین $V(\bar{X})<V(m)$ است.

برآوردگرهای نااریب μ و σ^2

برای داده های کمی این دو پارامتر بترتیب متوسط و پراکندگی توزیع داده ها را مشخص می کنند. فرض کنید $X_1,\dots,X_N$ اطلاعات حاصل از سرشماری جامعه باشند در اینصورت این پارامترها بصورت زیر مشخص می شوند:

$$\mu=\frac{\sum_{i=1}^{N}X_i}{N}$$

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

اگر $X_1,\dots,X_n$ اطلاعات حاصل از نمونه باشد میانگین و واریانس نمونه ای که هر دو برآوردگرهای نااریب μ و σ^2 هستند بصورت زیر هستند:

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

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

مثال ۷: در یک نمونه تصادفی ۹تایی از کارکنان یک اداره رضایت شغلی بصورت زیر بوده است میانگین و واریانس واقعی رضایت شغلی کارکنان را برآورد کنید؟

x_i:۱۲,۱۵,۱۰,۱٤,۱٦,۱۸,۱۱,۱۳,۸

حل:

x_i	۱۲	۱۵	۱۰	۱٤	۱٦	۱۸	۱۱	۱۳	۸	$\sum x_i=۱۱۷$
x_i^2	۱٤٤	۲۲۵	۱۰۰	۱۹٦	۲۵٦	۳۲٤	۱۲۱	۱٦۹	٦٤	$\sum x_i^2=۱۵۹۹$

$$\bar{x}=\frac{\sum x}{n}=\frac{۱۱۷}{۹}=۱۳$$

و

$$s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{1599 - 9(13)^2}{9-1} = \frac{78}{8} = 9/75$$

بنابراین $\hat{\mu} = 13$ و $\hat{\sigma}^2 = 9/75$.

برآوردگر ناریب نسبت

هنگامی که جامعه آماری بر اساس یک متغیر کیفی دو حالتی به دو طبقه افراز شده باشد شامل موفقیت و شکست پارامتر نسبت در جامعه بررسی می شود. فرض کنید $X_1,...,X_N$ اطلاعات حاصل از سرشماری جامعه باشند در اینصورت این پارامتر نسبت در جامعه بصورت زیر مشخص می شوند:

$$\pi = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X}{N}$$

که $X_i = 1$ در صورت که اطلاعات نمونه موفقیت باشد و اگر شکست باشد $X_i = 0$. بنابراین X تعداد موفقیت ها در جامعه است.

اگر $X_1,...,X_n$ اطلاعات حاصل از نمونه باشد نسبت نمونه ای که برآوردگر ناریب نسبت در جامعه است بصورت زیر است

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

که $\tilde{X}$ تعداد موفقیت ها در نمونه است.

مثال ۸: در یک نمونه ۱۲۰ از شرکت های بورس ۲۳ تای آنها ورشکسته بوده اند. اگر تعداد کل شرکت های این بورس ۴۷۰ شرکت باشند برآورد شرکت های ورشکسته در این بورس چه تعداد است؟

حل:

$$P = \frac{\tilde{x}}{n} = \frac{23}{120} = 0/192$$

بنابراین انتظار داریم نسبت واقعی شرکت های ورشکسته در این بورس $\hat{\pi} = 0/192$ باشد. بنابراین برای برآورد تعداد واقعی تعداد شرکت های ورشکسته خواهیم داشت:

$$\pi = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X}{N} \Rightarrow \hat{X} = N \times \hat{\pi} = ٤٧٠ \times ٠/١٩٢ = ٩٠/٢٤$$

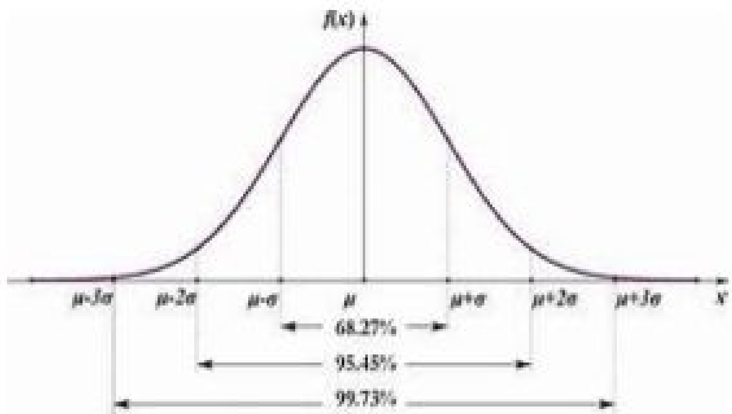
بنابراین انتظار داریم ۹۱ شرکت ورشکسته در این بورس باشد.

توزیع های نمونه گیری

۱-توزیع نرمال

متغیر تصادفی نرمال یک متغیر پیوسته می‌باشد که هر مقدار حقیقی را می تواند اختیار کند. متغیر نرمال دارای دو پارامتر است که به ترتیب میانگین μ و واریانس توزیع $\sigma^۲$ می باشند. در بسیاری از استنباط های آماری فرض می شود که توزیع جامعه نمونه گیری شده دارای توزیع نرمال است. استاندارد شده متغیر نرمال دارای توزیع نرمال است که با آن توزیع نرمال استاندارد گفته می شود.

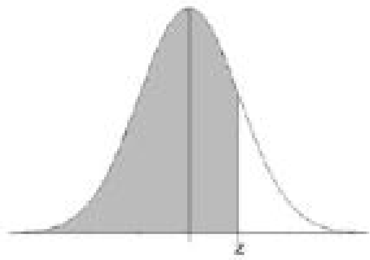
$$Z = \frac{X - \mu}{\sigma} \sim N(٠,١) \qquad \mu_Z = ٠, \quad \sigma_Z^۲ = ١$$



شکل ۱: منحنی تابع احتمال نرمال

در شکل ۱ تابع احتمال نرمال نشان داده شده است اگر در این شکل میانگین ۰ و انحراف معیار یک قرار داده شود بنابراین شکل توزیع نرمال استاندارد را داریم. احتمـالات تجمعـی نرمـال اسـتاندارد از طریـق جدول محاسبه شوند.

مثال۹: احتمال $P(Z < ١/٩٦)$ و $P(Z < -١/٦٤)$ زیر را بدست آورید؟



الف) $P(Z < ۱/۹۶) = ۰/۹۷۵$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

احتمال تجمعی توزیع نرمال استاندارد در نقطه ۱/۹۶ برابر ۰/۹۷۵ است($Z_{۰/۹۷۵} = ۱/۹۶$). همچنین

$Z_{۰/۰۲۵} = -۱/۹۶$

ب) $P(Z < -۱/۶۴) = ۰/۰۵۰۵$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

در نمونه گیری از توزیع نرمال میانگین نمونه ای دارای توزیع نرمال است یعنی

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(۰,۱)$$

برای مثال اگر یک نمونه ۳۶ تایی از توزیع نرمال با میانگین ۱۳ واریانس معلوم ۱۶ داشته باشیم احتمال اینکه میانگین نمونه کمتر ۱۲ باشد برابر است با

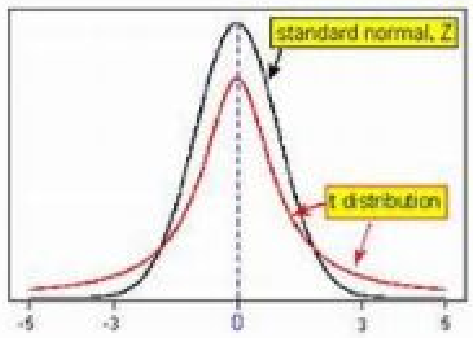
$P(\overline{X} < ۱۲) = P(Z < -۱/۵) = ۰/۰۶۶۸$

نکته: اگر توزیع جامعه نمونه گیری شده معلوم نباشد اگر اندازه نمونه $n \geq ۳۰$ باشد توزیع میانگین نمونه ای نرمال است.

۲-توزیع t استیودنت

شکلی شبیه توزیع نرمال استاندارد اما پراکندگی بیشتری نسبت به شکل توزیع نرمال دارد دارای یک پارامتر است که با آن درجه آزادی گفته می شود هنگامی که درجه آزادی توزیع t به سمت بینهایت میل

کند توزیع تقریبی Z را خواهد داشت. برای تقریب به عنوان یک قاعده سرانگشتی حداقل درجـه آزادی را ۳۰ فرض می شود.

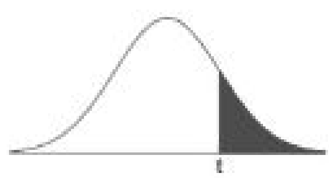


شکل: مقایسه منحنی نرمال استاندارد و توزیع t

مثال۱۰: احتمال های زیر را در توزیع t بدست آورید؟

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
∞	1.282	1.645	1.960	2.326	2.576

الف) $t_{/۰۲۵}(۹) = ۲/۲۶۲$



ب) $t_{/۰۲۵}(۵۸) \approx Z_{/۰۲۵} = ۱/۹۶$

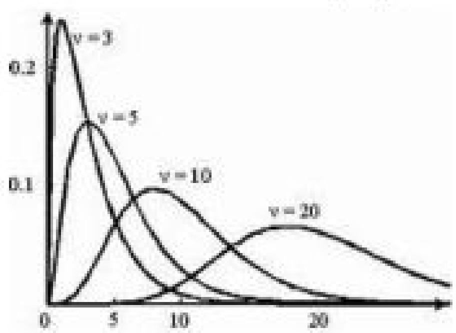
همانطور که از جدول مشاهده می شود در توزیع t با ۶ درجه آزادی ۹۹ درصد سطح زیر منحنی احتمال در بازه $(-۳/۷۰۷,۳/۷۰۷)$ است در حالیکه در توزیع نرمال این بازه $(-۲/۵۷۶,۲/۵۷۶)$ است.

نکته: در نمونه گیری از توزیع نرمال با واریانس مجهول $\sigma^۲$ متغیر زیر دارای توزیع t است

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

۳-توزیع کای دو

متغیر تصادفی کای دو یک متغیر پیوسته است که مقادیر نامنفی را می‌تواند داشته باشد یک پارامتر دارد که با آن درجه آزادی($df = v$) گفته می شود.

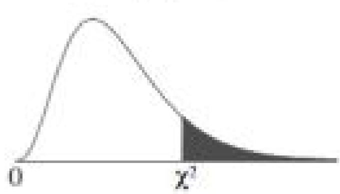


شکل۳: منحنی تابع احتمال کای دو با درجه آزادی های مختلف

میانگین $E(X) = v$ و واریانس توزیع $V(X) = ۲v$ هستند.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672

در جدول بالا $\chi^2_{۰/۰۲۵}(۱۰) = ۲۰/۴۸۳$ و $\chi^2_{۰/۹۷۵}(۱۰) = ۳/۲۴۷$ هستند.



نکته: در نمونه گیری از توزیع نرمال با واریانس مجهول σ^2 متغیر زیر دارای توزیع χ^2 است

$$\chi^2 = \frac{(n-۱)S^2}{\sigma^2} \sim \chi^2(n-۱)$$

که n اندازه نمونه است.

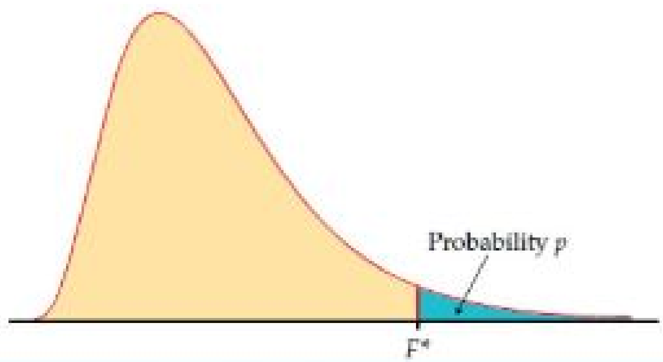
٤-توزیع F

متغیر تصادفی F مانند متغیر کای دو مقادیر نامنفی را می گیرد این متغیر حاصل تقسیم دو متغیر کای دو مستقل می باشد که بر درجه آزادی شان تقسیم شده باشند.یعنی اگر $X \sim \chi^2(v_1)$ و $Y \sim \chi^2(v_2)$ دو متغیر کای دومستقل باشند

$$F = \frac{X/v_1}{Y/v_2} = \frac{v_2 X}{v_1 Y} \sim F(v_1, v_2)$$

پارامترهای توزیع درجه آزادی صورت v_1 و مخرج v_2 هستند. در محاسبه احتمال ها به نکته زیر دقت می شود:

$$F_{\alpha}(v_1, v_2) = \frac{1}{F_{1-\alpha}(v_2, v_1)}$$



		Degrees of freedom in the numerator								
	p	1	2	3	4	5	6	7	8	9
Degrees of freedom in the denominator 3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24

با استفاده از جدول توزیع $F_{0.025}(۳,۵) = ۷/۷۶$ است و $F_{0.9}(۵,۴) = ۰/۲۸۴$ است(چرا؟).

نکته: فرض شود یک نمونه تصادفی n_1 تایی از یک جامعه نرمال با میانگین و واریانس μ_1 و σ_1^2 داشته باشیم همچنین یک نمونه تصادفی n_2 تایی از جامعه نرمال دیگر که مستقل از جامعه اول باشد با میانگین و واریانس μ_2 و σ_2^2 داشته باشیم آنگاه

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

که S_1^2 و S_2^2 واریانس نمونه ای یا برآوردگرهای نااریب σ_1^2 و σ_2^2 هستند.