بسم الله الرحمن الرحیم

موسسه آموزش عالی آزاد
ماهان
www.mahan.ac.ir

اقتصادسنجی

مجموعه علوم اقتصادی

دکتری

مؤلف: محیا مومنی

مومنی، محیا ۱۳۶۹

اقتصادسنجی رشته علوم اقتصادی / محیا مومنی

مشاوران صعود ماهان: ۱۴۰۱

۲۹۷ ص: جدول، نمودار (آمادگی آزمون دکتری علوم اقتصادی)

ISBN: 978-600-458-924-6

فهرستنویسی بر اساس اطلاعات فیپا.

فارسی ـ چاپ اول ـ ویرایش اول

۱-اقتصادسنجی

محیا مومنی

ج ـ عنوان

انتشارات مشاوران صعود ماهان

موسسه آموزش عالی آزاد ماهان

www.mahan.ac.ir

- **نام کتاب:** .... اقتصادسنجی
- **مدیران مسئول:** .... هادی و مجید سیّاری
- **مولف:** .... **محیا مؤمنی**
- **مسئول برنامه‌ریزی و تولید محتوا:** .... سمیه بیگی
- **ناشر:** .... مشاوران صعود ماهان
- **نوبت و تاریخ چاپ:** .... ویرایش اول /۱۴۰۱
- **تیراژ:** .... ۱۰۰۰ نسخه
- **قیمت:** ۲/۳۹۰/۰۰۰ریال
- **شابک:** ISBN۹۷۸-۶۰۰-۴۵۸-۹۲۶-۶

انتشارات مشاوران صعود ماهان:تهران ـ خیابان ولیعصر، بالاتر از تقاطع ولیعصر مطهری، پلاک ۲۰۵۰

تلفن: ۸۸۱۰۰۱۱۳ و ۸۸۴۰۱۳۱۳

مقدمه ناشر

بنام خدا

ایمان داریم که هر تغییر و تحول بزرگی در مسیر زندگی بدون تحول معرفت و نگرش میسر نخواهد بود. پس بیایید با اندیشه توکل، تفکر، تلاش و تحمل در توسعه دنیای فکریمان برای نیل به آرامش و آسایش توأمان اولین گام را برداریم. چون همگی یقین داریم دانایی، توانایی می‌آورد.

شاد باشید و دلی را شاد کنید

برادران سیاری

مقدمه مولف

اقتصادسنجی شاخه‌ای از علم اقتصاد است که به اندازه‌گیری و تجزیه و تحلیل روابط بین متغیّرهای اقتصادی می‌پردازد.اقتصادسنجی با مطالعه نظام‌مند پدیده‌های اقتصادی با استفاده از داده‌های مشاهده شده سر و کار دارد به عبارتی، اقتصادسنجی علم تحلیل‌های آماری از مدل‌های اقتصادی است. از آنجا که مباحث اقتصادسنجی وسیع و گسترده می‌باشد و در عین حال همواره در حال تحول و پیشرفت می‌باشد، جمع‌آوری تمام آنچه به عنوان اقتصادسنجی شناخته می‌شود در یک متن درسی غیر ممکن به نظر می‌رسد. آنچه مشخص است، آن است که بر اساس آنچه مورد علاقه مولف می‌باشد، نوع ارائه مطالب و بیان آنها پوشش داده می‌شود. و همین موضوع انتخاب مطالبی را که لازم است برای آمادگی در آزمون مورد مطالعه قرار گیرند با مشکل روبرو کرده است.کتاب فوق در مجالی اندک و با هدف آمادگی داوطلبان برای شرکت در آزمون دکتری آماده شده است و سعی شده است که مطالب جامعی از اقتصادسنجی ارائه شود. هدف اصلی کتاب فوق، ارائه مطالب مهم به منظور آمادگی در آزمون دکتری می‌باشد. این کتاب در هفت فصل اصلی طبقه‌بندی شده است. در فصل اول آمار در فصل دوم رگرسیون کلاسیک خطی دو متغیره مورد بررسی قرار گرفته است و مفاهیم اساسی آن ارائه شده است. فصل سوم به گسترش مدل رگرسیون کلاسیک و الگوی چند متغیره رگرسیون اختصاص دارد در فصل چهارم موارد نقض فروض کلاسیک رگرسیون، نتایج نقض فروض و همچنین راههای تشخیص و رفع آنها ارائه شده‌اند. . در فصل پنجم دو روش دیگر برآورد پارامترهای مدل رگرسیون یعنی روشهای حداکثر درستنمایی و حداقل مربعات تعمیم یافته معرفی شده‌اند. در فصل ششم الگوهای سری زمانی در فصل هفتم به مبحث سیستم معادلات همزمان و روشهای برآورد آن اختصاص دارد. معرفی شده‌اند. در انتها مجموعه سؤالات دکتری آزاد با حل آنها جهت تقویت بیشتر دانشجویان با حل عددی و تشریحی بیان شده است.

حسین نصیری

فهرست صفحه

۲) فصل سوم : مدل رگرسیون خطی کلاسیک چند متغیره

فصل چهارم: آزمونهای فروض کلاسیک و نقض آنها

آمار توصیفی

مفاهیم اساسی آمار: علم آمار نیز مانند سایر علوم بر پایه یک سری از مفاهیم بنا شده است که به مفاهیم اولیه معروفند و بدون تعریف، پذیرفته شده و بصورت زیر توصیف می‌شوند:

– **جامعه آماری:** به کلیه عناصر، افراد و یا اشیاء که حداقل در یک خاصیت مشترک باشند، جامعه آماری گویند.

– **صفت مشخصه:** به خاصیت یا خواصی که در همه عناصر، افراد یا اشیاء مشترک است، صفت مشخصه گفته می‌شود.

– **صفت متغیر:** خاصیت یا خواصی که از یک عنصر به عنصر دیگر، از یک فرد به فرد دیگر جامعه تغییر کند، صفت متغیر نامیده می‌شود.

صفاتی را که بتوان برای آنها اندازه تعریف کرد، صفات متغیر کمی و صفاتی را که نتوان برای آنها اندازه‌ای تعریف کرد، صفات کیفی نامند.

صفات متغیر کمی یا گسسته (نا پیوسته) هستند یا پیوسته.

– **طبقه‌بندی صفات:** بسته به اینکه صفت متغیر، گسسته یا پیوسته باشد، آنها را با توجه به ماهیتشان طبقه‌بندی می‌کنند.

در حالت گسسته آنها را به ترتیب صعودی یا نزولی مرتب می‌کنند و توزیع صفت را بر حسب فراوانی مطلق بدست می‌آورند.

در حالت پیوسته بودن صفت، آنها را در فاصله طبقه‌بندی کرده و توزیع صفت متغیر را بر حسب فراوانی مطلق بدست می‌آورند.

– **نمودارها**: برای بیان چگونگی توزیع صفت متغیر، می‌توان از نمودارها استفاده کرد. برای صفات متغیر نا پیوسته از نمودارهای میله‌ای و چند گوش (پلی‌گون) استفاده می‌شود.

و برای صفات متغیر پیوسته از نمودار هیستوگرام (بافت نگار) استفاده می‌شود.

و برای صفات کیفی از نمودارهای میله‌ای یا پلی‌گون و یا دایره‌ای استفاده می‌شود.

– **فراوانی تجمعی(انباشته):** فراوانی تجمعی مربوط به x_i برابر فراوانی تمامی مقادیر صفت x را گویند که از x_i تجاوز نکند:

$F(x_i) = n(x \le x_i)$

– **فراوانی تجمعی نسبی:** فراوانی تجمعی نسبی از جمع فراوانیهای نسبی صفت x تا نقطه x_i را فراوانی تجمعی نسبی در نقطه x_i نامند:

$$F'(x_i) = f(x \le x_i) = \frac{n(x \le x_i)}{n}$$

– **مفهوم اندازه‌گیری و مقیاس آنها**

در اجرای یک تحقیق آماری، اولین مرحله، مشاهده داده‌های آماری می‌باشد و آنهم اندازه‌گیری صفات در جامعه است.

«نسبت دادن اعداد به عناصر یا افراد جامعه بر طبق قواعد معینی، اندازه‌گیری نامیده می‌شود»

۲- مشخص‌کننده‌های عددی(آماری)

مشخص‌کننده‌های عددی، کمیتهائی (پارامترهائی) هستند که برای مقایسه بین چند جامعه بکار می‌روند.
و در سه گروه زیر طبقه‌بندی می‌شوند:

۱-۱- مشخص‌کننده‌های مرکزی: که مرکز تمرکز توزیع صفت را بیان می‌کنند و از واحد اندازه‌گیری صفت متغیر تبعیت دارند. مهمترین آنها عبارتند از:

نکته: میانگین حسابی یا متوسط حسابی: که از مجموع متغیرها به حجم جامعه به دست می‌آید:

$$\bar{x} = \frac{\sum x_i}{n}$$

در صورتیکه جدول بر حسب فراوانی مطلق یا فراوانی نسبی باشد، میانگین حسابی بصورت:

$$\bar{x} = \frac{\sum n_i x_i}{n} \qquad \text{یا} \qquad \bar{x} = \sum f_i x_i$$

محاسبه میشود که به میانگین وزنی معروفند. که دراین رابطه $f_i = \frac{n_i}{n}$ فراوانی نسبی صفت متغیر می‌باشد.

نکته: خواص میانگین:

۱- میانگین اعداد ثابت، خود عدد ثابت است: $\frac{\sum a}{n} = a$

۲- اگر از متغیرها، عدد ثابتی مانند a را کم یا اضافه کنیم از میانگین هم به اندازه‌a کم یا اضافه خواهد شد.

$$\frac{\sum (x_i \pm a)}{n} = \bar{x} \pm a$$

۳- اگرمتغیرها را در عدد ثابتی مانند k ضرب یا تقسیم کنیم، میانگین هم در عدد ثابت k ضرب و یا بر عدد ثابت k تقسیم می‌شود:

$$\frac{\sum (kx_i)}{n} = k\bar{x} \qquad \frac{\sum \left(\frac{x_i}{k}\right)}{n} = \frac{\bar{x}}{k}$$

۴- مجموع انحرافات متغیرها از میانگین مساوی صفر است:

$$\sum (x_i - \bar{x}) = 0 \qquad \sum n_i (x_i - \bar{x}) = 0 \qquad \sum f_i (x_i - \bar{x}) = 0$$

نکته: میانگین هندسی: اگر متغیرها بصورت نسبت یا درصد و یا شاخص بیان شده باشند، در این صورت، متوسط متغیرها بصورت میانگین هندسی محاسبه می‌شود.

$$\bar{x}_G = \sqrt[n]{x_1, x_2, x_3, \ldots, x_n} \qquad (x_i > 0)$$

نکته: میانگین همساز (هارمونیک): اگر صفت متغیر در ارتباط با زمان باشد و از دو واحد اندازه‌گیری تبعیت کند، دراین صورت برای محاسبه میانگین صفت متغیر، از میانگین همساز استفاده می‌شود:

$$\bar{x}_H = \frac{n}{\sum \frac{1}{x_i}} \qquad \text{یا} \qquad \bar{x}_H = \frac{n}{\sum n_i \frac{1}{x_i}}$$

همواره بین میانگین حسابی، میانگین هندسی و میانگین هارمونیک رابطهٔ زیر برقرار است:

$$\bar{x}_H < \bar{x}_G < \bar{x}_M$$

۲-۱- مشخص کننده‌های پراکندگی

نکته: دامنه تغییرات یا فاصله تغییرات(Range): اختلاف بزرگترین متغیر و کوچکترین متغیر را دامنه تغییرات نامند:

$R = X_{max} - X_{min}$

- انحراف چارکی(نیم دامنه چارکی): برای اندازه پراکندگی توزیعهای باز(توزیع‌هائی که حد پایین یا حد بالای آنها نامعلوم است). محاسبه می‌شود:

$$Q = \frac{Q_3 - Q_1}{2}$$

نکته: میانگین قدر مطلق انحرافات (انحراف متوسط): از مجموع قدر مطلق اختلاف متغیرها از میانگین تقسیم بر n (حجم جامعه) بدست می‌آید:

$$M.D = \frac{\sum |x_i - \bar{x}|}{n} \quad \text{یا} \quad M.D = \frac{\sum n_i |x_i - \bar{x}|}{n}$$

نکته: واریانس: میانگین مجذور انحرافات متغیرها از میانگین را واریانس می‌نامند:

$$D(x) = \sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2$$

در صورتیکه توزیع صفت متغیر بر حسب فراوانی بدست آمده باشد، واریانس مساوی خواهد شد با:

$$D(x) = \sigma^2 = \frac{\sum n_i (x_i - \bar{x})^2}{n} \quad \frac{\sum n_i x_i^2}{n} - \bar{x}^2$$

واریانس با مجذور واحد اندازه‌گیری صفت متغیر بیان می‌شود، به همین جهت از آن جذر گرفته و به‌عنوان معیار پراکندگی مناسب بدست می‌آورند.

نکته: انحراف معیار(انحراف استاندارد): از جذر واریانس بدست می‌آید و هم بعد صفت متغیر می‌باشد:

$$\sigma = +\sqrt{D(x)}$$

نکته:خواص واریانس:

۱- واریانس عدد ثابت مساوی با صفر است. $D(a) = 0$

۲- اگر از متغیرها عدد ثابتی مانند α را کم و یا به متغیرها عدد ثابتی مانند a را اضافه کنیم، در مقدار واریانس تأثیری نخواهد داشت: $D(X \pm a) = D(x)$

۳- اگر متغیرها را در عدد ثابتی مانند k ضرب کنیم، واریانس در k^2 ضرب خواهد شد: $D(kx) = k^2 D(X)$

۴- اگر متغیرها را بر عدد ثابتی مانند K تقسیم کنیم، واریانس بر k^2 تقسیم خواهد شد:

$$D\left(\frac{X}{K}\right) = \frac{1}{K^2} D(X)$$

۱-۳- مشخص‌کننده‌های نسبی پراکندگی

- ضریب تغییرات: از خارج قسمت انحراف معیار (σ) به میانگین $(\bar{X})$ بدست می‌آید و از واحد اندازه‌گیری مجرد است:

$$C.V = \frac{\sigma}{\bar{X}} \times 100$$

ضریب تغییرات، درصد تغییرات متغیرها را حول میانگین نشان می‌دهد.

- ضریب چولگی: با استفاده از روابط بین سه مشخص کننده مرکزی، ضریب چولگی پیرسن محاسبه می‌شود که از دقت بالائی برخوردار نیست:

$$A_s = \frac{\bar{X} - MO}{\sigma}$$

در صورتیکه متغیر دارای چولگی ملایم باشد، همواره رابطه زیر بین سه مشخصه مرکزی برقرار است:

$$\overline{X}-MO=۳(\overline{X}-Me)$$

ولی برای محاسبه دقیق‌تر، ضریب چولگی از خارج قسمت گشتاور مرکزی مرتبه سوم $(\mu_۳)$ به $\sigma^۳$ استفاده می‌شود:

$$A=\frac{\mu_۳}{\sigma^۳}$$

که درآن $\mu_۳$ بصورت زیر محاسبه می‌شود:

$$\mu_۳=\frac{\sum(x_i-\overline{x})^۳}{n} \quad \text{یا} \quad \mu_۳=\frac{\sum n_i(x_i-\overline{x})^۳}{n}$$

ضریب چولگی ممکن است مثبت و یا منفی بدست آید. اگر مثبت باشد، توزیع دارای چولگی راست و در صورتی که منفی باشد، توزیع دارای چولگی چپ می‌باشد. اگر $A=۰$ گردد، توزیع متقارن است. اگر $|A|<۰/۱$ باشد، توزیع تقریباً قرینه است و اگر $۰/۵\le|A|\le۰/۱$باشد، توزیع دارای چولگی ملایم و اگر $|A|>۰/۵$ باشد، توزیع شدیداً چولگی دارد.

- **ضریب کشیدگی:** از خارج قسمت گشتاور مرکزی مرتبه چهارم به $\sigma^۴$ بدست می‌آید و معمولاً کشیدگی توزیعها را با توزیع نرمال مقایسه می‌کنند و از طرفی ضریب کشیدگی توزیع نرمال مساوی با ۳ می‌باشد، لذا ضریب کشیدگی بصورت:

$$K=\frac{\mu_۴}{\sigma^۴}-۳$$

محاسبه می‌شود. اگر $K=۰$ باشد، کشیدگی توزیع به اندازه نرمال است و اگر $K<۰$ باشد، توزیع از نرمال کوتاهتر است. در صورتیکه $K>۰$ باشد، توزیع از توزیع نرمال کشیده‌تر است. گشتاور مرکزی مرتبه چهارم بصورت زیر بدست می‌آید:

$$\mu_۴=\frac{\sum(x_i-\overline{x})^۴}{n} \quad \text{یا} \quad \mu_۴=\frac{\sum n_i(x_i-\overline{x})^۴}{n}$$

- **گشتاورها:** گشتاورها، میانگین‌هائی هستند حول مرکز ثقل توزیع با مراتب مختلف.

چون مرکز ثقل توزیع می‌تواند مبدأ مختصات(صفر) و یا میانگین حسابی باشد، لذا دو نوع گشتاور دراینجا تعریف می‌کنیم:

۱- گشتاورهای اولیه: یا گشتاورهای نسبت به‌مبدأ صفر بصورت زیر محاسبه می‌شوند:

$$m_h=\frac{\sum x^h}{n} \quad \text{یا} \quad m_h=\frac{\sum n_i x_i^h}{n} \quad \text{یا} \quad m_h=\sum f_i x_i^h$$

و گشتاورهای مرتبه اول تا چهارم نسبت به‌مبدأ صفر مساوی است با:

$$m_۱=\frac{\sum x_i}{n}=\overline{x} \qquad m_۲=\frac{\sum x^۲}{n}=\overline{x}^۲$$

$$m_۳=\frac{\sum x_i^۳}{n}=\overline{x}^۳ \qquad m_۴=\frac{\sum x_i^۴}{n}=\overline{x}^۴$$

که به گشتاورهای توانی یا میانگینهای توانی معروفند.

۲- گشتاورهای مرکزی: یا گشتاورهای نسبت به‌مبدأ میانگین که به صورت زیر محاسبه می‌شوند:

$$\mu_h=\frac{\sum(x_i-\overline{x})^h}{n} \quad \text{یا} \quad \mu_h=\frac{\sum n_i(x_i-\overline{x})^h}{n} \quad \text{یا} \quad \mu_h=\sum f_i(x_i-\overline{x})^h$$

و گشتاورهای مرتبه اول تا چهارم نسبت به مبدأ میانگین بصورت زیر بدست می‌آیند:

$$\mu_۱=\frac{\sum(x-\overline{x})}{n}=۰$$

$$\mu_2 = \frac{\sum (x_i - \bar{x})^2}{n} = D(x)$$
$$\mu_3 = \frac{\sum (x_i - \bar{x})^3}{n}$$
$$\mu_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$

- رابطه تبدیل گشتاورهای اولیه به گشتاورهای مرکزی: برای تبدیل گشتاورهای اولیه به گشتاورهای مرکزی ثابت می‌شود که $\mu_h = \frac{\sum (x_i - \bar{x})^h}{n} = (m - m_1)^h$ و در نتیجه برای $h = 1,2,3,4$ روابط زیر بدست می‌آید:

$$\mu_1 = (m_0 - m_1) = 0$$
$$\mu_2 = (m - m_1)^2 = m_2 - m_1^2$$
$$\mu_3 = (m - m_1)^3 = m_3 - 3m_2 m_1 + 2m_1^3$$
$$\mu_4 = (m - m_1)^4 = m_4 - 4m_3 m_1 + 6m_2 m_1^2 - 3m_1^4$$

۳-توزیعهای دو بعدی

- دستگاه دو صفت متغیر: نتایج مشاهدات بر روی دو صفت متغیر y, x با هم مجموعه‌ای از زوجهای مرتب شده را تشکیل می‌دهند که دنباله زوجهای مرتب شده را سری همبستگی می‌نامند.

اگر توزیع فراوانیهای دو صفت را با هم در جدول نشان دهیم، جدول زیر بدست می‌آید:

$x \backslash y$	y_1	y_2	...	y_t	n_{i0}
x_1	n_{11}	n_{12}	...	n_{1t}	n_{10}
x_2	n_{21}	n_{22}	..	n_{2t}	n_{20}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_s	n_{s1}	n_{s2}	...	n_{st}	n_{s0}
n_{0j}	n_{01}	n_{02}	...	n_{0t}	n

اگر فراوانیهای توزیعهای دو بعدی را سطری جمع کنیم توزیع حاشیه‌ای صفت x و در صورتیکه فراوانیها را ستونی با یکدیگر جمع کنیم، توزیع حاشیه‌ای برای صفت y بدست می‌آید.

- **توزیع‌های شرطی:** چون به تعداد سطرها توزیع تکی برای صفت y و به تعداد ستونها، توزیع تکی برای صفت x در توزیعهای دو بعدی وجود دارد و هر یک از آنها به ازاء مقدار معینی از صفت x و با مقدار معینی از صفت y بوجود می‌آیند، آنها را توزیعهای شرطی می‌نامند.

- **مفهوم بستگی آماری:** رابطه بین صفت متغیر که از افزایش یا کاهش یکی از صفت‌ها و تغییرات توزیع‌های شرطی صفت دیگر بوجودمی‌آید، بستگی آماری نامیده می‌شود. در نتیجه هرگاه با تغییر یکی از صفتهای متغیر، توزیع‌های شرطی صفت متغیر دیگر تغییر کند، گویند بستگی آماری بین دو صفت وجود دارد.

نکته: همبستگی: هرگاه به ازاء افزایش یا کاهش یکی از صفت‌های متغیر، میانگین توزیع‌های شرطی صفت متغیر دیگر تغییر کند، بین دو صفت متغیر x و y همبستگی وجود خواهد داشت.

نکته: چون در مفهوم همبستگی، دو صفت x و y هم ارز نمی‌باشند، لذا:

الف- همبستگی صفت y بر حسب x را رگرسیون y روی x می‌نامند و بصورت $\bar{y}_x = f(x)$ نوشته می‌شود.

ب- همبستگی صفت x بر حسب y را رگرسیون x روی y می‌نامند و بصورت $\bar{x}_y = g(y)$ می‌نویسند. بنابراین اگر با تغییر صفت x، میانگین توزیعهای شرطی $y\,(\bar{y}_i)$ تغییر کنند، همبستگی y روی x وجود دارد و هم چنین اگر با تغییر صفت y میانگین توزیعهای شرطی $x\,(\bar{x}_j)$ تغییر کنند، دلیل بر وجود همبستگی یا رگرسیون x روی y خواهد بود.

اگر زوجهای مرتب شده $(x_i, \bar{y}_i)$ و یا $(\bar{x}_j, y_j)$ را در صفحه مختصات مشخص کنیم، از طرز قرار گرفتن نقاط در صفحه مختصات، نوع معادله‌ای که بیان کننده رابطه بین دو صفت(نوع معادله رگرسیون) می‌باشد، بدست می‌آید.

در صورتیکه رابطه رگرسیون به صورت معادله خطی مستقیم باشد، یعنی:

$$\bar{y}_x = f(x) = a + bx \quad \text{یا} \quad \bar{x}_y = g(y) = a' + b'y$$

درآن صورت ضرایب آن با استفاده از روش کمترین توانهای دوم بدست می‌آید:

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{sp_{xy}}{ss_x} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^{۲}} = \frac{\sum xy - \frac{\sum x.\sum y}{n}}{\sum x^{۲} - \frac{\left(\sum x\right)^{۲}}{n}} = \frac{cov(x,y)}{D(x)}$$

– **خطای معادله‌رگرسیون:** اگر معادله خط همبستگی(معادله خط رگرسیون) بصورت $\underset{x}{\bar{y}} = a + bx$ بیان شده باشد، آنگاه خطای معادله رگرسیون $(\sigma_{y.x})$ از جذر واریانس همبستگی یا واریانس رگرسیون بدست می‌آید:

$$\sigma^{۲}_{y.x} = \frac{\sum y^{۲} - a\sum y - b\sum xy}{n}$$

$$\sigma_{y.x} = \sqrt{\frac{\sum y^{۲} - a\sum y - b\sum xy}{n}}$$

از خطای معادله رگرسیون برای ارزیابی صفت y به ازاء یک مقدار معین از صفت x استفاده می‌کنند.

$$\underset{x_i}{\bar{y}} - \sigma_{y.x} < y < \bar{y}_{xi} + \sigma_{yx}$$

– **ضریب همبستگی:** به منظور تعیین شدت همبستگی بین دو صفت x و y از ضریب همبستگی استفاده می‌شود که: اولاً ضریب همبستگی بعد ندارد یعنی از هیچ واحد اندازه‌گیری تبعیت نمی‌کند، ثانیاً حدود تغییرات ضریب همبستگی همواره بین $(-۱,+۱)$ تغییر می‌کند.

نکته: اندازه ضریب همبستگی به توسط یکی از روابط زیر بدست می‌آید.

$$\rho = \frac{cov(x,y)}{\sigma_x.\sigma_y} = \frac{sp_{xy}}{\sqrt{ss_x.ss_y}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^{۲}.\sum(y-\bar{y})^{۲}}} = \sqrt{b.b'}$$

اگر $\rho = ۰$ باشد، همبستگی خطی بین دو صفت وجود ندارد.

اگر $\rho = +۱$ باشد، همبستگی کامل و مستقیم است.

اگر $\rho = -۱$ باشد، همبستگی کامل و معکوس است.

اگر $-۱ < \rho < +۱$ باشد، همبستگی را ضعیف و یا شدید می‌نامند.

– **ضریب تعیین یا ضریب تشخیص:** از خارج قسمت واریانس معادله رگرسیون $(\sigma^{۲}_{y.x})$ به واریانس y یعنی $\sigma^{۲}_y$ در معادله رگرسیون y بر حسب X بدست می‌آید:

$$R^{۲} = \frac{\sigma^{۲}_{y.x}}{\sigma^{۲}_y} = \frac{[cov(x,y)]^{۲}}{\sigma^{۲}_x.\sigma^{۲}_y} = \rho^{۲}$$

که در این رابطه $cov(x,y)$، کواریانس نامیده می‌شود که پراکندگی توأم دو متغیر y, x را نشان می‌دهد و به نام گشتاور همبستگی نیز موسوم است و به صورت زیر محاسبه می‌شود.

$$cov(x,y)=\frac{\sum(x-\bar{x})(y-\bar{y})}{n}=\frac{\sum xy}{n}-\overline{xy}$$

ضریب تعیین، میزان تغییرپذیری مقادیر مشاهده شده صفت y را حول معادله رگرسیون به میزان تغییرپذیری صفت x نشان می‌دهد. یعنی اگر $\rho^۲$ را در ۱۰۰ ضرب کنیم و بر حسب درصد بیان نمائیم، آنگاه $R^۲$ درصد تغییرات صفت y را ناشی از اثرات صفت متغیر x نشان می‌دهد.

نظریه احتمال

– **مفاهیم اساسی:** نظریه احتمال بر پایه دو مفهوم حادثه و آزمایش استوار است و بعنوان مفاهیم اولیه بکار می‌روند، به همین دلیل آنها را تعریف نکرده، بلکه بصورت زیر توصیف می‌کنند:

حادثه: هرنمود یا پدیده‌ای که درنتیجه آزمایش، وقوع یا عدم وقوع آن بتواند مورد بحث قرار گیرد، حادثه نامیده میشود.

آزمایش: برقرار کردن مجموعه شرطهای معین که بطور مکرر و به تعداد دلخواه زیادی عملی گردد، آزمایش نامند.

حادثه درنتیجه آزمایش به یکی از سه صورت زیر اتفاق می‌افتد:

۱- حادثه یقین　　　۲- حادثه غیر ممکن　　　۳- حادثه تصادفی

– **تعریف احتمال حادثه:** اندازه امکان وقوع هر حادثه را احتمال آن حادثه می‌نامند و با $P(A)$ نشان می‌دهند.

– **فضای نمونه:** مجموعه‌ای از عناصر دلخواه را فضای نمونه می‌نامند و این عناصر دلخواه، تمامی نتایج ممکن آزمایش هستند یا بعبارتی مجموعه تمامی حوادث ساده را در یک آزمایش فضای نمونه می‌نامند.

اندازه امکان وقوع حادثه به یکی از سه صورت زیر محاسبه می‌گردد.

– محاسبه احتمال بر اساس تعریف کلاسیک احتمال: تعریف کلاسیک احتمال بر پایه هم احتمال بودن حوادث و یا هم امکان بودن حوادث بصورت زیر بیان می‌شود:

احتمال حادثه A از خارج قسمت تعداد حالات مساعد بر حادثه A به تعداد تمامی نتایج ممکن آزمایش بدست می‌آید:

$$P(A)=\frac{\text{تعداد حالات مساعد بر حادثه } A}{\text{تعداد تمامی نتایج ممکن}}$$

– محاسبه احتمال بر اساس تعریف آماری احتمال- تعریف آماری احتمال برای حوادثی که هم احتمال نباشند بکار می‌رود و از خاصیت فراوانی نسبی حادثه استفاده می‌شود، زیرا فراوانی نسبی را فقط می‌توان پس از انجام آزمایش‌ها تعیین کرد، بنابراین در تکرار زیاد آزمایشها، فراوانی نسبی حادثه به سمت عددی میل می‌کند که آن عدد، اندازه امکان وقوع حادثه را بیان می‌کند و احتمال آن حادثه نامیده می‌شود:

$$P(A)=\underset{n\to\infty}{f}(A)=\frac{m}{n}$$

این تعریف بر اساس قانون اعداد بزرگ ضعیف بصورت برنولی، بیان می‌شود:

$$\lim_{n\to\infty} p\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=۱$$

یعنی یقیناً در تکرار زیاد آزمایش‌ها، فراوانی نسبی حادثه به سمت احتمال حادثه میل می‌کند.

و بر همین اساس قانون اعداد بزرگ قوی بر نولی بصورت

$$p\left(\lim_{n\to\infty}\frac{m}{n}=p\right)=۱$$

تعریف می‌شود، یعنی هرگاه حد فراوانی نسبی حادثه برابر p گردد، یک حادثه یقین است.

محاسبه احتمال بر اساس تعریف احتمال هندسی: اگر تعداد وقوع حوادث در نتیجه آزمایش شمارش ناپذیر باشد، در این صورت خارج قسمت اندازه امکان وقوع حادثه مورد نظر به اندازه امکان نتایج کل آزمایش را احتمال هندسی می‌نامند.

$$P(A)=\frac{mes(A)}{mes(S)}=\frac{(A)}{(S)}\quad \frac{\text{اندازه امکان وقوع حادثه}}{\text{اندازه امکان وقوع کل حادثه}}$$

نکته: اصول متعارف احتمال

۱- احتمال حادثه بین صفر تا ۱ تغییر می‌کند:

$$0\le P(A)\le 1$$

۲- مجموع احتمالها در فضای نمونه (گروه کامل حوادث) مساوی یک است:

$$\sum_{i=1}^{n}P(A_i)=1$$

۳- اگر دو حادثه A و B نسبت به یکدیگر ناسازگار باشند، در این صورت، احتمال اجتماع دو حادثه مساوی است با اجتماع احتمال آن دو حادثه:

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

۲-۱ قضایای احتمالات

- قضیه عمومی حاصل ضرب احتمالها: اگر حادثه A قابل تجزیه به حوادث جزئی مانند B و C باشد، بطوریکه حوادث B و C با هم بتوانند اتفاق بیفتند، در این صورت آنرا بصورت حاصلضرب دو حادثه می‌نویسند:

$$A=BC=B\cap C$$

و احتمال حادثه A مساوی خواهد شد با:

$$P(A)=P(B\cap C)=P(B).P(C|B)$$

و $P(C|B)$ را احتمال شرطی حادثه C بر حسب B می‌نامند و این در حالت دو حادثه B و C نسبت به یکدیگر وابسته هستند و درحالیکه دو حادثه B و C نسبت به یکدیگر مستقل باشند. احتمال حادثه A مساوی خواهد شد با:

$$P(A)=P(B\cap C)=P(B).P(C)$$

قضیه فوق قابل تعمیم برای بیش از دو حادثه است.

- قضیه عمومی حاصل جمع احتمالها: اگر حادثه A قابل تجزیه به حوادث جزء مانند B و C باشد، بطوریکه در نتیجه آزمایش، حداقل یکی از حوادث (B یا C) بتواند اتفاق بیفتد. درآن صورت حادثه A از اجتماع حوادث جزء بدست می‌آید و احتمال حادثه A برابر است با:

$$A=B+C=B\cup C$$

$$P(A)=P(B)+P(C)-P(BC)$$

در این حالت دو حادثه B و C نسبت به یکدیگر سازگارند، ولی در صورت ناسازگار بودن حوادث B و C نسبت به یکدیگر، احتمال حادثه A از اجتماع احتمالهای حوادث B و C بدست می‌آید:

$$P(A)=P(B)+P(C)$$

قضیه فوق قابل تعمیم برای بیش از دو حادثه می‌باشد.

نکته: قضیه احتمال متوسط: اگر حادثه H قابل تجزیه به حوادث جزء مانند $H_1,H_2,\dots,H_s$ باشد، که دو به دو نسبت به یکدیگر ناسازگارند و حادثه A بتواند با هر یک ازحوادث H_i اتفاق بیفتد، درآن صورت:

$$P(A)=\sum P(H_i).P(A|H_i)$$

نکته: قضیه احتمال فرضیه‌ها(قضیه بیز): چنانچه در قضیه بالا، درنتیجه آزمایش، حادثه A وقوع کرده باشد، احتمال اینکه این حادثه برای فرضیه، iام (H_i) اتفاق افتاده باشد، مساوی است با احتمال فرضیه i ام در احتمال حادثه A بشرط فرضیه iام تقسیم بر احتمال متوسط:

$$P(H_i|A)=\frac{P(H_i).P(A|H_i)}{P(A)} \quad \frac{P(H_i).P(A|H_i)}{\sum P(H_i).P(A|H_i)}$$

- آزمایشهای تکراری (فرمول برنولی): اگر احتمال حادثه همواره ثابت باشد و برابر $P(A)=P$ و آزمایش را n بار تکرار کنیم، احتمال اینکه درn آزمایش مستقل از هم، m بار حادثه مورد نظر وقوع کند، از فرمول برنولی بدست می‌آید:

$$P_n(m)=\binom{n}{m}P^m q^{n-m} \qquad m=۰,۱,۲,\dots,n$$

که درآن q احتمال عدم وقوع حادثه است.

$$p(\overline{A})=۱-P=q$$

- فرمول مجانبی لاپلاس: در صورتی که n عدد نسبتاً بزرگی باشد، محاسبه آن بر اساس فرمول برنولی مشکل خواهد بود، ولی بر اساس قضایای حدی، از فرمول مجانبی لاپلاس می‌توان احتمال حادثه را بدست آورد، زیر ثابت می‌شود که با افزایش n، این احتمال $\underset{n}{p}(m)$ تقریباً بر طبق قانون نرمال توزیع می‌شود:

$$P_n(m)=\frac{۱}{\sqrt{npq}}\varphi(u)=\frac{۱}{\sqrt{npq}},\frac{۱}{\sqrt{۲\pi}}e^{-\frac{u^۲}{۲}}$$

کمیت تصادفی(متغیر تصادفی)

مقادیر عددی، که درنتیجه آزمایش اتفاق می‌افتد، متغیر تصادفی یا کمیت تصادفی نامیده می‌شود، زیرا اولاً اندازه آنها ازیک آزمایش به آزمایش دیگر تغییر می‌کند و ثانیاً غیر قابل پیش‌بینی می‌باشند. تناظری که بین مقادیر کمیت (X) و اندازه امکان وقوع آنها در جامعه بوجود می‌آید، قانون توزیع کمیت تصادفی X یا متغیر تصادفی نامیده می‌شود. اگر کمیت تصادفی X ناپیوسته باشد، قانون توزیع آن به صورت زیر نوشته می‌شود:

x	$x_۱$	$x_۲$	$x_۳$	...	x_s
P(x)	$P_۱$	$P_۲$	$P_۳$	...	P_s

تابع توزیع کمیت تصادفی X: احتمال اینکه کمیت تصادفی X مقادیری کمتر از x اختیار کند، تابع توزیع در نقطه x نامیده می‌شود و بصورت زیر نوشته می‌شود:

$$F(X)=P(X<x)=\sum_{x_i<x}P(X=x_i)$$

نکته: خواص تابع توزیع:

۱- تابع توزیع کمیتی غیر منفی است:

$$F_x(x)\geq ۰$$

۲- تابع توزیع کمیت تصادفی، تابعی غیر نزولی است: یعنی اگر

$$x_۱<x_۲ \rightarrow F(X_۱)\leq F(X_۲)$$

۳- تابع توزیع به ازاء حد بالای کمیت همواره مساوی یک است:

$$\lim_{x\to+\infty}F(x)=۱$$

۴- تابع توزیع به ازاء حد پائین کمیت، همواره مساوی صفر است:

$$\lim_{x\to-\infty} F(x)=0$$

۵- احتمال اینکه کمیت تصادفی x بین دو مقدار X_1 و X_2 مقدار اختیار کند، مساوی است با تابع توزیع در نقطه X_2 منهای تابع توزیع در نقطه X_1:

$$P(x_1<X<x_2)=F(x_2)-F(x_1)$$

نکته: تابع چگالی احتمالها: اگر کمیت تصادفی X پیوسته باشد، تابع چگالی احتمالها از مشتق تابع توزیع کمیت تصادفی X بدست می‌آید، زیرا برای کمیتهای پیوسته، احتمال انتخاب یک نقطه X_i مساوی است با صفر:

$$P(X=x_i)=0$$

بنابراین برای کمیتهای پیوسته، احتمال را در فاصله محاسبه می‌کنند و براساس تعریف چگالی وقتی $\Delta x\to 0$ میل کند، در حد، تابع چگالی احتمالها بدست می‌آید:

$$f(x)=\varphi(x)=F'(x)=\lim_{\Delta x\to 0}\frac{\Delta F(x)}{\Delta x}$$

- خواص تابع چگالی احتمالها:

۱- تابع چگالی کمیتی غیر منفی است:

$$\varphi(x)\geq 0$$

۲- انتگرال از تابع چگالی احتمالها در فاصله $(-\infty,+\infty)$ مساوی است با یک:

$$\int_{-\infty}^{+\infty}\varphi(x)dx=1$$

بنابراین تابع توزیع کمیت تصادفی x (پیوسته) با توجه به تعریف آن مساوی است با:

$$F(x)=P(X<x)=\int_{-\infty}^{x}\varphi(x)dx$$

و احتمال اینکه کمیت تصادفی x در فاصله x_1 و x_2 مقدار اختیار کند، به توسط تابع چگالی بدست می‌آید:

$$P(x_1<X<x_2)=\int_{x_1}^{x_2}\varphi(x)dx$$

- مشخص‌کننده‌های عددی قانون توزیع متغیر تصادفی

الف- امید ریاضی کمیت تصادفی نا پیوسته x مساوی است با:

$$EX=\sum xP(x)$$

در صورتیکه، کمیت، پیوسته باشد و تابع چگالی احتمالها برای آن وجود داشته باشد، برابر با:

$$EX=\int_{-\infty}^{+\infty}x\varphi(x)dx$$

ب- واریانس کمیت تصادفی x: در صورتیکه ناپیوسته باشد، مساوی خواهد شد با:

$$\begin{aligned}D(X)&=E(X-EX)^2=EX^2-(EX)^2\\&=\sum X^2P(x)-\left[\sum xP(x)\right]^2\end{aligned}$$

و اگر کمیت تصادفی x پیوسته باشد، واریانس مساوی خواهد شد با:

$$D(X)=\int_{-\infty}^{+\infty}X^2\varphi(x)dx-(EX)^2$$

- گشتاور مرتبه hام نسبت به مبداء صفر (گشتاورهای اولیه)

اگر کمیت تصادفی x نا پیوسته باشد، گشتاور مرتبه hام (m_h) مساوی خواهدشد با:

$$m_h=EX^h=\sum x^h p(x)$$

در صورتیکه پیوسته باشد وتابع چگالی احتمالها برای آن وجود داشته باشد، m_h مساوی خواهد شد با:

$m_h = EX^h = \int_{-\infty}^{+\infty} x^h \varphi(x)dx$

– **گشتاور مرتبه hام مرکزی:** اگر کمیت تصادفی x ناپیوسته باشد، گشتاور مرکزی مرتبه hام از رابطه زیر بدست می‌آید:

$\mu_h = E(X-EX)^h = \sum (x-EX)^h P(x)$

و اگر کمیت پیوسته باشد و تابع چگالی احتمالها برای آن وجود داشته باشد آنگاه

$\mu_h = \int_{-\infty}^{+\infty} (x-EX)^h \varphi(x)dx$

– **تابع مولد گشتاورها:** تابعی بصورت Ee^{xt} که از پارامتر t تبعیت می‌کند، تابع مولد گشتاورها نامیده می‌شود و بصورت:

$m_X(t) = Ee^{xt}$

نوشته می‌شود، زیرا تولید کننده گشتاورهای مراتب مختلف نسبت به مبداء صفر است.

نکته: خواص تابع مولد گشتاورها:

۱- خاصیت یگانگی- تابع مولد گشتاورها بطور یگانه قانون توزیع کمیت تصادفی X را معین می‌کند، یعنی هر قانون توزیعی فقط یک تابع مولد گشتاورها دارد و هر تابع مولد گشتاوری منحصراً متعلق به یک قانون توزیع است.

۲- اگر کمیت تصادفی X در عدد ثابتی مانند C ضرب شود، آنگاه:

$m_{cx}(t) = m_x(ct)$

۳- اگر کمیتهای $X_1, X_2, ..., X_n$,... مستقل از یکدیگر با تابع مولد گشتاورهای $M_{xi}(t)$ توزیع شده باشند، آنگاه تابع مولد گشتاورهای

$y = x_1 + x_2 + ... + x_n$

برابر خواهد شد با: $m_y(t) = m_{x_1}(t).m_{x_2}(t) \times ... \times m_{x_n}(t)$

۴- گشتاور مرتبه hام نسبت به مبداء صفر، از مشتق مرتبه hام تابع مولد گشتاورها نسبت به t وقتی‌که پس از عمل مشتق‌گیری $t=0$ شود، بدست می‌آید:

$$m_h = \left.\frac{d^h m \times (t)}{dt^h}\right|_{t=0}$$

بنابراین گشتاور مرتبه اول (m_1) بصورت زیر محاسبه می‌شود.

$$m_1 = \left.\frac{dm \times (t)}{dt}\right|_{t=0}$$

و گشتاور مرتبه دوم (m_2) خواهد شد:

$$m_2 = \left.\frac{d^2 m \times (t)}{dt^2}\right|_{t=0}$$

قوانین توزیعهای مهم برای کمیت تصادفی

در این قسمت فقط راجع به بعضی از انواع مهم قوانین توزیع کمیتهای تصادفی نا پیوسته و پیوسته بحث می‌شود که در جامعه بیشتر تکرار می‌شوند، لازم به ذکر است که تعداد قوانین توزیع کمیتهای تصادفی بسیارند که در کاربرد نظریه احتمال در رشته‌های مختلف علوم مورد استفاده قرار می‌گیرند.

۱- قانون توزیع دو نقطه‌ای (برنولی): اگر کمیت تصادفی x مقادیر ۰ و۱ را با احتمالهای $P(X=1)=P$ و $P(X=0)=q$ اختیار کند، در این صورت کمیت تصادفی X بر طبق قانون دو نقطه‌ای توزیع می‌شود:

$P(x) = P^x q^{1-x} \qquad x = 0/1$

امید ریاضی آن برابر است با:

$$EX=\sum xP(x)=\sum xp^{x}q^{1-x}=P$$

و واریانس آن برابر است با:

$$D(x)=EX^{2}-(EX)^{2}=\sum x^{2}P^{x}q^{1-x}-P^{2}=p-p^{2}=p(1-p)=pq$$

۲- قانون توزیع دو جمله‌ای: اگر احتمال حادثه همواره ثابت و برابر p باشد و آزمایش را n^{1} بار تکرار کنیم و کمیت تصادفی x تعداد وقوع حادثه در n آزمایش باشد، آنگاه کمیت تصادفی x بر طبق قانون دو جمله‌ای (قانون برنولی) توزیع می‌شود:

$$P(x)=\binom{n}{x}P^{x}q^{n-x}\qquad x=0,1,2,..,n$$

امید ریاضی آن برابر با:

$$EX=\sum xp(x)=\sum x\binom{n}{x}P^{x}\varphi^{n-x}=nP$$

و واریانس آن برابر با:

$$D(X)=E(X-EX)^{2}=nPq$$

۳- قانون توزیع پواسن: اگر احتمال حادثه ثابت و برابر p باشد و P عدد کوچکی باشد. $(P<0/1)$، بطوریکه $\lambda=nP\le 5$، آنگاه کمیت تصادفی X ، تعداد وقوع حادثه در n آزمایش، بر طبق قانون پواسن توزیع می‌شود:

$$P(x)=\frac{\lambda^{x}.e^{-\lambda}}{x!}\quad x=0,1,2,...$$

امید ریاضی آن برابر با:

$$EX=\sum xP(x)=\sum x\frac{\lambda^{x}.e^{-\lambda}}{x!}=\lambda$$

و واریانس آن برابر با:

$$D(X)=EX^{2}-(EX)^{2}=\lambda$$

خواهد بود.

۴- قانون توزیع هیپرژئومتریک (فوق هندسی): اگر احتمال حادثه از یک آزمایش به آزمایش دیگر تغییر کند و آزمایش را n بار تکرار کنیم و کمیت تصادفی x تعداد وقوع حادثه در n آزمایش باشد، آنگاه کمیت تصادفی x بر طبق قانون هیپرژئومتریک توزیع خواهد شد:

$$P(x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}\qquad x=0,1,2,...,n$$

اگر مجموعه‌ای از N عنصر که M عنصر آن دارای خاصیت مورد نظر باشند وآزمایش n بار انجام گیرد، امید ریاضی برای این کمیت مساوی است با:

$$EX=nP=n.\frac{M}{N}$$

که $\frac{M}{N}$ اولین احتمال حادثه است.

و واریانس آن مساوی است با:

$$D(X)=EX^{2}-(EX)^{2}$$

$$\frac{N-n}{N-1}.n\frac{M}{N}\left(1-\frac{M}{N}\right)$$

بطوریکه ملاحظه می‌شود، قانون توزیع هیپرژئو متریک از ۳ پارامتر N، M و n تبعیت دارد.

۵- قانون توزیع هندسی: اگر احتمال حادثه همواره ثابت و برابر P باشد و آزمایش را آنقدر تکرار کنیم تا حادثه مورد نظر اتفاق بیفتد، کمیتX تعداد آزمایش تا وقوع حادثه باشد، در آن صورت کمیت تصادفی X بر طبق قانون هندسی توزیع می‌شود:

$$P(x) = q^{x-1}.P \qquad x = 1,2,3,...$$

امید ریاضی آن مساویست با: $EX\sum xP(x) = \frac{1}{P}$

و واریانس آن مساوی خواهد شد با: $D(X) = EX^2 - (EX)^2 = \frac{q}{P^2}$

۶- قانون توزیع مستطیلی (یکنواخت) پیوسته: اگر کمیت تصادفی X مقادیری در فاصله (a,b) را با احتمال ثابت $\frac{1}{b-a}$ بتواند اختیار کند، آنگاه کمیت تصادفی x بر طبق قانون یکنواخت با تابع چگالی احتمالهای زیر توزیع می‌شود:

$$\varphi(X) = \frac{1}{b-a} \qquad a < x < b$$

تابع توزیع برای این کمیت بصورت: $F(x) = \frac{x-a}{b-a}$ خواهد بود.

امید ریاضی و واریانس آن به ترتیب برابرند با:

$$EX = \frac{a+b}{2}$$

$$D(X) = \frac{(b-a)^2}{12}$$

۷- قانون توزیع نرمال: اگر کمیت تصادفی x مقادیر خود را در فاصله $(-\infty,+\infty)$ با تابعی به صورت

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty$$

اختیار کند، دراین صورت کمیت X بر طبق قانون نرمال توزیع خواهد شد.

امید ریاضی و واریانس آن به ترتیب برابرند با:

$$EX = \mu$$

$$D(X) = \sigma^2$$

برای اینکه بتوانیم احتمال حادثه را با توزیع نرمال محاسبه کنیم، بهتر است که توزیع نرمال را به توزیع نرمال استاندارد تبدیل کرده و با استفاده از جداول تهیه شده برای توزیع نرمال استاندارد، احتمال حوادث را محاسبه کنیم. توزیع نرمال به توسط رابطه $U = \frac{X-\mu}{\sigma}$ به نرمال استاندارد تبدیل می‌شود که همواره دارای امید ریاضی صفر و واریانس یک است و تابع چگالی احتمالهای آن بصورت.

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \qquad -\infty < U < +\infty$$

تعریف می‌شود که آنرا بصورت زیر می‌نویسند:

$$U \sim N(0,1)$$

تابع توزیع کمیت تصادفی U بصورت:

$$F(u) = \int_{-\infty}^{U} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

بدست می‌آید که تابعی نامتقارن است.

اگر بخواهیم تابع توزیع نرمال استاندارد شده را بصورت متقارن مورد استفاده قرار دهیم، از تابع لاپلاس استفاده می‌کنیم که بصورت زیر تعریف می‌شود:

$$\varphi(U)=\int_{0}^{u}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du \qquad -\infty<U<+\infty$$

زیرا تابع لاپلاس، تابعی فرد است: $\phi(-u)=-\phi(u)$ و به ازاء $U=0$، $\phi(0)=0$ و به‌ازاء $U\geq 4$ $\phi(U\geq 4)=\frac{1}{2}$ است، به‌همین دلیل جدول تابع لاپلاس در فاصله صفر تا ۴ تهیه شده است.

۸- قانون توزیع گاما: اگر کمیت تصادفی X مقادیر خود را درفاصله صفر تا بی‌نهایت اختیار کند و تابع چگالی احتمالهای آن بصورت:

$$\varphi(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}.e^{-\beta x} \quad x>0$$

باشد، کمیت تصادفی x بر طبق قانون گاما توزیع خواهد شد. که در آن

$$\Gamma(\alpha)=\int_{0}^{\infty}x^{\alpha-1}.e^{-x}dx$$

تابع گاما یا انتگرال نوع دوم اولر نامیده می‌شود:

خواص تابع گاما:

$$1\text{-}\ \Gamma(\alpha+1)=\alpha\Gamma(\alpha)=\alpha!$$
$$2\text{-}\ \Gamma(2)=\Gamma(1)=1$$
$$3\text{-}\ \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

امید ریاضی و واریانس توزیع گاما برابر است با:

$$EX=\frac{\alpha}{\beta}$$
$$D(X)=\frac{\alpha}{\beta^2}$$

۹- قانون توزیع نمائی: اگر در توزیع گاما، $\alpha=1$ قرار داده شود، در آن صورت کمیت تصادفی X دارای توزیع نمائی با تابع چگالی احتمالهای زیر خواهد بود:

$$\varphi(X)=\beta e^{-\beta x} \qquad x>0$$

امید ریاضی و واریانس آن بصورت زیر در خواهد آمد:

$$EX=\frac{1}{\beta}$$
$$\phi(X)=\frac{1}{\beta^2}$$

که در این رابطه $\mathrm{cov}(x,y)$، کواریانس نامیده می‌شود که پراکندگی توأم دو متغیر x و y را نشان می‌دهد و به نام گشتاور همبستگی نیز موسوم است و بصورت زیر محاسبه می‌شود.

$$\mathrm{Cov}(x,y)=\frac{\sum(x-\bar{x})(y-\bar{y})}{n}=\frac{\sum xy}{n}-\bar{x}\,\bar{y}$$

ضریب تعیین، میزان تغییر پذیری مقادیر مشاهده شده صفت y را حول معادله رگرسیون به میزان تغییرپذیری صفت x نشان می‌دهد. یعنی اگر ρ^2 را در ۱۰۰ ضرب کنیم و بر حسب درصد بیان نمائیم، آنگاه R^2 درصد تغییرات صفت y را ناشی از اثرات صفت متغیر x نشان می‌دهد.

قانون توزیع دو بعدی

اگر کمیتهای x و y ناپیوسته بوده و توزیع مشترک آنها بصورت

$$p_{xy}(x,y)=p(x=x:,y=y_j)=p_{ij}$$

و یا بصورت جدول قانون توزیع دو کمیت x و y باشد:

x　y	$y_۱$	$y_۲$	…	y_t	p(x)
$x_۱$	$p_{۱۱}$	$p_{۱۲}$	…	$p_{۱t}$	$p_{۱۰}$
$x_۲$	$p_{۲۱}$	$p_{۲۲}$	…	$p_{۲t}$	$p_{۲۰}$
⋮	⋮	⋮	⋮	⋮	⋮
x_s	$p_{s۱}$	$p_{s۲}$	…	p_{st}	$p_{s۰}$
p(y)	$p_{۰۱}$	$p_{۰۲}$	…	p.t	۱

توزیعهای حاشیه ای x و حاشیه ای y و هم چنین توزیع شرطی x برحسب y

و y بر حسب x به توسط جدول دو بعدی و یا قانون توزیع توأم آنها بدست می‌آید.

اگر کمیتهای x و y پیوسته باشند و تابع چگالی احتمالها برای آنها وجود داشته باشد، آنرا بصورت

$$\varphi(x,y)=F''_{xy}(x,y)=\frac{\Delta^۲F(x:y)}{\Delta x\Delta y}$$

$$\Delta x\to ۰$$
$$\Delta y\to ۰$$

می‌نویسد. همواره شرایط زیر برای تابع چگالی احتمالهای دو کمیت x و y نیز صادق است:

$$\varphi(x,y)\geq ۰$$
$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\varphi(x,y)dx\ \ dy=۱$$

و تابع توزیع احتمالهای آن

$$F_{xy}(x,y)=P(X<x,Y<y)=\int_{-\infty}^{x}\int_{-\infty}^{y}\varphi(x,y)dxd_y$$

شرط لازم و کافی برای اینکه دو کمیت x و y از یکدیگر مستقل باشند باید تابع توزیع دو کمیت x و y مساوی باشد با حاصل ضرب تابع توزیع های دو کمیت x و y یعنی

$$\underset{xy}{\overset{\circ}{F}}(x,y)=F_X(x).F_Y(y)$$

شرط فوق هم برای تابع چگالی احتمالهای دو کمیت و هم برای قانون توزیع مشترک دو کمیت x و y صادق است.

در صورتیکه دو کمیت x و y پیوسته بوده و تابع چگالی احتمالها برای آنها وجود داشته باشد، آنگاه توزیع‌های حاشیه‌ای x و y با انتگرال گیری از تابع چگالی دو کمیت یکبار نسبت به y و یکبار هم نسبت به x بدست می‌آید:

$$\varphi(x)=\int_y\varphi(x,y)dy$$
$$\varphi(y)=\int_x\varphi(x,y)dx$$

توزیع‌های شرطی y برحسب x به صورت

$$\varphi(y\mid x)=\frac{\varphi(x,y)}{\varphi(x)}$$

و x برحسب y به صورت زیر بدست می‌آید:

$$\varphi\left(x/y\right)=\frac{\varphi(x,y)}{\varphi(y)}$$

امید ریاضی کمیت x بتوسط توزیع حاشیه‌ای x $\quad EX=\int x\varphi(x)dx,$

و امید ریاضی کمیت y به توسط توزیع حاشیه‌ای y $\quad Ey=\int_y^x y\varphi(y)dy$

و واریانس برای کمیت x و برای کمیت y به صورت

$$D(x)=EX^{۲}-(EX)^{۲}=\int X^{۲}\varphi(x)dx-(EX)^{۲}$$
$$D(y)=Ey^{۲}-(Ey)^{۲x}=\int y^{۲}\varphi(y)dy-(Ey)^{۲}$$

گشتاور همبستگی یا کواریانس بین y دو کمیت x و y به صورت زیر محاسبه می‌شود.

$$\mu_{۱,۱}=\mathrm{cov}(x,y)=Exy-Ex.Ey$$

امید ریاضی $x-y, x+y$ عبارتنداز:

$$E(x+y)=EX+Ey$$
$$E(x-y)=EX-Ey$$

و واریانس آنها به صورت

$$D(x+y)=D(x)+D(y)+۲\,\mathrm{cov}(x,y)$$
$$D(x-y)=D(x)+D(y)-۲\,\mathrm{cov}(x,y)$$

اگر دو کمیت x و y از یکدیگر مستقل باشند، آنگاه $Exy=Ex.Ey$

قضایای حدی

تجارب گذشته نشان داده است که حوادث با احتمالهای نزدیک به یک تقریباً بطور یقین وقوع خواهند کرد و به همان ترتیب حوادث با احتمالهای کوچک نزدیک به صفر به ندرت وقوع می‌کنند. یکی از هدفهای نظریه احتمال برقرار کردن نظم‌ها یا قوانینی است که با احتمالهای نزدیک به یک حوادث بتوانند وقوع کنند و اهمیت این نظم‌ها را قانون اعداد بزرگ مشخص می‌سازد.

۱- لم چه بی شف- کمیت تصادفی X با قانون نامشخص توزیع شده است و امید ریاضی برای آن وجود دارد، احتمال اینکه کمیت تصادفی X مقادیری بیش از a اختیار کند، از $\frac{Ex}{a}$ تجاوز نخواهد کرد:

$$P(x\geq a)\leq\frac{EX}{a}$$

که هم برای کمیتهای ناپیوسته و هم برای کمیتهای پیوسته صادق است. در نتیجه با توجه به تعریف تابع توزیع، خواهیم داشت:

$$F(a)=P(X<a)\geq ۱-\frac{EX}{a}$$

۲- نامساوی نوع اول چه بی شف:

کمیت تصادفی X بر طبق قانون $f(x)$ با واریانس معلوم $D(x)$ توزیع شده است، احتمال اینکه قدر مطلق تفاضل کمیت X از امید ریاضی آن، از عدد مثبتی مانند ε کوچکتر باشد، مساوی است با:

$$P(|X-EX|<\varepsilon)\geq ۱-\frac{D(X)}{\varepsilon^{۲}}$$

بر اساس قضیه عکس، می‌توان نوشت:

$$P(|X-EX|>\varepsilon)\le\frac{D(X)}{\varepsilon^{2}}$$

۳- نامساوی نوع دوم چه بی شف:

اگر کمیتهای $x_1,x_2,\dots,x_n$ مستقل از یکدیگر با امیدهای ریاضی $EX_1=a_1, EX_2=a_2,\dots,EX_n=a_n$ و واریانس‌های $D(x_1)=\sigma_1^2$، $D(X_2)=\sigma_2^2,\dots,D(X_n)\sigma_n^2$ توزیع شده باشند، احتمال اینکه قدر مطلق تفاضل میانگین کمیتهای $x_1,x_2,\dots,x_n$ از میانگین امیدهای ریاضی این کمیتها، کوچکتر از هر عدد مثبتی مانند ε باشد، مساوی است با:

$$P\left(\left|\frac{\sum X_i}{n}-\frac{\sum a_i}{n}\right|<\varepsilon\right)\ge 1-\frac{\sum\sigma_i^{2}}{n^{2}\varepsilon^{2}}$$

در حالت خاص، اگر کمیتها از یک جامعه انتخاب شود، آنگاه $EX_1=EX_2=\dots=EX_n=\mu$ و $D(x_1)=D(x_2)=\dots=D(X_n)=\sigma^2$ باشد، رابطه فوق بصورت زیر بدست می‌آید:

$$P\left(\left|\frac{\sum X_i}{n}-\mu\right|<\varepsilon\right)\ge 1-\frac{\sigma^{2}}{n\varepsilon^{2}}$$

– **قانون اعداد بزرگ به صورت چه بی شف:** در نامساوی نوع دوم چه بی شف، اگر n بطور نامحدود افزایش یابد، این احتمال مساوی با یک خواهد شد.

$$\lim_{n\to\infty} p\left(\left|\frac{\sum x_i}{n}-\frac{a_1+a_2+\dots+a_n}{n}\right|<\varepsilon\right)=1$$

- قانون اعداد بزرگ قوی به صورت چه بی شف:

$$p\left(\lim_{n\to\infty}\frac{\sum x_i}{n}=a\right)=1$$

– **قضیه حدی مرکزی لیاپانوف:** اگر کمیتهای $x_1,x_2,\dots,x_n$ ، ... دارای توزیع یکسان نباشند، قانون نرمال بعنوان توزیع حدی حاصل جمع یا میانگین حسابی دنباله‌ای از کمیتهای تصادفی بیان می‌شود:

$$Y=x_1+x_2+\dots+x_n$$

و کمیت Z_n کمیت استاندارد شده y_n باشد، آنگاه:

$$Z_n=\frac{y_n-Ey_n}{\sigma_{y_n}}$$

بر طبق قانون نرمال استاندارد توزیع خواهد شد:

$$\lim_{n\to\infty}F(Z_n)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^{2}}{2}}dz$$

به شرط اینکه حد گشتاور مطلق مرکزی مرتبه سوم مساوی صفر باشد.

$$\lim_{n\to\infty}\frac{\sqrt[3]{\sum|\mu_{3i}|}}{\sqrt{\sum S_i^{2}}}=0$$

استنتاج آماری

– **نمونه و تابع نمونه‌ای:** از آنجا که برای شناسائی صفت یا صفات متغیر در جامعه مورد مطالعه، امکان مشاهده همه عناصر جامعه وجود ندارد (به دلایل بالا بودن هزینه، عدم تأمین نیروی انسانی لازم، انهدام جامعه آماری و از همه مهمتر زمان

جمع‌آوری اطلاعات)، لذا جزئی از جامعه مورد مطالعه را بطور تصادفی انتخاب کرده و نتایج مشاهدات را به جامعه کل تعمیم می‌دهیم که این روش را روش استقرائی می‌نامند.

روش انتخاب نمونه‌ها به دو طریق (با جایگذاری) و (بدون جایگذاری) انجام می‌گیرد.

– **انتخاب تصادفی :** انتخابی را گویند که تمامی عناصر جامعه امکان معین و از قبل تعیین شده برای انتخاب داشته باشند.

– **تعریف نمونه تصادفی:** نمونه تصادفی به حجم n از یک کمیت تصادفی X با توزیع $f(x)$، مجموعه‌ای از n کمیت تصادفی مستقل از هم $x_1, x_2, ..., x_n$ است که هر یک توزیع یکسان با جامعه اصلی $f(x)$ داشته باشند.

–**تابع نمونه‌ای (statistic)** یا آماره تابعی از متغیرهای $x_1, x_2, ..., x_n$ را در نمونه‌ایی به حجم n، تابع نمونه‌ای یا Statistic می‌نامند و بصورت زیر نشان می‌دهند:

$$\theta = f(x_1, x_2, ..., x_n)$$

و چون تابعی از کمیتهای تصادفی است، لذا تابع نمونه‌ای خود نیز یک کمیت تصادفی خواهد بود. در نمونه‌ای به حجم n می‌توان توابع مختلف و بسیار زیادی تعریف کرد، از جمله:

$$\theta = f(x_1, x_2, ..., x_n) = x_1 + x_2 + ... + x_n = \sum x_i$$

$$\theta = f(x_1, x_2, ..., x_n) = \frac{\sum x_i}{n} = \bar{x}$$

$$\theta = f(x_1, x_2, ..., x_n) = \frac{\sum (X - \bar{X})^2}{n} = \tilde{S}^2$$

$$\theta = f(x_1, x_2, ..., x_n) = \frac{\sum (X - \bar{X})^2}{n-1} = S^2$$

نکته: نظریه توزیعهای نمونه به دو گروه تقسیم می‌شوند:

الف- نظریه توزیع‌های دقیق توابع نمونه‌ای

ب- نظریه توزیعهای مجانبی توابع نمونه‌ای

قانون توزیع توابع نمونه‌ای را دقیق گویند، هرگاه قانون توزیع $f(x_1, ..., x_n)$ برای تمامی اعداد طبیعی n صحیح باشد.

قانون توزیع $f(x_1, ..., x_n)$ را توزیع مجانبی نامند، هر گاه توزیع دقیق تابع نمونه‌ای برحسب $n \to \infty$ به این قانون توزیع گرایش داشته باشد.

– قوانین توزیع‌های مهم

۱- قانون توزیع χ^2 (کای دو): کمیتی که از مجموع مجذورات استاندارد شده متغیرهای $x_1, x_2, ..., x_n$ که بر طبق قانون نرمال با امید ریاضی μ و واریانس σ^2 توزیع شده‌اند، بدست آید، برطبق قانون، χ^2 توزیع می‌شود:

$$\chi^2 = \sum U_i^2 = \sum \left(\frac{x_i - \mu}{\sigma} \right)^2$$

و تابع چگالی احتمالهای آن بصورت

$$\varphi(x) = \frac{1}{2^{\frac{\upsilon}{2}} \Gamma\left(\frac{\upsilon}{2}\right)} x^{\frac{\upsilon}{2}-1} . e^{-\frac{x}{2}} \qquad 0 < x < \infty$$

بدست می‌آید که فقط از یک پارامتر υ تبعیت می‌کند و آنهم درجه آزادی توزیع χ^2 است:

اگر در کمیت χ^2 پارامتر μ مشخص باشد، آنگاه $\upsilon = n$ و در صورتی که پارامتر μ نامعلوم باشد، آنگاه درجه آزادی χ^2 $\upsilon = n - 1$ خواهد شد.

امید ریاضی توزیع $\chi^{۲}$ مساوی است با درجه آزادی آن:

$$E(X_{\upsilon}^{۲})=\upsilon$$

و واریانس کمیت $\chi^{۲}$ مساوی است با دو برابر درجه آزادی آن:

$$E\left(X_{v}^{۲}\right)=۲\upsilon$$

تابع مولد گشتاورها برای کمیت $\chi^{۲}$ مساوی است با:

$$m_{x}(t)=Ee^{xt}=(۱-t)^{-\frac{\upsilon}{۲}}$$

بر اساس قضیه حدی مرکزی، نتیجه می‌شود که با افزایش درجه آزادی $\chi^{۲}$، تابع توزیع کمیت تصادفی به تابع توزیع نرمال استاندارد گرایش پیدا می‌کند:

$$\lim_{n\to\infty} F\left(\frac{X^{۲}-\upsilon}{\sqrt{۲\upsilon}}\right)=\int_{-\infty}^{u}\frac{۱}{\sqrt{۲\pi}}e^{-\frac{u^{۲}}{۲}}du$$

۲-توزیع t - استودنت: کمیتی که از متغیرهای استاندارد شده $U_{۱}, U_{۲},..., U_{n}$ مستقل از هم بصورت:

$$t=\frac{U}{\sqrt{\frac{\chi_{\upsilon}^{۲}}{\upsilon}}}$$

توزیع شده باشد، کمیت تصادفی با قانون t - استودنت نامیده می‌شود و تابع چگالی احتمالهای آن بصورت زیر بدست می‌آید:

$$\varphi_{t}(x)=\frac{\Gamma\left(\frac{\upsilon+۱}{۲}\right)}{\sqrt{\pi\upsilon}\,\Gamma\left(\frac{\upsilon}{۲}\right)}.\left(۱+\frac{\chi^{۲}}{\upsilon}\right)^{-\frac{\upsilon+۱}{۲}} \qquad -\infty<x<+\infty$$

این توزیع فقط از یک پارامتر یعنی υ تبعیت می‌کند که آنهم درجه آزادی t - استودنت می‌باشد و با افزایش آن توزیع t - استودنت به توزیع نرمال تبدیل می‌شود $(V\geq ۳۰)$.

امید ریاضی و واریانس برای توزیع t - استودنت برابر است با:

$$E(t)=۰$$

$$D(t)=\frac{\upsilon}{\upsilon-۲}$$

۳- توزیع F: کمیت تصادفی که از خارج قسمت دو توزیع $\chi^{۲}$ بدست می‌آید، برطبق قانون F توزیع می‌شود:

$$F=\frac{\frac{\chi_{\upsilon_{۱}}^{۲}}{\upsilon_{۱}}}{\frac{\chi_{\upsilon_{۲}}^{۲}}{\upsilon_{۲}}}$$

که در آن $\upsilon_{۱}$ و $\upsilon_{۲}$ درجات آزادی کمیتهای $\chi_{\upsilon_{۱}}^{۲}$ و $\chi_{\upsilon_{۲}}^{۲}$ می‌باشند و درجه آزادی صورت و مخرج توزیع F نامیده می‌شوند.

برای $P>۰/۵$ مقادیر کرانیتهای توزیع F در جدول آورده شده است و برای $P<۰/۵$ مقادیر کرانیتهای توزیع F از رابطه زیر بدست می‌آید:

$$F_{\alpha};F_{\alpha;\upsilon_{۱},\upsilon_{۲}}=\frac{۱}{F_{۱-\alpha,\upsilon_{۲},\upsilon_{۱}}}$$

با تغییر مقادیر $\upsilon_۱$ و $\upsilon_۲$ در توزیع F می‌توان به توزیعهای دیگری دست یافت.

– قانون توزیع واریانس‌های نمونه‌ای: واریانسهای نمونه‌ای، $S^۲$، $\tilde{S}^۲$ و $S_*^۲$ که از جامعه نمونه بدست می‌آیند، به دلیل اینکه متغیرهای تصادفی می‌باشند و اندازه آنها از یک آزمایش به آزمایش دیگر تغییر می‌کند، لذا باید قانون توزیع داشته باشند که بصورت زیر بدست می‌آیند:

الف- واریانس نمونه‌ای در حجمهای کوچک $(n<۳۰)$ بصورت زیر تعریف می‌گردد:

$$\theta = f(x_۱, x_۲, \ldots, x_n) = \frac{\sum (X_i - \overline{X})^۲}{n-۱} = S^۲$$

که بر طبق قانون $\chi^۲$، با درجه آزادی $n-۱$ توزیع می‌شود.

$$S^۲ = \frac{\sum (X-\overline{X})^۲}{n-۱} = \frac{\sigma^۲ \sum (X-\overline{X})^۲}{\sigma^۲ (n-۱)} = \frac{\sigma^۲}{n-۱} \sum \left(\frac{X_i - \overline{X}}{\sigma} \right)^۲ = \frac{\sigma^۲}{n-۱} \chi_{n-۱}^۲$$

با امید ریاضی

$$ES^۲ = E \frac{\sigma^۲}{n-۱} \chi_{n-۱}^۲ = \sigma^۲$$

و واریانس

$$D(S^۲) = D\left(\frac{\sigma^۲}{n-۱} \chi_{n-۱}^۲ \right) = \frac{۲\sigma^۴}{n-۱}$$

ب- واریانس نمونه‌ای در حجم های بزرگ $n>۳۰$ بصورت زیر تعریف می‌گردد:

$$\theta = f(x_۱, x_۲, \ldots, x_n) = \frac{\sum (X-\overline{X})^۲}{n} = \tilde{S}^۲$$

این پارامتر بر طبق قانون $\chi^۲$ با درجه آزادی $n-۱$ توزیع می‌شود:

$$\tilde{S}^۲ = \frac{\sigma^۲ \sum (X-\overline{X})^۲}{\sigma^۲ n} = \frac{\sigma^۲}{n} \chi_{n-۱}^۲$$

با امید ریاضی

$$E(\tilde{S}^۲) = E \frac{\sigma^۲}{n} \chi_{n-۱}^۲ = \sigma^۲ - \frac{\sigma^۲}{n}$$

و واریانس

$$D(\tilde{S}^۲) = D\left(\frac{\sigma^۲}{n} \chi_{n-۱}^۲ \right) = \frac{۲(n-۱)\sigma^۴}{n^۲}$$

ج- واریانس نمونه‌ای وقتی امید ریاضی جامعه معلوم باشد: اگر در توزیع نرمال امید ریاضی μ معلوم باشد، واریانس نمونه‌ای بصورت زیر تعریف می‌شود.

$$\theta = f(x_۱, \ldots, x_۲) = \frac{\sum (X-\mu)^۲}{n} = S_*^۲$$

این پارامتر بر طبق قانون $\chi^۲$ با درجه آزادی n توزیع می‌شود:

$$S_*^۲ = \frac{\sigma^۲ \sum (X=\mu)^۲}{\sigma^۲ n} = \frac{\sigma^۲}{n} \chi_n^۲$$

با امید ریاضی $E(S_*^۲) = \sigma^۲$ و واریانس

$$D(S_*^۲) = \frac{۲\sigma^۴}{n}$$