مدرسان شریف

فصل اول

«آمـــار توصیفی»

درسنامه (۱): تعاریف پایه‌ای

علم آمار: آمار عبارت است از مجموعه‌ای از روش‌ها و تکنیک‌هایی که در جمع‌آوری، خلاصه کردن، طبقه‌بندی، پردازش، نمایش، تجزیه و تحلیل و تفسیر اطلاعات آماری و نهایتاً نتیجه‌گیری و تعمیم نتایج، مورد استفاده قرار می‌گیرد.

علم آمار به سه شاخه آمار توصیفی، آمار استنباطی (پارامتری) و آمار ناپارامتری تقسیم می‌شود.

آمار توصیفی: مجموعه‌ای از روش‌هایی است که برای سازماندهی، خلاصه کردن، تهیه جدول، رسم نمودار، تبیین و توصیف خصوصیات مهم مجموعه داده‌ها به کار برده می‌شود.

آمار استنباطی (پارامتری): آمار استنباطی به معرفی فنون و روش‌هایی می‌پردازد که با استفاده از آن اطلاعات و نتایج حاصل از نمونه به کل جامعه آماری تعمیم داده می‌شود. به عبارت دیگر آمار استنباطی روشی برای تعمیم نتایج و استنباط حاصل‌شده از نمونه به کل جامعه آماری برمبنای تئوری احتمالات می‌باشد.

آمار ناپارامتری: این نوع آمار وقتی به‌کار می‌رود که توزیع جامعه نامعلوم باشد یا همه مشاهدات در اختیار نباشد.

مثال ۱: در کدام نوع از روش‌های آماری هیچ‌گونه تعمیمی اتفاق نمی‌افتد؟ (مدیریت و حسابداری ـ سراسری ۱۴۰۰)

۱) پارامتریک ۲) ناپارامتریک ۳) استنباطی ۴) توصیفی

☑ **پاسخ:** گزینه «۴» در آمار توصیفی ابتدا به روش سرشماری، اطلاعات تمام اعضای جامعه جمع‌آوری می‌شود. سپس صفت تمام داده‌ها مورد بررسی قرار می‌گیرد و هیچ‌گونه استنباطی بر روی آن‌ها انجام نمی‌شود.

جامعه آماری: جامعه آماری مجموعه‌ای از افراد، اشیاء یا عناصر می‌باشد که حداقل در یک صفت مشترک هستند. به عنوان مثال، جامعه دانشجویان یک دانشگاه در صفت دانشجو، مشترک‌اند. حجم جامعه را معمولاً با N نشان می‌دهند.

نمونه: هر بخش از یک جامعه را که بر طبق قانون خاصی انتخاب می‌شود و مطالعه آن به جای مطالعه تمام اعضای جامعه انجام می‌شود، نمونه‌ای از جامعه می‌نامند. به عنوان مثال انتخاب تعداد ۱۰ کالا از تولیدات یک شرکت تشکیل یک نمونه آماری به اندازه ۱۰ را می‌دهد. حجم نمونه را معمولاً با n نشان می‌دهند.

صفت مشخصه: صفتی است که اعضای جامعه آماری در آن مشترک هستند. به عنوان مثال در جامعه آماری دانشجویان یک دانشگاه، تمام اعضا در صفت دانشجو مشترک‌اند اما در صفاتی نظیر قد، وزن، گروه خونی، سن، میزان علاقه به سینما و موارد مشابه متغیر هستند. به این صفات که از یک دانشجو به دانشجوی دیگر تغییر می‌کند، **صفت متغیر** یا **متغیر** می‌گویند. با توجه به کاربرد صفت متغیر، به دانشجویان توصیه می‌شود مفهوم آن را به طور عمیق یاد بگیرند.

صفت متغیر (متغیر): به صفتی که از یک فرد به فرد دیگر یا از یک مشاهده به مشاهده دیگر تغییر می‌کند، صفت متغیر می‌گویند.

مثال ۲: کدام‌یک از گزینه‌های زیر تعریف صفت مشخصه است؟ (مدیریت و حسابداری ـ سراسری ۹۳)

۱) از فردی به فرد دیگر تغییر می‌کند.
۲) صفت مشترک بین کلیه افراد جامعه است.
۳) متمایزکننده عناصر جامعه از یکدیگر است.
۴) عنصر مشترک جوامع آماری مختلف است.

☑ **پاسخ:** گزینه «۲» صفت مشخصه، صفت مشترک بین اعضای جامعه است.

صفات متغیر به دو دسته کیفی و کمّی تقسیم می‌شوند.

متغیرهای کیفی: متغیرهایی هستند که بدون واحد بوده و اندازه‌گیری آن‌ها با مقیاس‌ها و واحدهای اندازه‌گیری مرسوم امکان‌پذیر نیست. به‌عنوان مثال گروه خونی افراد متغیر کیفی است.

متغیرهای کمّی: متغیرهایی هستند که قابل اندازه‌گیری بوده و واحد اندازه‌گیری دارند یا قابل شمارش بوده و دارای واحد شمارش هستند. به عنوان مثال قد، وزن، سن، مساحت، حجم، سرعت و موارد مشابه متغیرهای کمّی محسوب می‌شوند.

مقیاس‌های اندازه‌گیری متغیرها: مقیاس‌های اندازه‌گیری متغیرها شامل موارد زیر است:

۱ـ مقیاس اسمی: از این مقیاس برای اندازه‌گیری متغیرهای کیفی نظیر رنگ، طعم، نژاد، گروه خونی، مذهب، دین، جنسیت، زبان مادری و موارد مشابه استفاده می‌شود. به عنوان مثال وضعیتی را در نظر بگیرید که به گزینه‌های مربوط به رنگ‌های آبی، قرمز و سبز به ترتیب کدهای ۱، ۲ و ۳ تخصیص یابد. در این صورت ویژگی‌های زیر برقرار است.

الف) جای کدها را می‌توان تغییر داد. ب) این اعداد اسماً عدد هستند نه واقعاً، زیرا می‌توان برای این رنگ‌ها هر کد دیگر را در نظر گرفت.

ج) روی این اعداد نمی‌توان عملیات ریاضی انجام داد. در اینجا ۲ + ۱ بی‌معنی است.

۲ـ مقیاس ترتیبی: از این مقیاس نیز برای اندازه‌گیری متغیرهای کیفی استفاده می‌شود اما در این مقیاس کدهای درنظر گرفته‌شده برای گزینه‌های یک سؤال یا یک متغیر دارای اهمیت است. از این مقیاس برای اندازه‌گیری متغیرهایی (سؤالاتی) نظیر سطح طبقات اجتماعی افراد، فرهنگ خانواده، میزان علاقه به یک موضوع خاص، میزان آگاهی و موارد مشابه استفاده می‌شود.

به عنوان مثال، اگر گزینه‌های مربوط به سؤال علاقه به سینما را به صورت: ۱ـ بی‌علاقه، ۲ـ تقریباً علاقه‌مند، ۳ـ علاقه‌مند و ۴ـ بسیار علاقه‌مند، کدبندی کنیم، در این صورت ویژگی‌های مقابل برقرار است: الف) این اعداد اسمی هستند و واقعی نیستند. ب) نمی‌توان ترتیب اعداد را به هم زد یا جای اعداد را عوض کرد.

ج) نمی‌توان اغلب عملیات ریاضی را بر روی آن‌ها انجام داد.

۳ـ مقیاس فاصله‌ای: از این مقیاس برای اندازه‌گیری متغیرهای کمّی استفاده می‌شود. این مقیاس اغلب وقتی به کار می‌رود که اندازه‌ها یا مقادیر متغیر به‌صورت فاصله بیان شده باشند و طول این فاصله‌ها برابر باشد. به عنوان مثال ۰ تا ۳، ۳ تا ۶ و ۶ تا ۹.

برای این مقیاس ویژگی‌های زیر برقرار است:

الف) تنها جمع و تفریق انجام می‌شود. ضرب و تقسیم برای این مقیاس قابل انجام نیست. ب) صفر در این مقیاس قراردادی است و ذاتی نیست.

ج) صفر به معنای عدم وجود نیست بلکه یک داده به مقدار صفر موجود است. به‌عنوان مثال ۱ لیتر آب با دمای صفر درجه به معنای این نیست که آب دما ندارد بلکه به معنای وجود دما به میزان صفر است.

۴ـ مقیاس نسبتی: از این مقیاس برای اندازه‌گیری متغیرهای کمیّ نظیر قد، وزن، سن، سرعت، مسافت، فشار، حجم، مساحت و موارد مشابه استفاده می‌شود و ویژگی‌های زیر برقرار است:

۱ـ چهار عمل اصلی قابل انجام است. ۲ـ صفر، مبدأ واقعی است و قراردادی نیست.

به عنوان مثال متحرکی با سرعت صفر به معنای عدم وجود سرعت برای این متحرک می‌باشد.

مثال ۳: کدام‌یک از مقیاس‌ها دارای صفر قراردادی است؟ (مدیریت و حسابداری ـ سراسری ۹۴)

۱) نسبی ۲) فاصله‌ای ۳) اسمی ۴) رتبه‌ای

☑ **پاسخ:** گزینه «۲» بر طبق تعریف مقیاس‌های اندازه‌گیری متغیرها، در مقیاس فاصله‌ای صفر، قراردادی است.

مثال ۴: درخصوص مقیاس داده‌ها کدام گزینه صحیح است؟ (علوم اقتصادی ـ سراسری ۱۴۰۰)

۱) داده‌ها در مقیاس فاصله‌ای و ترتیبی می‌توانند غیرعددی نیز باشند.

۲) داده‌ها در مقیاس اسمی و ترتیبی می‌توانند غیرعددی نیز باشند.

۳) داده‌ها در مقیاس فاصله‌ای و نسبتی، غیرعددی هستند.

۴) داده‌ها در مقیاس اسمی و ترتیبی تنها عددی هستند.

☑ **پاسخ:** گزینه «۲» از مقیاس‌های اسمی و ترتیبی برای اندازه‌گیری داده‌های کیفی استفاده می‌شود و از مقیاس‌های فاصله‌ای و نسبتی برای اندازه‌گیری داده‌های کمّی استفاده می‌شود.

مطالعه توصیفی داده‌ها (آمار توصیفی)

نماد سیگما (Σ)

از نماد Σ (سیگما) برای ساده‌نویسی عمل جمع به شکل زیر استفاده می‌شود.

$$x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$$

به عنوان مثال:

$$x_1 + x_2 + x_3 + x_4 = \sum_{i=1}^{4} x_i$$

ویژگی‌های Σ :

$$۱)\ \sum_{i=۱}^{n} k = \underbrace{k+k+\cdots+k}_{n\text{ مرتبه}} = nk \qquad ۲)\ \sum_{i=۱}^{n} ax_i = a\sum_{i=۱}^{n} x_i$$

$$۳)\ \sum_{i=۱}^{n}(x_i+y_i) = \sum_{i=۱}^{n} x_i + \sum_{i=۱}^{n} y_i \qquad ۴)\ \sum_{i=۱}^{n}(ax_i+b) = a\sum_{i=۱}^{n} x_i + nb$$

$$۵)\ \sum_{i=۱}^{n} x_i^۲ = x_۱^۲ + x_۲^۲ + \cdots + x_n^۲ \qquad ۶)\ (\sum_{i=۱}^{n} x_i)^۲ = (x_۱ + x_۲ + \cdots + x_n)^۲$$

$$۷)\ \sum(ax_i + by_i) = a\sum_{i=۱}^{n} x_i + b\sum_{i=۱}^{n} y_i$$

(a و b و k اعداد ثابت هستند.)

مثال۵: حاصل $A = \sum_{i=۱}^{۴}(۲x_i + ۳)$ را به ازای x_i : ۵,۲,۱,۷ بیابید.

☑ **پاسخ:** در اینجا $x_۱ = ۵\ ,\ x_۲ = ۲\ ,\ x_۳ = ۱$ و $x_۴ = ۷$ است. لذا حاصل عبارت $A = \sum_{i=۱}^{۴}(۲x_i + ۳)$ با توجه به رابطه $\sum(ax_i + b) = a\sum x_i + nb$ برابر است با:

$$A = ۲\sum_{i=۱}^{۴} x_i + ۴(۳) = ۲(۵+۲+۱+۷) + ۱۲ = ۴۲$$

جدول توزیع فراوانی (طبقه‌بندی مشاهدات)

اگر تعداد داده‌ها و مشاهدات زیاد باشد برای تفسیر آسان اطلاعات، آن‌ها را طبقه‌بندی می‌کنیم. برای طبقه‌بندی داده‌های $x_۱, x_۲, \ldots, x_N$ مراحل زیر را انجام می‌دهیم: (N برابر با تعداد داده‌ها است.)

۱ـ دامنه تغییرات را با رابطه مقابل حساب می‌کنیم: $R = \max_i x_i - \min_i x_i$

۲ـ تعداد طبقات را با دستور $k = \sqrt{N}$ یا $k = ۱ + ۳/۳۲۲\log N$ تعیین می‌کنیم.

۳ـ فاصله طبقات یا طول هر طبقه را با رابطه زیر مشخص می‌کنیم: $I = \frac{R}{K} = \frac{\text{دامنه تغییرات}}{\text{تعداد طبقات}}$

۴ـ حدود طبقات را به صورت فاصله $[a,b)$ در نظر می‌گیریم. دقت نمایید حد پایین طبقه اول را معمولاً کوچک‌ترین داده قرار می‌دهند سپس با اضافه کردن فاصله طبقات (I) به آن، حد بالای همان طبقه به دست می‌آید.

همچنین حد پایین طبقه بعد برابر حد بالای طبقه قبل در نظر گرفته می‌شود. به مثال زیر توجه فرمایید.

مثال۶: داده‌های زیر را طبقه‌بندی نمایید.

۲۰ – ۷ – ۱۵ – ۱۶ – ۹ – ۲۳ – ۲۹ – ۷ – ۱۳ – ۱۶ – ۱۷ – ۱۸ – ۲۶ – ۸ – ۱۵ – ۱۷ – ۲۰ – ۴ – ۱۲ – ۲۸ – ۱۳ – ۱۱ – ۱۹ – ۱۲ – ۱۵

☑ **پاسخ:** برای طبقه‌بندی تعداد $N = ۲۵$ داده، مراحل زیر را انجام می‌دهیم:

۱) دامنه تغییرات را با رابطه $R = \max_i x_i - \min_i x_i$ مشخص می‌کنیم. در اینجا بزرگ‌ترین داده برابر ۲۹ و کوچک‌ترین داده برابر ۴ است. بنابراین مقدار دامنه تغییرات برابر است با: $R = ۲۹ - ۴ = ۲۵$

۲) تعداد طبقات را با رابطه $k = \sqrt{N}$ به دست می‌آوریم. $N = ۲۵ \Rightarrow k = \sqrt{۲۵} = ۵$

۳) فاصله طبقات را با دستور $I = \frac{R}{K}$ حساب می‌کنیم: $R = ۲۵\ ,K = ۵ \Rightarrow I = \frac{۲۵}{۵} = ۵$

اکنون با انتخاب کوچک‌ترین داده یعنی عدد ۴ به‌عنوان حد پایین طبقه اول، جدول طبقه‌بندی را در تعداد $k = ۵$ طبقه به شکل زیر تشکیل می‌دهیم:

شماره طبقه	۱	۲	۳	۴	۵
حدود طبقات	۴–۹	۹–۱۴	۱۴–۱۹	۱۹–۲۴	۲۴–۲۹

دقت نمایید برای طبقه اول مقدار $I = ۵$ (فاصله طبقات) را به حد پایین اضافه نمودیم و حد بالای طبقه اول برابر $۴ + ۵ = ۹$ به دست آمده است.

حد پایین طبقه دوم برابر حد بالای طبقه اول یعنی ۹ در نظر گرفته می‌شود. مجدداً به حد پایین هر طبقه، فاصله طبقات یعنی $I = ۵$ اضافه می‌شود تا حد بالای همان طبقه به دست آید.

پارامترهای جدول طبقه‌بندی

جدول طبقه‌بندی داده‌ها، ساده‌ترین و متداول‌ترین جدول آماری است که شامل پارامترهای زیر می‌باشد.

فراوانی مطلق: فراوانی مطلق طبقه iام را با F_i نشان می‌دهیم و برابر تعداد مشاهدات موجود در طبقه iام است.

نکته۱: حاصل‌جمع فراوانی‌های طبقات برابر تعداد کل مشاهدات (N) است. $F_1 + F_2 + \cdots + F_k = N$

فراوانی نسبی: اگر فراوانی مطلق هر طبقه را به تعداد کل مشاهدات تقسیم کنیم، فراوانی نسبی آن طبقه به دست می‌آید. فراوانی نسبی طبقه iام را با f_i نشان می‌دهیم:

$$f_i = \frac{F_i}{N} \quad ; \quad i = 1,\dots,k$$

نکته۲: حاصل‌جمع فراوانی نسبی طبقات برابر با عدد ۱ است. $f_1 + f_2 + \cdots + f_k = 1$

درصد فراوانی نسبی: برای به دست آوردن درصد فراوانی نسبی هر طبقه، کافی است فراوانی نسبی هر طبقه را در عدد ۱۰۰ ضرب کنیم.

$$\text{درصد فراوانی نسبی طبقه } i\text{ام} = f_i \times 100 \quad ; \quad i = 1,\dots,k$$

فراوانی تجمعی: فراوانی تجمعی طبقه iام (Fc_i)، برابر حاصل‌جمع فراوانی مطلق از طبقه اول تا طبقه iام است.

$$Fc_i = \sum_{j=1}^{i} F_j$$

نکته۳: فراوانی تجمعی آخرین طبقه برابر تعداد کل مشاهدات (N) است. یعنی: $Fc_k = N$

فراوانی نسبی تجمعی: فراوانی نسبی تجمعی طبقه iام (fc_i) از تقسیم فراوانی تجمعی طبقه iام (Fc_i) به کل مشاهدات (N) به دست می‌آید.

$$fc_i = \frac{Fc_i}{N} \quad ; \quad i = 1,\dots,k$$

نکته۴: فراوانی نسبی تجمعی طبقه آخر برابر با عدد ۱ است، یعنی: $fc_k = 1$

درصد فراوانی نسبی تجمعی: اگر فراوانی نسبی تجمعی هر طبقه را در عدد ۱۰۰ ضرب کنیم، درصد فراوانی نسبی تجمعی به دست می‌آید.

$$\text{درصد فراوانی نسبی تجمعی} = fc_i \times 100 \quad ; \quad i = 1,\dots,k$$

نماینده طبقات: نماینده طبقه iام را با x_i نشان می‌دهیم و به‌صورت مقابل به دست می‌آید:

$$x_i = \frac{\text{حد بالا} + \text{حد پایین}}{2} \quad ; \quad i = 1,2,\dots,k$$

مثال۷: برای داده‌های طبقه‌بندی شده، فراوانی مطلق، فراوانی نسبی، درصد فراوانی نسبی، فراوانی تجمعی، فراوانی نسبی تجمعی، درصد فراوانی نسبی تجمعی و نماینده طبقات را بیابید.

۲۰ – ۷ – ۱۵ – ۱۶ – ۹ – ۲۳ – ۲۹ – ۷ – ۱۳ – ۱۶ – ۱۷ – ۱۸ – ۲۶ – ۸ – ۱۵ – ۱۷ – ۲۰ – ۴ – ۱۲ – ۲۸ – ۱۳ – ۱۱ – ۱۹ – ۱۲ – ۱۵

☑ **پاسخ:**

نماینده طبقات	درصد فراوانی نسبی تجمعی	فراوانی نسبی تجمعی fc_i	فراوانی تجمعی Fc_i	درصد فراوانی نسبی	f_i	F_i	حدود طبقات
۶/۵	۱۶	$\frac{۴}{۲۵} = ۰/۱۶$	۴	۱۶	$\frac{۴}{۲۵} = ۰/۱۶$	۴	۴-۹
۱۱/۵	۴۰	$\frac{۱۰}{۲۵} = ۰/۴۰$	۱۰	۲۴	$\frac{۶}{۲۵} = ۰/۲۴$	۶	۹-۱۴
۱۶/۵	۷۲	$\frac{۱۸}{۲۵} = ۰/۷۲$	۱۸	۳۲	$\frac{۸}{۲۵} = ۰/۳۲$	۸	۱۴-۱۹
۲۱/۵	۸۸	$\frac{۲۲}{۲۵} = ۰/۸۸$	۲۲	۱۶	$\frac{۴}{۲۵} = ۰/۱۶$	۴	۱۹-۲۴
۲۶/۵	۱۰۰	۱	۲۵	۱۲	$\frac{۳}{۲۵} = ۰/۱۲$	۳	۲۴-۲۹
						$N = ۲۵$	

مدرسان شریف

فصل دوم

«تئوری احتمال»

درسنامه (۱): آنالیز ترکیبی

مقدمه

نیاز به آمار به دلیل رشد و پیشرفت آن در بسیاری از شاخه‌ها از قبیل علوم مهندسی، اقتصاد، مدیریت و ... احساس می‌شود و دیگر به آمار به عنوان مجموعه‌ای از داده‌های خام دسته‌بندی شده، نگاه نمی‌شود. امروزه پژوهشگران براساس مشاهدات و داده‌های جمع‌آوری شده با کمک گرفتن از علم آمار در برخورد با عدم قطعیت در وقوع رخدادها و اتفاقات اقدام به استنباط و تصمیم‌گیری می‌کنند.

موضوع عدم قطعیت، دامنه وسیعی از مسائل روزمره را دربرمی‌گیرد. به عنوان مثال، کارشناس بیمه در زمان تعیین حق بیمه، پژوهشگر دارو به هنگام آزمایش مواد افزوده شده به دارو یا یک اقتصاددان برای پیش‌بینی رویدادهای اقتصادی به نوعی با عدم قطعیت در وقوع رخدادها مواجه هستند. مطالبی که در این فصل ارائه شده است شامل اصول و تکنیک‌های شمارش، تعریف مقدماتی احتمال، احتمال شرطی، قضایای احتمال، قانون احتمال کل و دستور بیز همراه با مثال‌های متنوع آزمون‌های سراسری است.

اصل شمارش ضرب (اصل اساسی شمارش)

فرض کنید آزمایشی در دو مرحله قابل انجام باشد به‌طوری که اولین مرحله به m طریق ممکن و در مرحله دوم برای هر طریق مرحله اول، n طریق ممکن وجود داشته باشد. آنگاه برای انجام دو مرحله با هم، $m \times n$ طریق ممکن وجود دارد.

توجه: کلمه «و» در اصل شمارش، متناظر ضرب است.

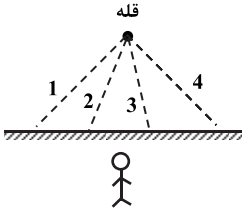
مثال ۱: شخصی دارای ۵ پیراهن و ۳ شلوار است. این شخص به چند طریق می‌تواند، لباس‌های متفاوت بپوشد؟

۱) ۲۵ ۲) ۱۰ ۳) ۸ ۴) ۱۵

☑ **پاسخ:** گزینه «۴» در اینجا تهیه لباس شامل دو مرحله انتخاب پیراهن و انتخاب شلوار است. در مرحله اول شخص به $m = 5$ طریق می‌تواند یکی از ۵ پیراهن را انتخاب نماید. در مرحله دوم به $n = 3$ طریق می‌تواند یکی از ۳ شلوار را انتخاب کند. بنابراین طبق اصل شمارش ضرب، این شخص می‌تواند به $m \times n = 5 \times 3 = 15$ طریق لباس‌های متفاوت بپوشد.

مثال ۲: ۴ جاده به قله کوهی می‌رود، یک کوهنورد به چند طریق می‌تواند بالا رفته و پایین بیاید به شرط آنکه مسیر رفت و برگشت یکی نباشد؟

☑ **پاسخ:** کوهنورد عزیز ما از $n = 4$ مسیر به قله کوه می‌تواند برسد ولی چون مسیر رفت و برگشت نباید یکی باشد بنابراین فقط از $m = 3$ مسیر می‌تواند از قله کوه پایین بیاید؛ در نتیجه تعداد حالت‌ها برابر $n \times m = 3 \times 4 = 12$ است.



توجه: برای درک بهتر توجه شما را به شکل روبرو جلب می‌کنم. کوهنورد ما از یکی از مسیرهای ۱ یا ۲ یا ۳ یا ۴ به قله می‌رسد. فرض کنید از مسیر ۲ به قله کوه رسیده باشد، چون نمی‌تواند از مسیر ۲ برای برگشت استفاده کند در نتیجه فقط می‌تواند یکی از مسیرهای ۱ یا ۳ یا ۴ انتخاب کند.

تعمیم اصل شمارش ضرب

فرض کنید یک آزمایش در k مرحله قابل انجام باشد به‌طوری که:

(مرحله kام به n_k طریق) و... و (مرحله دوم به n_2 طریق) و (مرحله اول به n_1 طریق)

بتواند انجام شود. در این صورت تعداد کل راه‌های انجام این k مرحله برابر است با: $n_1 \times n_2 \times \ldots \times n_k$

برای درک بهتر، به مثال‌های زیر توجه کنید.

مثال ۳: فرض کنید A، B، C، D و E پنج شهر باشند که مطابق شکل زیر توسط جاده‌هایی به‌هم وصل شده‌اند، به چند طریق می‌توان از A به E با عبور از شهرهای B، C و D سفر کرد؟

۱) ۲۷ ۲) ۱۴ ۳) ۲۸ ۴) ۲۴

☑ **پاسخ:** گزینه «۴» مسافرت از A به E، در ۴ مرحله انجام می‌شود. در مرحله اول سفر از A به B به $n_1 = 3$ طریق، در مرحله دوم سفر از B به C به $n_2 = 2$ طریق، در مرحله سوم سفر از C به D به $n_3 = 4$ طریق و در مرحله چهارم سفر از D به E به $n_4 = 1$ طریق قابل انجام است.

بنابراین طبق اصل شمارش ضرب تعداد راه‌های مسافرت از A به E برابر است با:

$$n_1 \times n_2 \times n_3 \times n_4 = 3 \times 2 \times 4 \times 1 = 24$$

مثال ۴: اگر دو تاس و یک سکه پرتاب شود، چند حالت ممکن می‌تواند رخ دهد؟

۱) ۴۸ ۲) ۸۱ ۳) ۷۲ ۴) ۱۴

☑ **پاسخ:** گزینه «۳» در اینجا آزمایش شامل ۳ مرحله پرتاب تاس اول و پرتاب تاس دوم و پرتاب سکه می‌باشد. در پرتاب تاس اول یکی از اعداد ۱، ۲، ۳، ۴، ۵ و ۶ ظاهر می‌گردد. پس در مرحله اول $n_1 = 6$ برآمد ممکن و مشابهاً در پرتاب تاس دوم در مرحله دوم $n_2 = 6$ برآمد ممکن وجود دارد و در مرحله سوم که شامل پرتاب سکه است، $n_3 = 2$ برآمد وجود دارد.

بنابراین طبق اصل شمارش ضرب، تعداد کل برآمدها برابر است با: $n_1 \times n_2 \times n_3 = 6 \times 6 \times 2 = 72$

نکات مربوط به شمارش اعداد یا تشکیل کدها، رمزها و سریال‌ها:

برای شمارش اعداد یا ساخت کدها، رمزها و سریال‌ها توجه شما را به نکات زیر جلب می‌کنم.

۱) اگر هدف مسئله تشکیل عدد چندرقمی باشد، صفر نمی‌تواند در اولین رقم سمت چپ قرار بگیرد؛ به عنوان مثال ۰۷۵ عدد سه رقمی نیست.

۲) اگر هدف تشکیل کد، رمز یا سریال چندرقمی باشد، صفر می‌تواند در اولین رقم سمت چپ قرار بگیرد، به‌عنوان مثال ۰۲۱ یک کد سه رقمی است.

۳) در ساخت کدها، رمزها و اعداد از سمت چپ به راست حرکت می‌کنیم مگر آنکه در صورت مسئله شرطی مانند فرد بودن، زوج بودن و در حالت کلی شرطی که به ارقام سمت راست مربوط باشد ذکر شده باشد، در این حالت باید ابتدا شرایط داده شده در مسئله مانند فرد بودن یا زوج بودن را لحاظ کنیم سپس مجدداً از سمت چپ به سمت راست حرکت کنیم.

۴) کدها، رمزها یا اعداد می‌توانند دارای ارقام تکراری باشند یا نباشند، مانند «۳۳۱» یا «۱۲۳». اگر در صورت سؤال درباره مجاز بودن یا نبودن تکرار ارقام مطلبی گفته نشود و شرطی بیان نشود پیش‌فرض آن است که تکرار ارقام مجاز است.

مثال ۵: با استفاده از ارقام $\{1,2,3,4\}$ چند عدد چهار رقمی بدون تکرار می‌توان نوشت؟

۱) ۲۴ ۲) ۴۸ ۳) ۳۶ ۴) ۱۲

☑ **پاسخ:** گزینه «۱» فرض کنید عدد چهاررقمی مطابق خانه‌های A، B، C و D مقابل باشد. با توجه به توضیحات داده شده از سمت چپ به راست حرکت می‌کنیم. چون در خانه A یکی از ارقام $\{1,2,3,4\}$ را می‌توان قرار داد پس برای خانه A دارای ۴ انتخاب هستیم. دقت کنید چون تکرار مجاز نیست بنابراین برای خانه‌ی B دارای ۳ انتخاب و به همین ترتیب برای خانه‌ی C دارای ۲ انتخاب و برای خانه D دارای ۱ انتخاب هستیم. در نتیجه با استفاده از اصل شمارش ضرب تعداد اعداد چهاررقمی خواسته شده برابر $4 \times 3 \times 2 \times 1 = 24$ است.

A	B	C	D
☐	☐	☐	☐

مثال ۶: با استفاده از ارقام $\{0,1,2,3\}$ چند عدد چهاررقمی بدون تکرار می‌توان نوشت؟

۱) ۲۴ ۲) ۳۲ ۳) ۱۸ ۴) ۴۸

A	B	C	D

پاسخ: گزینه «۳» فرض کنید عدد ۴ رقمی مطابق خانه‌های D,C,B,A مقابل باشد. با توجه به توضیحات داده شده از سمت چپ به راست حرکت می‌کنیم، دقت کنید در خانه A دارای ۳ انتخاب هستیم، به عبارت دیگر یکی از ارقام $\{1,2,3\}$ را می‌توان در خانه A قرار داد ولی در خانه B صفر را می‌توانیم قرار بدهیم. بنابراین برای این خانه هم با توجه به مجاز نبودن تکرار دارای ۳ انتخاب هستیم. برای خانه C، ۲ انتخاب و برای خانه D دارای یک انتخاب هستیم در نتیجه با استفاده از اصل شمارش ضرب تعداد اعداد چهاررقمی خواسته شده برابر $3\times3\times2\times1=18$ است.

اصل شمارش جمع

فرض کنید کاری را بتوان در دو حالت A یا B انجام داد. اگر عمل A به m طریق و عمل B به n طریق انجام شود و رخ دادن هم‌زمان این دو عمل ممکن نباشد، آنگاه این کار به $m+n$ طریق انجام می‌پذیرد. تفاوت اصل شمارش جمع با اصل شمارش ضرب این است که در اصل شمارش ضرب عمل‌ها به‌صورت سری و پی‌درپی انجام می‌شود در حالی‌که در اصل شمارش جمع عمل‌ها به‌صورت موازی انجام می‌پذیرد. در اصول شمارش «کلمه «و» متناظر با ضرب و کلمه «یا» متناظر جمع است.» به مثال‌های زیر توجه نمایید.

مثال ۷: به چند طریق می‌توان از بین ۶ دانشجوی برق و ۵ دانشجوی مکانیک و ۴ دانشجوی کامپیوتر دو دانشجو انتخاب نمود طوری‌که هر دو نفر انتخابی از یک رشته نباشند؟

۱) ۷۴ ۲) ۶۸ ۳) ۸۴ ۴) ۶۶

پاسخ: گزینه «۱» به‌دلیل اینکه دو دانشجو نباید از یک رشته باشند، بنابراین لازم است انتخاب‌ها به‌صورت زیر باشد.

«یکی کامپیوتر و یکی مکانیک» یا «یکی کامپیوتر و یکی برق» یا «یکی مکانیک و یکی برق»

به $6\times5=30$ حالت یک دانشجوی برق و یک دانشجوی مکانیک انتخاب می‌شود. به $6\times4=24$ حالت یک دانشجوی برق و یک دانشجوی کامپیوتر انتخاب می‌شود. به $5\times4=20$ حالت یک دانشجوی مکانیک و یک دانشجوی کامپیوتر انتخاب می‌شود. اکنون تعداد راه‌های انتخاب خواسته شده طبق اصل شمارش جمع برابر است با: $30+24+20=74$

مثال ۸: چند عدد زوج سه رقمی با ارقام متمایز وجود دارد؟

۱) ۲۵۶ ۲) ۳۲۸ ۳) ۴۰۵ ۴) ۳۶۰

پاسخ: گزینه «۲» هر عدد سه رقمی زوج به کمک ارقام $\{0,1,2,\ldots,9\}$ نوشته می‌شود. فرض کنید عدد سه رقمی مطابق خانه‌های A، B و C باشد. اگر عدد سه رقمی خواسته شده زوج باشد، لازم است در خانه C یکی از ارقام $0,2,4,6$ یا ۸ قرار بگیرد. اگر صفر در خانه C قرار نگیرد، در خانه A نیز نمی‌تواند باشد، لذا دو حالت زیر را در نظر می‌گیریم.

A	B	C

حالت اول: عدد صفر در خانه C قرار بگیرد. که این به یک طریق امکان‌پذیر است. اکنون در خانه A یکی از ارقام $\{1,2,3,\ldots,9\}$ را می‌توان قرار داد که این به ۹ طریق امکان‌پذیر است. چون تکرار ارقام مجاز نیست برای خانه B، ۸ انتخاب داریم در نتیجه تعداد اعداد سه رقمی زوج با ارقام متمایز در این حالت برابر $9\times8\times1=72$ است.

A	B	C
۹	۸	۱

حالت دوم: در خانه C یکی از ارقام $\{2,4,6,8\}$ قرار بگیرد و عدد صفر در خانه C قرار نگیرد.

بنابراین برای خانه C، ۴ انتخاب داریم و چون رقم صفر نمی‌تواند در خانه A قرار بگیرد، پس برای خانه A، ۸ انتخاب باقی می‌ماند. همچنین برای خانه B عدد صفر می‌تواند قرار بگیرد پس برای خانه B نیز ۸ انتخاب داریم. در نتیجه تعداد اعداد سه رقمی با ارقام متمایز در این حالت برابر $8\times8\times4=256$ می‌شود.

A	B	C
۸	۸	۴

اکنون با استفاده از اصل شمارش جمع تعداد کل اعداد سه رقمی مدنظر برابر $256+72=328$ می‌شود.

مثال۹: با ارقام ۵، ۴، ۳، ۲، ۱ و ۰، چند عدد سه رقمی مضرب ۵ بدون تکرار رقم‌ها می‌توان نوشت؟

۱) ۳۰ ۲) ۳۲ ۳) ۳۶ ۴) ۴۲

پاسخ: گزینه «۳» اعدادی مضرب ۵ هستند که رقم یکان یا همان خانه C صفر یا ۵ باشد. با توجه به اینکه در خانه C یکی از ارقام $\{0,5\}$ را می‌توان قرار داد دو حالت در نظر می‌گیریم.

A B C

حالت اول: فرض می‌کنیم رقم یکان یا همان خانه C صفر باشد. با توجه به مجاز نبودن تکرار ارقام وقتی صفر را در خانه C قرار دادیم برای خانه A دارای ۵ انتخاب هستیم. به‌عبارت دیگر در خانه A یکی از ارقام $\{1,2,3,4,5\}$ را می‌توان قرار داد. وقتی یکی از این ارقام را در خانه A قرار بدهیم برای خانه B دارای ۴ انتخاب هستیم و چون صفر را در خانه C قرار دادیم پس برای خانه C دارای یک انتخاب می‌باشیم. با توجه به اصل شمارش ضرب تعداد اعداد سه رقمی که مضرب ۵ هستند و رقم یکان آن‌ها صفر است برابر $5\times4\times1=20$ است.

حالت دوم: فرض می‌کنیم رقم یکان یا همان خانه C عدد ۵ باشد. در خانه A نمی‌توان صفر را قرار داد و از طرفی عدد ۵ را در خانه C قرار دادیم در نتیجه برای خانه A دارای ۴ انتخاب هستیم؛ به عبارت دیگر در خانه A یکی از ارقام $\{1,2,3,4\}$ را می‌توان قرار داد. دقت کنید در خانه B عدد صفر را می‌توانیم قرار بدهیم در نتیجه برای خانه B هم دارای ۴ انتخاب هستیم. با توجه به توضیحات داده شده، تعداد اعداد سه رقمی مضرب ۵ که رقم یکان آنها عدد ۵ است طبق اصل شمارش ضرب برابر $4\times4\times1=16$ است. در پایان برای محاسبه کل تعداد اعداد خواسته شده در صورت سؤال باید دو مقداری را که از حالت‌های اول و دوم به دست آوردیم طبق اصل جمع با یکدیگر جمع کنیم در نتیجه تعداد کل برابر $16+20=36$ است.

مثال۱۰: چند کلمه ۵ حرفی با استفاده از حروف a و b و c می‌توان نوشت به طوری که هیچ دو حرف مجاور یکسان نباشد؟

۱) ۵۲ ۲) ۴۸ ۳) ۶۴ ۴) ۳۲

پاسخ: گزینه «۲» فرض کنید کلمه ۵ حرفی مطابق خانه‌های A، B، C، D، E باشد. برای خانه A یکی از سه حرف $\{a,b,c\}$ به ۳ طریق انتخاب می‌شود با توجه به شرط مسئله اگر برای خانه A حرف a را انتخاب کنیم

A B C D E

در این صورت برای خانه B یکی از حروف $\{b,c\}$ به ۲ طریق انتخاب می‌شود. اکنون اگر برای خانه B، b انتخاب می‌شود آنگاه برای خانه C یکی از حروف $\{a,c\}$ به ۲ طریق انتخاب می‌شود. با همین استدلال برای خانه D، ۲ انتخاب و برای خانه E نیز ۲ انتخاب داریم.

در نتیجه تعداد کل کلمات ۵ حرفی با استفاده از اصل شمارش ضرب برابر $3\times2\times2\times2\times2=48$ است.

مثال۱۱: چند عدد سه رقمی وجود دارد که حداقل یک رقم آن عدد ۱ باشد؟

۱) ۹۰ ۲) ۱۸۰ ۳) ۲۷۰ ۴) ۲۵۲

پاسخ: گزینه «۴» برای حل این سؤال با توجه به اینکه کلمه حداقل ذکر شده کافی است تعداد اعداد ۳ رقمی بدون ۱ از تعداد کل اعداد ۳ رقمی کم شود. برای این منظور تعداد اعداد سه رقمی که اصلاً ۱ ندارد به‌صورت زیر محاسبه می‌شود.

A B C

اعداد ۳ رقمی خواسته شده با استفاده از ارقام $\{0,2,3,4,5,6,7,8,9\}$ ساخته می‌شوند در خانه A صفر را نمی‌توان قرار داد بنابراین برای خانه A دارای ۸ انتخاب هستیم؛ به عبارت دیگر یکی از ارقام $\{2,3,4,5,6,7,8,9\}$ را می‌توان در خانه A قرار داد.

دقت کنید وقتی کلمه حداقل ذکر شده عددی که شامل دوتا یک هم است قابل قبول است؛ مانند عدد ۳۱۱. با توجه به این توضیحات نتیجه می‌گیریم که تکرار مجاز است بنابراین برای خانه B دارای ۹ انتخاب و برای خانه C نیز دارای ۹ انتخاب هستیم. در نتیجه طبق اصل شمارش ضرب تعداد اعداد سه رقمی که اصلاً ۱ ندارند برابر $8\times9\times9=648$ است.

اکنون تعداد کل اعداد سه رقمی را شمارش می‌کنیم. تعداد کل اعداد سه رقمی با ارقام $\{0,1,2,...,9\}$ نوشته می‌شود. برای خانه A، ۹ انتخاب، و برای خانه B، ۱۰ انتخاب و برای خانه C نیز ۱۰ انتخاب داریم. لذا تعداد کل اعداد سه رقمی برابر $9\times10\times10=900$ می‌شود. در نتیجه تعداد اعداد سه رقمی که حداقل یک رقم آن یک باشد برابر است با:

اعداد سه رقمی که حداقل یک رقم آن‌ها ۱ است = اعدادی سه رقمی − کل اعداد سه رقمی که اصلاً ۱ ندارند $=900-648=252$

فاکتوریل

❖ **تعریف:** حاصل‌ضرب اعداد طبیعی متوالی از ۱ تا n را n فاکتوریل می‌گویند و با نماد $n!$ نشان می‌دهند. $n! = n \times (n-1) \times \cdots \times ۳ \times ۲ \times ۱$

به‌عنوان مثال: $۴! = ۴ \times ۳ \times ۲ \times ۱ = ۲۴$

تذکر: $۱! = ۱ , ۰! = ۱$

نکته۱: $k! = k \times \underbrace{(k-۱) \times (k-۲) \times \cdots \times ۲ \times ۱}_{(k-۱)!} = k(k-۱)!$

به‌عنوان مثال: $۱۸! = ۱۸ \times ۱۷!$

جایگشت (تبدیل)

به تعداد راه‌های قرار گرفتن اشیای متمایز کنارهم، جایگشت (تبدیل) می‌گویند. جایگشت انواع مختلف دارد که در ادامه توضیح داده شده است.

جایگشت خطی

منظور از جایگشت خطی n شیء متمایز، یعنی n شیء متمایز را به چند طریق می‌توان کنار هم در یک صف قرار داد. به عنوان مثال سه حرف a، b و c را به شش حالت (بدون تکرار) می‌توان کنار هم قرار داد. $abc, acb, bac, bca, cab, cba$

در حالت کلی، طبق اصل ضرب n شیء متمایز را بدون تکرار به $n!$ حالت می‌توان کنار هم در یک صف قرار داد. اگر تکرار اشیاء مجاز باشد تعداد جایگشت‌ها برابر n^n خواهد بود.

مثال۱۲: با حروف کلمه Black چند رمز پنج حرفی بدون تکرار می‌توان نوشت؟

۱) ۱۲۰ ۲) ۲۴۰ ۳) ۳۶۰ ۴) ۴۸۰

☑ **پاسخ:** گزینه «۱» **روش اول:** تعداد رمزهای پنج حرفی با استفاده از حروف $\{B-l-a-c-k\}$ معادل این است که به چند طریق می‌توان این پنج حرف را بدون تکرار کنار هم قرار داد. با استفاده از دستور جایگشت خطی تعداد آن‌ها برابر $n! = ۵! = ۱۲۰$ است.

روش دوم: از سمت چپ به راست حرکت می‌کنیم:

A	B	C	D	E
□	□	□	□	□

$۵ \times ۴ \times ۳ \times ۲ \times ۱ = ۱۲۰$

جایگشت یک در میان

می‌خواهیم تعداد راه‌های قرار گرفتن m شیء متمایز از یک گروه و n شیء متمایز از گروه دیگر را به‌صورت یک در میان کنار هم مشخص کنیم. واضح است این کار وقتی قابل انجام است که $m = n$ باشد یا m و n به اندازه یک واحد اختلاف داشته باشند:

حالت اول: اگر تعداد دو گروه با هم برابر باشند: اگر $m = n \Rightarrow$ تعداد حالت‌ها $= ۲ \times m! \times n!$

حالت دوم: تعداد اشیای دو گروه به اندازه یک واحد با یکدیگر اختلاف داشته باشند: اگر $m = n + ۱ \Rightarrow$ تعداد حالت‌ها $= m! \times n!$

برای درک بهتر به مثال‌های زیر توجه کنید.

مثال۱۳: ۳ دختر و ۳ پسر به چند حالت می‌توانند به صورت یک در میان در کنار هم قرار بگیرند؟

۱) ۳۶ ۲) ۷۲ ۳) ۱۲۸ ۴) ۲۴

☑ **پاسخ:** گزینه «۲» تعداد هر دو گروه برابر $m = n = ۳$ است. در نتیجه تعداد حالات خواسته شده برابر $۲ \times ۳! \times ۳! = ۲ \times m! \times n!$ است.

مثال۱۴: ۳ پسر و ۴ دختر به چند طریق می‌توانند به‌صورت یک در میان در کنار هم بنشینند؟

۱) ۵۷۶ ۲) ۳۶ ۳) ۱۴۴ ۴) ۷۲

☑ **پاسخ:** گزینه «۳» $m = ۴$ و $n = ۳$، در نتیجه تعداد حالت‌هایی که این دو گروه می‌توانند به‌صورت یک در میان قرار بگیرند برابر است با:

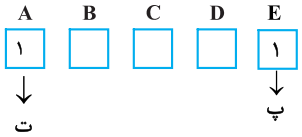
تعداد کل حالت‌ها $= m! \times n! = ۴! \times ۳! = ۱۴۴$

نکته۲: اگر در محاسبه جایگشت n شی، قرار باشد تعدادی از اشیاء در مکان‌های مشخصی قرار بگیرند، ابتدا آن‌ها را در جایگاه‌های از پیش تعیین شده قرار می‌دهیم سپس جایگشت سایر اشیاء را طبق روال عادی حساب می‌کنیم.

مثال ۱۵: با حروف کلمه «پرستو» چند جایگشت ۵ حرفی می‌توان ساخت که با حرف «ت» شروع و به «پ» ختم شود؟

۱) ۲۴ ۲) ۴۸ ۳) ۱۲۰ ۴) ۶

پاسخ: گزینه «۴» با توجه به توضیحات داده شده در بالا و صورت سؤال حرف «ت» را به ۱ حالت در خانه A و حرف «پ» را نیز به ۱ حالت در خانه E قرار می‌دهیم. با قراردادن این دو حرف، سه حرف {ر، و، س} باقی می‌ماند.



در نتیجه برای خانه B دارای ۳ انتخاب و برای خانه C دارای ۲ انتخاب و برای خانه D دارای یک انتخاب هستیم. در نتیجه طبق اصل شمارش ضرب تعداد جایگشت‌ها برابر $۱\times۳\times۲\times۱\times۱=۶$ است.

مثال ۱۶: یک سرمایه‌گذار ۱۰ واحد سرمایه دارد که قصد دارد آن را به ۵ سهم مختلف اختصاص دهد. مقدار سرمایه اختصاص‌یافته به سهم شماره k برابر x_k واحد ($۰\le x_k\le ۱۰$) و عددی صحیح است. همه ۱۰ واحد باید سرمایه‌گذاری شود و سرمایه اختصاص‌یافته به هیچ دو سهمی، با یکدیگر برابر نیست. چند روش برای اختصاص سرمایه به سهم‌ها وجود دارد؟ (مدیریت و حسابداری ـ سراسری ۱۴۰۲)

۱) ۸۰ ۲) ۱۲۰ ۳) ۸۰۰ ۴) ۱۲۰۰

پاسخ: گزینه «۲» فرض کنید $x_۱, x_۲, x_۳, x_۴$ و $x_۵$ نشان‌دهنده تعداد سرمایه تخصیص‌یافته به پنج سهم باشد. بنابراین داریم:

$$x_۱+x_۲+x_۳+x_۴+x_۵=۱۰$$

از طرفی تعداد سرمایه هیچ دو سهم نباید یکسان باشد، لذا در تساوی بالا متغیرها می‌توانند تنها مقادیر $۰,۱,۲,۳,۴$ باشند. یعنی: $x_i=۰,۱,۲,۳,۴$

اکنون تعداد رأس‌های تخصیص پنج عدد $۰,۱,۲,۳,۴$ بین ۵ متغیر بدون تکرار با استفاده از دستور جایگشت برابر است با: $n!=۵!=۱۲۰$

جایگشت n شی متمایز با این فرض که k شیء مشخص در کنار هم باشند.

اگر قرار باشد در محاسبه جایگشت n شیء، k شیء مشخص کنار هم باشند کافی است آن k شیء را درون یک جعبه قرار بدهیم و مانند یک شیء درنظر بگیریم. سپس این جعبه را در کنار سایر اشیاء (یعنی n-k شیء دیگر) قرار می‌دهیم و تعداد جایگشت‌های $n-k+۱$ شیء را حساب کرده و در آخر این عدد به‌دست آمده را در جایگشت k شیء درون جعبه یعنی $k!$ ضرب کنیم.

نکته ۳: با توضیحات داده شده می‌توان نتیجه گرفت برای محاسبه جایگشت n شی متمایز به شرط اینکه k شیء مشخص کنار هم باشند از رابطه $(n-k+۱)!\times k!$ استفاده می‌کنیم.

مثال ۱۷: تعداد جایگشت‌های حروف کلمه «Modern» به طوری که سه حرف (m-n-o) در کنار هم باشند، برابر چند است؟

۱) $۴!\times۳!$ ۲) $۶!$ ۳) $۴!$ ۴) $۳!\times۳!$

پاسخ: گزینه «۱» سه حرف $(m-n-o)$ را در یک بسته قرار می‌دهیم و این بسته را یک شیء در نظر می‌گیریم؛ وقتی این ۳ حرف کنار هم قرار بگیرند ۳ حرف $\{d,e,r\}$ باقی می‌ماند. این ۳ حرف و یک بسته را ۴ شیء متمایز درنظر می‌گیریم. این ۴ شیء به $۴!$ حالت در کنار هم قرار می‌گیرند. از طرفی ۳ حرف $(m-n-o)$ هم به $۳!$ حالت در کنار هم جابه‌جا می‌شوند. در نتیجه طبق اصل شمارش ضرب تعداد حالات برابر $۳!\times۴!$ است.

برای محاسبه جایگشت n شیء متمایز با این فرض که k شیء مشخص در کنار هم نباشند؛ ابتدا تعداد کل جایگشت‌های این n شیء را به دست می‌آوریم سپس تعداد حالاتی که این k شیء کنار هم باشند را محاسبه کرده و از تعداد کل $(n!)$ کم کنیم.

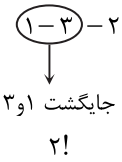
نکته ۴: تعداد حالت‌هایی که n شیء متمایز می‌توانند در کنار هم قرار بگیرند به شرط آنکه k شیء مشخص در کنارهم نباشند برابر است با:

$$n!-(n-k+۱)!\times k!$$

مثال ۱۸: با ارقام ۴ – ۶ – ۳ – ۲ – ۱ چند عدد ۵ رقمی بدون تکرار می‌توان نوشت به شرط آنکه هیچ دو عدد فرد کنار هم نباشد؟

۱) ۶۴ ۲) ۹۶ ۳) ۱۰۸ ۴) ۷۲

پاسخ: گزینه «۴» تعداد کل اعداد ۵ رقمی را که می‌توان نوشت برابر $۵!$ است. تعداد کل اعداد ۵ رقمی به شرط اینکه ۱و۳ کنار هم باشند برابر $۴!\times۲!$ است.



با توجه به توضیحات داده شده در نکته بالا تعداد اعداد ۵ رقمی با شرط خواسته شده برابر $۵!-۲!\times۴!=۷۲$ است.

مدرسان شریف

فصل سوم

«متغیرهای تصادفی»

مقدمه

در فصل قبل بحث احتمال تنها محدود به یک یا چند جنبه خاص یک آزمایش تصادفی بود نه تمام خصوصیات آن‌ها. به عنوان مثال در پرتاب یک تاس، فقط برآمدی از قبیل «رو شدن عدد ۳» مورد بررسی قرار گرفت حال آنکه ممکن است بخواهیم برآمد «رو شدن هریک از اعداد» را مورد بررسی قرار دهیم.

در چنین مسائلی اگر عناصر و اعضای فضای نمونه‌ای مربوط به آزمایش تصادفی، عدد نباشد، توصیف و بررسی هر خصوصیت از آن‌ها بسیار دشوار خواهد بود. لذا نیاز است روشی ارائه گردد که به جای عناصر فضای نمونه اعداد قرار بگیرد. به عنوان مثال در پرتاب تاس، برآمد «رو شدن هریک از اعداد» متغیر تصادفی است که می‌تواند مقادیر ۱، ۲، ... ، ۶ را اختیار نماید. اکنون با توجه به توضیحات بالا متغیر تصادفی را تعریف می‌کنیم.

متغیر تصادفی

فرض کنید فضای نمونه‌ای یک آزمایش تصادفی S باشد. در این صورت متغیر تصادفی تابعی است از فضای نمونه‌ای S به مجموعه اعداد حقیقی. به عبارت دیگر تابعی است که به هرکدام از اعضای فضای نمونه‌ای S، یک عدد نسبت می‌دهد و صفت تصادفی به این دلیل به کار می‌رود که از قبل نمی‌دانیم کدام پیشامد رخ خواهد داد و متغیر مورد نظر چه مقداری را اختیار خواهد کرد. معمولاً متغیرهای تصادفی را با حروف بزرگ لاتین مانند X و Y و مقادیری را که اختیار می‌کنند با حروف کوچک مانند x و y نشان می‌دهند. به عنوان مثال، فرض کنید سکه‌ای دو مرتبه پرتاب شود و متغیر تصادفی X نشان‌دهنده تعداد شیرهای ظاهر شده باشد. در این صورت مقادیری که X می‌تواند اختیار نماید، به شکل زیر تعیین می‌شود.

در اینجا فضای نمونه‌ای برابر $S=\{TT,TH,HT,HH\}$ است. همان‌طور که ملاحظه می‌شود S شامل چهار برآمد است. برآمد TT به معنای آن است که در پرتاب دو مرتبه سکه، اصلاً شیر ظاهر نشود و چون X نشان‌دهنده تعداد شیرهای ظاهر شده است، لذا برای این برآمد X مقدار صفر را اختیار می‌کند. همچنین برای برآمد TH و HT، چون تعداد شیرهای ظاهر شده برابر یک است، لذا برای هریک از این دو برآمد، X مقدار ۱را اختیار می‌کند و در برآمد HH، چون دو مرتبه شیر ظاهر شده، بنابراین X مقدار ۲ را اختیار می‌کند. برآمدهای آزمایش تصادفی و مقادیری که X اختیار می‌کند در جدول زیر آمده است.

برآمد	TT	TH	HT	HH
مقادیر متغیر تصادفی X	۰	۱	۱	۲

بنابراین نتیجه می‌شود $x=0,1,2$ می‌باشد.

انواع متغیرهای تصادفی

متغیرهای تصادفی به سه دسته گسسته، پیوسته و آمیخته تقسیم می‌شوند که هریک را در زیر توضیح می‌دهیم:

۱ـ متغیر تصادفی گسسته: متغیر تصادفی را گسسته گویند هرگاه مقادیری که اختیار می‌کند، شمارای متناهی یا شمارای نامتناهی باشد. به عنوان مثال، فرض کنید تاسی دو مرتبه پرتاب شود و متغیر تصادفی X نشان‌دهنده مجموع دو عدد رو شده باشد، در این صورت X با مقادیر گسسته $x=2,3,...,12$ می‌باشد. همان‌طور که ملاحظه می‌کنید $\{2,3,...,12\}$ مجموعه شمارای متناهی است.

۲ـ متغیر تصادفی پیوسته: متغیری است که هر مقدار حقیقی (اعم از اعشاری یا کسری) را می‌توان به آن اختصاص داد. به بیان دیگر متغیری است که مقادیر آن در یک فاصله عددی به شکل (a,b) می‌باشد. به عنوان مثال اگر X نشان‌دهنده وزن، قد، سرعت، میزان بارندگی در هریک از شهرها، طول عمر یک لامپ الکتریکی، میزان مصرف آب، برق، گاز باشد، در این صورت متغیر تصادفی X پیوسته است.

۳ـ متغیر تصادفی آمیخته: هر متغیر تصادفی که در فاصله‌هایی از دامنه خود مقادیر عددی پیوسته و گسسته اختیار نماید و به‌صورت ترکیبی از متغیرهای تصادفی گسسته و پیوسته ظاهر شود، به آن متغیر تصادفی آمیخته می‌گویند. این متغیرها معمولاً چند ضابطه هستند. به عنوان مثال اگر X نشان‌دهنده هزینه استفاده از تلفن همراه باشد، در این صورت X شامل هزینه پیامک‌های تلفن همراه برحسب تعداد پیامک و هزینه مکالمات برحسب زمان می‌باشد. لذا می‌توان گفت X یک متغیر تصادفی آمیخته است که شامل مقادیر گسسته هزینه پیامک‌ها و مقادیر پیوسته هزینه مدت زمان مکالمات است.

مثال ۱: سکه‌ای سه مرتبه پرتاب می‌شود. اگر متغیر تصادفی X نشان‌دهنده تعداد شیرهای ظاهر شده باشد، در این صورت مقادیری را که X اختیار می‌کند، بیابید.

☑ **پاسخ:** سکه‌ای که سه مرتبه پرتاب می‌شود، فضای نمونه‌ای آن به‌صورت زیر است:

$$S=\{TTT,TTH,THT,HTT,THH,HTH,HHT,HHH\}$$

چون X نشان‌دهنده تعداد شیرهای ظاهر شده است، بنابراین برای برآمد TTT، مقدار X برابر صفر و برای هریک از برآمدهای TTH، THT و HTT، مقدار X برابر ۱ و برای هریک از برآمدهای THH، HTH و HHT، مقدار X برابر ۲ و برای برآمد HHH مقدار X برابر ۳ خواهد بود. برآمدها و مقادیر متناظر هر برآمد در جدول زیر آمده است.

برآمد	TTT	TTH	THT	HTT	THH	HTH	HHT	HHH
مقدار متغیر تصادفی X	۰	۱	۱	۱	۲	۲	۲	۳

لذا نتیجه می‌شود $x=0,1,2,3$ می‌باشد.

درسنامه (۱): متغیرهای تصادفی گسسته و توابع احتمال

تابع احتمال متغیرهای تصادفی گسسته

تابع احتمال برای متغیرهای تصادفی گسسته جدول یا فرمولی است که برای هریک از مقادیر متغیر تصادفی، احتمال مربوط به آن را مشخص می‌کند. برای متغیر تصادفی گسسته X با مقادیر $x=x_1,x_2,...,x_n$، احتمال آنکه X مقدار x_i را اختیار نماید با نماد $P(X=x_i)$ یا $f(x_i)$ نشان می‌دهیم.

به عنوان مثال، اگر از بین چهار کارت به شماره ۱ تا ۴ به تصادف کارتی انتخاب کنیم و X نشان‌دهنده عدد روی کارت باشد، در این صورت x مقادیر ۱، ۲، ۳ یا ۴ را اختیار می‌کند. احتمال آنکه $x=1$ باشد برابر $P(X=1)=\frac{1}{4}$، احتمال آنکه $x=2$ باشد برابر $P(X=2)=\frac{1}{4}$، احتمال آنکه $x=3$ باشد، برابر $P(X=3)=\frac{1}{4}$ و احتمال آنکه $x=4$ باشد برابر $P(X=4)=\frac{1}{4}$ است. لذا تابع احتمال X به فرم جدولی به‌صورت زیر است:

x	۱	۲	۳	۴
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

❖ **تعریف:** برای متغیر تصادفی X با مقادیر گسسته $x_1,x_2,...,x_n$، عبارت $f(x)$ را تابع احتمال متغیر تصادفی گسسته X گویند، اگر و فقط اگر در شرایط زیر صدق نماید:

$$۱)\ 0\le f(x_i)\le 1 \quad ;\quad i=1,2,...,n \qquad ۲)\ \sum_{i=1}^{n} f(x_i)=\sum_{i=1}^{n} P(X=x_i)=1$$

مثال ۲: به ازای کدام مقدار k، عبارت زیر تابع احتمال برای متغیر تصادفی X است؟

$$f(x)=k\binom{3}{x} \quad ;\quad x=0,1,2,3$$

۱) $\frac{1}{16}$ ۲) $\frac{1}{8}$ ۳) $\frac{1}{4}$ ۴) $\frac{1}{10}$

☑ **پاسخ:** گزینه «۲» در اینجا X گسسته با مقادیر $x=0,1,2,3$ است. طبق تعریف، شرط اینکه $f(x)$ تابع احتمال باشد، این است که:

$$f(0)+f(1)+f(2)+f(3)=1$$

با توجه به تابع داده شده داریم:

$$f(0)=k\binom{3}{0}=k \ ,\ f(1)=k\binom{3}{1}=3k$$

$$f(2)=k\binom{3}{2}=3k \ ,\ f(3)=k\binom{3}{3}=k$$

در نتیجه مقدار خواسته شده برابر است با:

$$k+3k+3k+k=1 \Rightarrow 8k=1 \Rightarrow k=\frac{1}{8}$$

نکته۱: برای عبارت e^x داریم:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

تساوی بالا بسط مک‌لورن تابع e^x است که در برخی مسائل می‌توان از آن استفاده کرد. توجه نمایید که x متغیر است و به جای آن می‌توان هر عدد حقیقی قرار داد.

به عنوان مثال اگر $x = 1$ باشد، تساوی بالا به صورت مقابل خواهد بود.

$$e^1 = \sum_{n=0}^{\infty} \frac{(1)^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} = (\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \cdots)$$

نکته۲: فرض کنید a عدد ثابت و معلوم باشد، در این صورت عبارت $f(x) = k\frac{a^x}{x!}$; $x = 0,1,2,\ldots$ وقتی تابع احتمال است که $k = e^{-a}$ باشد.

مثال۳: اگر تابع توزیع احتمال متغیر تصادفی گسسته X به‌صورت $f(x) = C\frac{2^x}{x!}; x = 0,1,2,3,4,\cdots$ باشد، عدد ثابت C کدام است؟

(مدیریت و حسابداری ـ سراسری ۹۳)

۱) e^2 ۲) $2e$ ۳) e^{-2} ۴) $Ln2$

☑ **پاسخ:** گزینه «۳» **روش اول:** برای تعیین عدد ثابت C از ویژگی مقابل برای تابع احتمال استفاده می‌کنیم: $f(0) + f(1) + f(2) + f(3) + \cdots = 1$

چون $f(x) = C\frac{2^x}{x!}$، لذا داریم:

$$C\frac{2^0}{0!} + C\frac{2^1}{1!} + C\frac{2^2}{2!} + C\frac{2^3}{3!} + \cdots = 1$$

از عبارت C فاکتور می‌گیریم. لذا تساوی مقابل به دست می‌آید.

$$C\sum_{n=0}^{\infty} \frac{2^n}{n!} = 1$$

اکنون در بسط $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ قرار می‌دهیم $x = 2$. داریم:

$$e^2 = \sum_{n=0}^{\infty} \frac{2^n}{n!}$$

بنابراین در تساوی به‌دست آمده $C\sum_{n=0}^{\infty} \frac{2^n}{n!} = 1$، به جای $\sum_{n=0}^{\infty} \frac{2^n}{n!}$ مساوی آن e^2 قرار می‌دهیم. در نتیجه:

$$C(e^2) = 1 \Rightarrow C = \frac{1}{e^2} = e^{-2}$$

روش دوم: استفاده از نکته بالا:

$$a = 2 \Rightarrow c = e^{-a} = e^{-2}$$

مثال۴: فرض کنید X یک متغیر تصادفی با تابع احتمال زیر باشد. مقدار c کدام است؟

(علوم اقتصادی ـ دکتری ۱۴۰۲)

$$P(X = x) = \begin{cases} c(\frac{2}{3})^x & ; x = 1,2,3,\ldots \\ 0 & ; \text{سایر جاها} \end{cases}$$

۱) ۲ ۲) ۳ ۳) $\frac{3}{2}$ ۴) $\frac{1}{2}$

☑ **پاسخ:** گزینه «۴» می‌دانیم که متغیر x یک متغیر تصادفی گسسته می‌باشد و از طرفی طبق خواص تابع احتمال داریم:

$$\sum_{x=1}^{\infty} P(X = x) = 1 \Rightarrow \sum_{x=1}^{\infty} C(\frac{2}{3})^x = 1 \Rightarrow C\sum_{x=1}^{\infty} (\frac{2}{3})^x = 1$$

از طرفی سری فوق یک سری هندسی می‌باشد. در سری هندسی داریم:

$$\sum_{x=1}^{\infty} q^x = \frac{q}{1-q}$$

بنابراین به ازای $q = \frac{2}{3}$ داریم:

$$C\frac{\frac{2}{3}}{1-\frac{2}{3}} = 1 \Rightarrow C\frac{\frac{2}{3}}{\frac{1}{3}} = 1 \Rightarrow C = \frac{1}{2}$$

مثال۵: متغیر تصادفی X با تابع احتمال زیر داده شده است. حاصل $P(X \geq 3)$ کدام است؟

x	۱	۲	۳	۴
$f(x)$	۰/۱	۰/۲	k	۰/۲

۱) ۰/۴ ۲) ۰/۳ ۳) ۰/۵ ۴) ۰/۷

☑ **پاسخ:** گزینه «۴» برای پاسخ به این سؤال ابتدا مقدار k را حساب می‌کنیم. برای این کار از ویژگی $\sum_x f(x) = 1$ استفاده می‌کنیم.

$$0/1 + 0/2 + k + 0/2 = 1 \Rightarrow k = 1 - 0/5 = 0/5$$

با محاسبه مقدار k، حاصل احتمال خواسته شده برابر است با:

$$P(X \geq 3) = P(X = 3) + P(X = 4) = 0/5 + 0/2 = 0/7$$

مثال ۶: اگر $x \in \mathbb{N}$; $P(X=x)=\frac{۱}{x^۲+x}$ تابع احتمال متغیر تصادفی X باشد، حاصل $P(۲ \le X \le ۱۹)$ کدام است؟

۱) ۰/۳۶ ۲) ۰/۴۵ ۳) ۰/۵۴ ۴) ۰/۶۳

پاسخ: گزینه «۲» طبق فرض مسئله $x \in \mathbb{N}$، این نشان می‌دهد متغیر تصادفی X گسسته است و مقادیر طبیعی $x=۱,۲,۳,...$ را اختیار می‌کند.

بنابراین: $P(۲ \le X \le ۱۹)=P(X=۲)+P(X=۳)+\cdots+P(X=۱۹)$

در اینجا برای سادگی محاسبات، عبارت کسری $f(x)=\frac{۱}{x^۲+x}$ را به کسرهای جزئی تجزیه می‌کنیم: $P(X=x)=\frac{۱}{x^۲+x}=\frac{۱}{x(x+۱)}=\frac{۱}{x}-\frac{۱}{x+۱}$

اکنون حاصل احتمال خواسته شده برابر است با: $P(۲ \le X \le ۱۹)=(\frac{۱}{۲}-\frac{۱}{۳})+(\frac{۱}{۳}-\frac{۱}{۴})+(\frac{۱}{۴}-\frac{۱}{۵})+\cdots+(\frac{۱}{۱۹}-\frac{۱}{۲۰})$

همان‌طور که ملاحظه می‌کنید جملات به‌صورت تلسکوپی ساده شده و تنها جمله اول و جمله آخر باقی می‌ماند. $P(۲ \le X \le ۱۹)=\frac{۱}{۲}-\frac{۱}{۲۰}=\frac{۹}{۲۰}=۰/۴۵$

مثال ۷: **پرونده‌های خسارتی رسیده به شرکت بیمه متغیر تصادفی با تابع احتمال $P[N=n]=\frac{۱}{۲^{n+۱}}$ برای $n \ge ۰$ در طول یک هفته است. پرونده‌های رسیده در طول یک هفته مستقل از پرونده‌ها در هفته‌های دیگر است. احتمال این‌که در یک دوره دو هفته‌ای فقط ۷ پرونده به شرکت بیمه برسد، کدام است؟**

(علوم اقتصادی ـ سراسری ۱۴۰۰)

۱) $\frac{۱}{۲۴}$ ۲) $\frac{۱}{۳۲}$ ۳) $\frac{۱}{۴۸}$ ۴) $\frac{۱}{۶۴}$

پاسخ: گزینه «۴» طبق فرض مسئله X نشان‌دهنده تعداد پرونده‌ها در یک هفته است. اکنون اگر Y نشان‌دهنده تعداد پرونده‌ها در دو هفته باشد، داریم:

$$Y=X-۲ \Rightarrow X=Y+۲$$

$$P(X=n)=\frac{۱}{۲^{n+۱}} \Rightarrow P(Y+۲=n)=\frac{۱}{۲^{n+۱}} \Rightarrow P(Y=n-۲)=\frac{۱}{۲^{n+۱}} \Rightarrow P(Y=n)=\frac{۱}{۲^{n-۱}}$$

در نتیجه احتمال آنکه در دو هفته، ۷ پرونده به شرکت بیمه برسد، برابر است با: $P=(Y=۷)=\frac{۱}{۲^{۷-۱}}=\frac{۱}{۲^۶}=\frac{۱}{۶۴}$

تابع توزیع تجمعی (تابع احتمال تجمعی)

در برخی از مسائل، تعیین احتمال آنکه متغیر تصادفی X، کوچک‌تر یا بزرگ‌تر از x باشد، دارای اهمیت است. به عنوان مثال، یک تولیدکننده قصد دارد تعیین نماید، احتمال اینکه در هفته جاری حداقل ۱۰۰ واحد سفارش فروش داشته باشد، چقدر است؟ یا احتمال آنکه در یک روز خاص حداکثر ۱۰ نفر به یک مرکز درمانی بهداشتی مراجعه نمایند، چقدر است؟ برای پاسخ به این‌گونه سؤالات، می‌توان از تابع توزیع تجمعی متغیر تصادفی که از روی تابع احتمال به دست می‌آید، استفاده کرد. برای این منظور احتمال آنکه متغیر تصادفی X مقداری کوچک‌تر و یا مساوی x اختیار نماید، به‌صورت $F(x)=P(X \le x)$ نوشته می‌شود و این تابع که برای تمام اعداد حقیقی x تعریف شده است، **تابع توزیع تجمعی** نامیده می‌شود.

❖ **تعریف:** متغیر تصادفی گسسته X با تابع احتمال $f(x)$ را در نظر بگیرید.

تابع توزیع تجمعی X را با نماد $F(x)$ نشان می‌دهیم و به‌صورت زیر تعریف می‌شود:

$$F(x)=P(X \le x)=\sum_{t \le x} f(t)$$

و این بدان معناست که تابع توزیع در هر نقطه که خواسته شود، باید احتمال‌های آن نقطه به پایین را با هم جمع کنیم.

مثال ۸: **تابع توزیع تجمعی X، با تابع احتمال داده شده زیر را بیابید.**

x	-۱	۰	۱	۲
$f(x)$	$\frac{۲}{۷}$	$\frac{۱}{۷}$	$\frac{۳}{۷}$	$\frac{۱}{۷}$

☑ **پاسخ:** برای متغیر تصادفی گسسته X ، تابع توزیع تجمعی با رابطه $F(x)=P(X\le x)=\sum_{t\le x} f(t)$ به دست می‌آید. در اینجا $x=-1,0,1,2$ می‌باشد. بنابراین مقدار تابع توزیع تجمعی در نقاط داده شده، برابر است با:

$$F(-1)=P(X\le -1)=P(X=-1)=\frac{2}{7}$$

$$F(0)=P(X\le 0)=P(X=0)+P(X=-1)=\frac{1}{7}+\frac{2}{7}=\frac{3}{7}$$

$$F(1)=P(X\le 1)=P(X=1)+P(X=0)+P(X=-1)=\frac{3}{7}+\frac{1}{7}+\frac{2}{7}=\frac{6}{7}$$

$$F(2)=P(X\le 2)=P(X=2)+P(X=1)+P(X=0)+P(X=-1)=\frac{1}{7}+\frac{3}{7}+\frac{1}{7}+\frac{2}{7}=1$$

x	-1	0	1	2
$F(x)$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{6}{7}$	1

مثال ۹: برای متغیر تصادفی X با تابع احتمال زیر، مقدار $F(\sqrt{3})$ کدام است؟

x	0	1	2	3
$f(x)$	$0/2$	$0/25$	$0/25$	$0/3$

۱) $0/45$ ۲) $0/25$

۳) $0/7$ ۴) $0/55$

☑ **پاسخ:** گزینه «۱» تابع توزیع تجمعی در نقطه X به‌صورت $F(x)=P(X\le x)$ تعریف می‌شود. بنابراین $F(\sqrt{3})$، برابر است با:

$$F(\sqrt{3})=P(X\le \sqrt{3})$$

در اینجا $x=0,1,2,3$ است و چون تنها اعداد کمتر از $\sqrt{3}\simeq 1/7$ برابر $x=0,1$ است، لذا حاصل $F(\sqrt{3})$ برابر می‌شود با:

$$F(\sqrt{3})=P(X\le 1/7)=f(0)+f(1)=0/2+0/25=0/45$$

نکته ۳: تابع توزیع تجمعی را می‌توان به شکل تابع چند ضابطه نوشت. به مثال زیر توجه کنید:

مثال ۱۰: متغیر تصادفی X با تابع احتمال زیر را در نظر بگیرید.

x	0	1	2	3	4
$f(x)$	$0/15$	$0/05$	$0/3$	$0/4$	$0/1$

الف) نمودار بافت‌نمای احتمال $f(x)$ را رسم کنید.

ب) تابع توزیع تجمعی X را به صورت تابع چند ضابطه بنویسید.

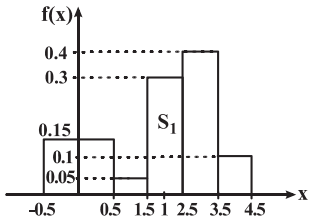
ج) نمودار تابع توزیع تجمعی X را به شکل پله‌ای رسم کنید.

☑ **پاسخ: الف)** برای رسم نمودار بافت‌نمای احتمال، با توجه به اینکه فاصله نقاط یک واحد است، ابتدا به اندازه $0/5$ از هر نقطه کم و $0/5$ به آن نقطه اضافه می‌کنیم تا برای هر نقطه یک فاصله به‌صورت زیر به دست آید:

فاصله عددی	$(-0/5)-0/5$	$0/5-1/5$	$1/5-2/5$	$2/5-3/5$	$3/5-4/5$
احتمال	$0/15$	$0/05$	$0/3$	$0/4$	$0/1$

اکنون در دستگاه مختصات قائم، مستطیل‌هایی به عرض فاصله عددی به موازات محور افقی و به طول احتمال هر فاصله به موازات محور قائم رسم می‌کنیم. توجه نمایید در اینجا مساحت هر مستطیل برابر است با احتمال آنکه متغیر تصادفی X برابر نقطه میانی قاعده پایین مستطیل باشد.

$$S_1=P(X=1)=1\times 0/3=0/3$$



ب) برای به‌دست آوردن تابع توزیع تجمعی X از رابطه $F(x)=P(X\le x)$ استفاده می‌کنیم و $F(a)$ به معنای جمع احتمالات از اولین نقطه تا نقطه $x=a$ می‌باشد. بنابراین با توجه به این توضیحات تابع توزیع تجمعی X به‌صورت زیر می‌باشد.

$$F(0)=P(X\le 0)=P(X=0)=0/15$$

$$F(1)=P(X\le 1)=P(X=1)+P(X=0)=0/05+0/15=0/2$$

$$F(2)=P(X\le 2)=P(X=2)+P(X=1)+P(X=0)=0/3+0/05+0/15=0/5$$

$$F(3)=P(X\le 3)=P(X=3)+P(X=2)+P(X=1)+P(X=0)=0/4+0/3+0/05+0/15=0/9$$

$$F(4)=P(X\le 4)=P(X=4)+P(X=3)+P(X=2)+P(X=1)+P(X=0)=0/1+0/4+0/3+0/05+0/15=1$$

x	0	1	2	3	4
$F(x)$	$0/15$	$0/2$	$0/5$	$0/9$	1

مدرسان شریف

فصل چهارم

«توزیع‌های احتمال خاص»

مقدمه

در این فصل توزیع‌های احتمال خاص که دارای نام استاندارد هستند، به همراه پارامترهای آن‌ها از قبیل میانگین، واریانس و تابع مولد گشتاور ارائه می‌شود. این توزیع‌ها شامل دو دسته گسسته و پیوسته می‌شود که در نظریه آمار دارای کاربردهای فراوان هستند. در این فصل دو مبحث در دو درسنامه جداگانه ارائه شده است. درسنامه اول توابع احتمال با متغیرهای تصادفی گسسته و درسنامه دوم توابع چگالی احتمال با متغیرهای تصادفی پیوسته است.

درسنامه (۱): توزیع‌های آماری گسسته

توزیع‌های احتمال گسسته: این توزیع‌ها مربوط به متغیرهای تصادفی با مقادیر گسسته بوده و شامل موارد زیر است:

توزیع احتمال یکنواخت گسسته

اگر متغیر تصادفی X مقادیر گسسته $x_1, x_2, \ldots, x_n$ را با احتمال‌های برابر $\frac{1}{n}$ اختیار نماید، در این صورت X دارای توزیع یکنواخت گسسته است.

به عنوان مثال فرض کنید تاسی یک مرتبه پرتاب شود و X نشان‌دهنده عدد رو شده باشد. در این صورت متغیر تصادفی X هریک از مقادیر $1, 2, \ldots, 6$ را با احتمال برابر $\frac{1}{6}$ اختیار می‌کند. لذا تابع احتمال X به‌صورت مقابل می‌باشد. $f(x) = \frac{1}{6} \; ; \; x = 1, 2, 3, 4, 5, 6$

❖ **تعریف:** متغیر تصادفی X دارای توزیع یکنواخت گسسته است، اگر و تنها اگر توزیع احتمال آن به‌صورت زیر باشد:

$$f(x) = \begin{cases} \frac{1}{n} & ; \; x = x_1, x_2, \cdots, x_n \\ 0 & ; \; \text{سایرجاها} \end{cases}$$

و در حالتی‌که X مقادیر $1, 2, \ldots, n$ را اختیار نماید، تابع احتمال X به‌صورت زیر است:

$$f(x) = \begin{cases} \frac{1}{n} & ; \; x = 1, 2, \cdots, n \\ 0 & ; \; \text{سایرجاها} \end{cases}$$

که به آن تابع احتمال یکنواخت گسسته استاندارد می‌گویند.

ویژگی‌های توزیع یکنواخت گسسته استاندارد: فرض کنید X دارای توزیع یکنواخت گسسته با مقادیر $1, 2, \ldots, n$ باشد، در این صورت داریم:

$$۱) \; E(X) = \frac{n+1}{2} \qquad ۲) \; Var(X) = \frac{n^2 - 1}{12} \qquad ۳) \; M_X(t) = \frac{e^t(1 - e^{nt})}{n(1 - e^t)}$$

فرم دوم توزیع یکنواخت گسسته: اگر متغیر تصادفی X مقادیر $0, 1, 2, \ldots, n$ را با احتمال‌های برابر اختیار نماید، در این صورت تابع احتمال X به‌صورت مقابل است:

$$f(x) = \frac{1}{n+1} \quad ; \quad x = 0, 1, 2, \ldots, n$$

در این حالت میانگین و واریانس X برابر است با:

$$۱) E(X) = \frac{n}{2} \qquad ۲) Var(X) = \frac{n(n+2)}{12}$$

مثال ۱: اگر متغیر تصادفی X به‌صورت $f(x)=\begin{cases}\frac{1}{9} & ; \quad x=1,2,\ldots,9 \\ 0 & ; \quad \text{سایر جاها}\end{cases}$ توزیع شده باشد، میانگین و واریانس آن برابر است با: (علوم اقتصادی ـ سراسری ۸۴)

۱) $E(X)=\frac{9}{2}\ ,\ Var(X)=\frac{81}{12}$

۲) $E(X)=5\ ,\ Var(X)=\frac{20}{3}$

۳) $E(X)=5\ ,\ Var(X)=\frac{12}{3}$

۴) $E(X)=\frac{71}{10}\ ,\ Var(X)=\frac{9}{12}$

☑ **پاسخ:** گزینه «۲» در اینجا تابع احتمال یکنواخت گسسته با $n=9$ داریم. لذا میانگین و واریانس X برابر است با:

$$E(X)=\frac{n+1}{2}=\frac{9+1}{2}=5$$

$$Var(X)=\frac{n^2-1}{12}=\frac{81-1}{12}=\frac{80}{12}=\frac{20}{3}$$

مثال ۲: امید ریاضی و واریانس متغیر تصادفی X که در پرتاب تاس نشان‌دهنده شماره‌های روی تاس و دارای توزیع یکنواخت می‌باشند، به ترتیب برابرند با: (علوم اقتصادی ـ سراسری ۹۱)

۱) $\mu=3\ ,\ \sigma^2=2/5$

۲) $\mu=3/5\ ,\ \sigma^2=\frac{35}{12}$

۳) $\mu=2/5\ ,\ \sigma^2=3$

۴) $\mu=3/5\ ,\ \sigma^2=3$

☑ **پاسخ:** گزینه «۲» وقتی تاسی یک مرتبه پرتاب و X نشان‌دهنده عدد رو شده باشد، در این صورت متغیر تصادفی X هریک از اعداد ۱،۲،۳،۴،۵ و ۶ را با احتمال $\frac{1}{6}$ اختیار می‌کند. بنابراین تابع احتمال X به‌صورت $f(x)=\frac{1}{6}\ ;\ x=1,2,3,4,5,6$ یا

x	۱	۲	۳	۴	۵	۶
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

است.

این نشان می‌دهد X دارای توزیع یکنواخت گسسته با $n=6$ است. در نتیجه میانگین و واریانس X طبق دستور زیر حساب می‌شود:

$$E(X)=\frac{n+1}{2}=\frac{6+1}{2}=3/5$$

$$Var(X)=\frac{n^2-1}{12}=\frac{36-1}{12}=\frac{35}{12}$$

مثال ۳: یک عدد تصادفی از مجموعه $\{1,2,\ldots,n\}$ انتخاب می‌کنیم. اگر عدد انتخابی k باشد، آنگاه عدد دوم را به تصادف از $\{1,\ldots,k\}$ انتخاب می‌کنیم. اگر X عدد دوم باشد، مقدار $E(X)$ کدام است؟ (علوم اقتصادی ـ سراسری ۱۴۰۲))

۱) $\frac{n+1}{4}$

۲) $\frac{n+3}{4}$

۳) $\frac{n+1}{2}$

۴) $\frac{n+3}{2}$

☑ **پاسخ:** گزینه «۲» طبق فرض مسئله متغیر تصادفی k دارای توزیع یکنواخت گسسته با مقادیر $k=1,2,\ldots,n$ می‌باشد. بنابراین: $E(k)=\frac{n+1}{2}$

همچنین متغیر تصادفی X به شرط انتخاب k یعنی $X|k$ دارای توزیع یکنواخت گسسته با مقادیر $X|k=1,2,\ldots,k$ می‌باشد. لذا داریم: $E(X|k)=\frac{k+1}{2}$

اکنون حاصل $E(X)$ با توجه به رابطه $E(X)=E(E(X|k))$ برابر است با:

$$E(X)=E(\frac{k+1}{2})=\frac{1}{2}E(k)+\frac{1}{2}=\frac{1}{2}(\frac{n+1}{2})+\frac{1}{2}=\frac{n+1}{4}+\frac{1}{2}=\frac{n+1+2}{4}=\frac{n+3}{4}$$

توزیع برنولی

هر آزمایش تصادفی که نتیجه آن تنها شامل دو برآمد پیروزی یا شکست باشد، به آن آزمایش برنولی می‌گویند. به عنوان مثال اگر سکه‌ای یک مرتبه پرتاب شود، در این صورت نتیجه آن شامل تنها دو برآمد شیر یا خط است. اگر برآمد شیر پیروزی تلقی شود، در این صورت برآمد خط شکست خواهد بود.

اکنون اگر متغیر تصادفی X نشان‌دهنده تعداد پیروزی‌ها باشد، آنگاه X تنها مقادیر ۰ یا ۱ را اختیار می‌کند.

❖ **تعریف:** متغیر تصادفی X دارای توزیع برنولی است اگر و تنها اگر تابع احتمال آن به‌صورت مقابل باشد: $f(x)=p^x q^{1-x}\ ;\ x=0,1$

که در آن p احتمال پیروزی و q احتمال شکست است. بنابراین:

$P(X=1)=p$; احتمال پیروزی

$P(X=0)=q$; احتمال شکست

همچنین داریم: $p+q=1\ ;\ 0\le p\le 1\ ,\ 0\le q\le 1$

ویژگی‌های توزیع برنولی: فرض کنید X دارای توزیع برنولی با پارامتر p باشد $(X \sim B(p,1))$، در این صورت داریم:

۱) $E(X)=p$	۲) $Var(X)=pq$
۳) $E(X^k)=p$	۴) $M_X(t)=q+pe^t$

مثال ۴: در آزمایش تصادفی که دارای توزیع برنولی است و احتمال پیروزی $\frac{۱}{۳}$ باشد، آنگاه کدام گزینه صحیح نیست؟ (علوم اقتصادی ـ دکتری ۹۵)

۱) واریانس $\frac{۲}{۹}$ ۲) میانه $\frac{۱}{۲}$ ۳) میانگین $\frac{۱}{۳}$ ۴) احتمال شکست $\frac{۲}{۳}$

☑ **پاسخ:** گزینه «۲» در توزیع برنولی وقتی احتمال پیروزی $p=\frac{۱}{۳}$ باشد، در این صورت احتمال شکست (q) با توجه به رابطه $p+q=۱$ برابر است با:

$$q=۱-p=۱-\frac{۱}{۳}=\frac{۲}{۳}$$

در نتیجه داریم:

$$E(X)=p=\frac{۱}{۳} \quad , \quad Var(X)=pq=\frac{۱}{۳}\times\frac{۲}{۳}=\frac{۲}{۹}$$

این نشان می‌دهد گزینه‌های ۱، ۳ و ۴ صحیح است و لذا گزینه (۲) صحیح نیست.

توزیع دوجمله‌ای

اگر آزمایش برنولی n مرتبه و به‌طور مستقل از هم تکرار شود و X نشان‌دهنده تعداد پیروزی‌ها در n آزمایش برنولی باشد، در این حالت X دارای توزیع دوجمله‌ای با پارامتر n و p است. به عنوان مثال اگر سکه‌ای ۵ مرتبه پرتاب شود و X نشان‌دهنده تعداد شیرهای ظاهر شده باشد، در این صورت X دارای توزیع دوجمله‌ای است.

❖ **تعریف:** متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامتر n و p است اگر و تنها اگر تابع احتمال آن به‌صورت زیر باشد:

$$P(X=x)=f(x)=\binom{n}{x}p^x q^{n-x} \quad ; \quad x=۰,۱,۲,\ldots,n$$

که در آن $P(X=x)$ برابر احتمال کسب x پیروزی در n آزمایش برنولی می‌باشد.

ویژگی‌های توزیع دوجمله‌ای: فرض کنید X دارای توزیع دوجمله‌ای با پارامتر n و p باشد $(X \sim B(n,p))$، در این صورت داریم:

۱) $E(X)=np$	۲) $Var(X)=npq$	۳) $M_X(t)=(q+pe^t)^n$

مثال ۵: ۹۰ درصد کارکنان یک شرکت باسوادند. اگر ۳ نفر از کارکنان این شرکت به تصادف انتخاب شوند، احتمال آنکه حداقل یک نفر باسواد باشد، کدام است؟ (مدیریت ـ دکتری ۹۵)

۱) ۰/۹۹۹ ۲) ۰/۹۸۳ ۳) ۰/۹۸۰ ۴) ۰/۹۷۲

☑ **پاسخ:** گزینه «۱» در اینجا احتمال کسب حداقل یک پیروزی (حداقل یک باسواد) از تعداد $n=۳$ نفر مدنظر است. لذا با توزیع دوجمله‌ای با احتمال پیروزی $p=۰/۹$ و احتمال شکست $q=۱-۰/۹=۰/۱$ مواجه هستیم. بنابراین اگر X نشان‌دهنده تعداد افراد باسواد در بین $n=۳$ نفر باشد، در این صورت احتمال آنکه حداقل یک نفر باسواد باشد، با توجه به دستور احتمال متمم و دستور $P(X=x)=\binom{n}{x}p^x q^{n-x}$ برابر است با:

$$P(X\geq ۱)=۱-P(X<۱)=۱-P(X=۰)=۱-\binom{۳}{۰}(۰/۹)^۰(۰/۱)^۳=۱-۰/۰۰۱=۰/۹۹۹$$

مثال ۶: قیمت سهام شرکت A در هر روز با احتمال $\frac{۱}{۳}$ بالا رفته و با احتمال $\frac{۲}{۳}$ پایین یا ثابت می‌ماند. احتمال آنکه قیمت سهام مذکور طی ۵ روز، فقط چهار بار بالا رود، کدام است؟ (علوم اقتصادی ـ دکتری ۹۵)

۱) $\frac{۱۰}{۸۱}$ ۲) $\frac{۲}{۲۴۳}$ ۳) $\frac{۵}{۸۱}$ ۴) $\frac{۱۰}{۲۴۳}$

☑ **پاسخ:** گزینه «۴» اگر X نشان‌دهنده تعداد روزهایی باشد که در ۵ روز قیمت سهام شرکت A بالا می‌رود، در این‌صورت احتمال آنکه در طی ۵ روز $x=۴$ بار بالا رود، با استفاده از تابع احتمال دوجمله‌ای با احتمال پیروزی $p=\frac{۱}{۳}$ و احتمال شکست $q=\frac{۲}{۳}$ برابر است با:

$$P(X=۴)=\binom{n}{x}p^x q^{n-x}=\binom{۵}{۴}(\frac{۱}{۳})^۴(\frac{۲}{۳})^۱=۵\times\frac{۲}{۲۴۳}=\frac{۱۰}{۲۴۳}$$

مثال ۷: قانون توزیع متغیر تصادفی X در جامعه‌ای به‌صورت زیر حاصل شده است. ضریب تغییرات متغیر تصادفی X کدام است؟ (مدیریت ـ دکتری ۹۹)

$$f(x)=\binom{8}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{8-x}$$

۱) $\frac{2}{4}$　　۲) $\frac{1}{3}$

۳) $\frac{1}{4}$　　۴) $\frac{2}{3}$

پاسخ: گزینه «۳» تابع احتمال داده‌شده به فرم $f(x)=\binom{n}{x}P^xq^{n-x}$ است. لذا X دارای توزیع دوجمله‌ای با پارامتر $n=8$ ، $P=\frac{2}{3}$ و $q=\frac{1}{3}$ است. در توزیع دوجمله‌ای میانگین و واریانس به شکل مقابل حساب می‌شود. $\mu=np=8\times\frac{2}{3}=\frac{16}{3}$ ، $\sigma^2=npq=\frac{16}{3}\times\frac{1}{3}=\frac{16}{9}\Rightarrow\sigma=\frac{4}{3}$

در نتیجه ضریب تغییرات برابر است با: $CV=\frac{\sigma}{\mu}=\frac{\frac{4}{3}}{\frac{16}{3}}=\frac{1}{4}$

مثال ۸: احتمال اینکه فردی در هر بار تیراندازی هدف را بزند ۸۰% است. اگر وی ۲۵ بار به هدف شلیک کند، میانگین و انحراف معیار تیرهای به هدف خورده به ترتیب کدام است؟ (علوم اقتصادی ـ دکتری ۹۶)

۱) ۲۰ و ۲　　۲) ۱۸ و ۳

۳) ۲۰ و ۴　　۴) ۲۲ و ۴

پاسخ: گزینه «۱» فرض کنید X نشان‌دهنده تعداد تیرهای به هدف خورده در $n=25$ شلیک باشد. در این صورت X دارای توزیع دوجمله‌ای با احتمال پیروزی $P=0/8$ و احتمال شکست $q=0/2$ است. لذا میانگین و انحراف معیار X برابر است با:

$$E(X)=np=25\times0/8=20$$

$$Var(X)=npq=25\times0/8\times0/2=4\Rightarrow\sigma=\sqrt{4}=2$$

مثال ۹: احتمال اینکه هدفی مورد اصابت قرار گیرد $\frac{3}{4}$ است. ۸ بار به سمت این هدف شلیک می‌کنیم. احتمال اینکه حداقل یکبار به هدف بزنیم، برابر کدام است؟ (علوم اقتصادی ـ سراسری ۱۴۰۱)

۱) $\frac{1}{2^{16}}$　　۲) $(\frac{3}{4})^7$

۳) $1-(\frac{3}{4})^7$　　۴) $1-\frac{1}{2^{16}}$

پاسخ: گزینه «۴» برای پاسخ به این سؤال از توزیع دو جمله‌ای با پارامتر $n=8$ و $p=\frac{3}{4}$ و $q=1-\frac{3}{4}=\frac{1}{4}$ استفاده می‌کنیم. برای این منظور داریم:

$$P(X\geq1)=1-P(X=0)=1-\binom{8}{0}(\frac{3}{4})^0(\frac{1}{4})^8=1-(\frac{1}{4})^8=1-(\frac{1}{2})^{16}$$

مثال ۱۰: متغیر تصادفی X دارای توزیع دوجمله‌ای با میانگین ۲/۵ و واریانس ۱/۲۵ می‌باشد. $P(1\leq X\leq4)$ کدام است؟ (مدیریت و حسابداری ـ سراسری ۹۸)

۱) ۰/۹۰۲۵　　۲) ۰/۸۹۷۵

۳) ۰/۹۱۲۵　　۴) ۰/۹۳۷۵

پاسخ: گزینه «۴» برای محاسبه احتمال موردنظر ابتدا مقدار n و p را حساب می‌کنیم. طبق فرض مسئله داریم:

$$\left.\begin{matrix}E(X)=np=2/5\\Var(X)=npq=1/25\end{matrix}\right\}\Rightarrow2/5q=1/25\Rightarrow q=\frac{1/25}{2/5}=0/5\Rightarrow p=1-q=1-0/5=0/5$$

اکنون مقدار n را از رابطه $np=2/5$ حساب می‌کنیم.

$$0/5n=2/5\Rightarrow n=5$$

با محاسبه پارامترهای توزیع دوجمله‌ای حاصل $P(1\leq X\leq4)$ با استفاده از احتمال متمم برابر است با:

$$P(1\leq X\leq4)=1-(P(X=0)+P(X=5))=1-\binom{5}{0}(0/5)^0(0/5)^5-\binom{5}{5}(0/5)^5(0/5)^0=1-0/03125-0/03125=0/9375$$

دقت نمایید که متغیر تصادفی X تنها مقادیر $x=0,1,2,3,4,5$ را اختیار می‌کند.

مدرسان شریف

فصل پنجم

«نمونه‌گیری و توزیع‌های نمونه‌ای»

درسنامه (۱): روش‌های نمونه‌گیری

آمار عمدتاً به نتایج و پیشگویی‌های حاصل از برآمدهای شانسی می‌پردازد که این برآمدها معمولاً در آزمایش‌ها و تحقیقات که به دقت طرح‌ریزی شده‌اند، پیش می‌آیند. در حالت شانسی، این برآمدهای شانسی تشکیل زیرمجموعه یا نمونه‌ای از اندازه‌گیری‌ها یا مشاهداتی از مجموعهٔ بزرگ‌تر به نام جامعه را می‌دهند. در حالت پیوسته، آن‌ها معمولاً مقادیر متغیرهای تصادفی هم توزیع‌اند که این توزیع را توزیع جامعه یا جامعهٔ نامتناهی مورد نمونه‌گیری می‌نامیم. کلمه «نامتناهی» به این معنی است که، از لحاظ منطقی، حدی بر تعداد متغیرهای تصادفی که مقادیر آن‌ها قابل مشاهده است، متصور نیست. به عنوان مثال فرض کنید یک پژوهشگر از ۵۰ خرگوش آزمایشگاهی، تعداد چهار خرگوش را انتخاب و سپس آن‌ها را وزن کند.

در اینجا نمونه انتخابی وزن خرگوش‌های انتخابی از بین وزن ۵۰ خرگوش می‌باشد که وزن آنها تشکیل جامعهٔ نامتناهی را می‌دهد.

در این‌گونه مسائل پژوهشگران به دنبال تعیین پارامترهای جامعه از قبیل میانگین، واریانس و نسبت هستند اما در اغلب موارد، به‌دست آوردن پارامترهای جامعه با سرشماری جامعهٔ آماری دشوار و یا غیرممکن است. بنابراین آن‌ها ناچار هستند به نمونه‌هایی از جامعه آماری برای تخمین پارامترهای موردنظر اکتفا کنند.

پارامتر

هر ویژگی از جامعه آماری را یک پارامتر جامعه می‌گویند. به‌عنوان مثال میانگین جامعه (μ)، واریانس جامعه (σ^2) و نسبت جامعه (p) پارامترهای جامعه هستند.

آماره

ویژگی و شاخصی است که بر اساس یک نمونه تصادفی n تایی انتخابی از جامعه، حاصل می‌شود. به عنوان مثال میانگین نمونه‌ای $(\bar{X})$، واریانس نمونه‌ای (S^2) و نسبت نمونه‌ای $(\bar{p})$ آماره هستند.

تعریف: اگر $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی باشند، آنگاه $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ میانگین نمونه‌ای و $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$ واریانس نمونه‌ای نامیده می‌شود.

از آنجا که اندازه‌های نمونه، از نمونه‌ای به نمونه دیگر تغییر می‌کند، بنابراین مقدار آماره نیز از یک نمونه به نمونه دیگر تغییر می‌کند لذا آماره متغیر تصادفی است.

برای استنباط هر پارامتر نامعلوم جامعه از یک آماره استفاده می‌شود. بنابراین متناظر هر پارامتر جامعه لااقل یک آماره وجود دارد که خود آماره متغیر تصادفی است و دارای توزیع احتمال است. توزیع احتمال یک آماره تابعی است که براساس نمونه‌های تصادفی n تایی که به‌طور مکرر از جامعه آماری انتخاب شده‌اند، به‌دست می‌آیند. این تابع را توزیع نمونه‌گیری آماری می‌گویند.

دلایل نمونه‌گیری

دلایل مختلفی وجود دارد که معمولاً پژوهشگران از سرشماری کل جامعه اجتناب کرده و به تعداد محدودی از عناصر جامعه یعنی نمونه اکتفا می‌کنند. این دلایل به پنج دسته به شرح زیر می‌باشد:

۱ـ کاهش هزینه: اگر داده‌ها از بخش کوچکی از جامعه تأمین شوند، بدیهی است که هزینه تهیه آن‌ها به مراتب کمتر از هزینه سرشماری است.

۲ـ سرعت بیشتر و کاهش زمان: چون حجم نمونه از حجم جامعه کمتر است، بنابراین جمع‌آوری و اندازه‌گیری داده‌ها در نمونه‌گیری با سرعت بیشتر و با صرف وقت کمتر انجام می‌شود.

۳ـ قابلیت اعتماد: چون تعداد داده‌ها در نمونه‌گیری کمتر از سرشماری است، بنابراین اندازه‌گیری آن‌ها با دقت بیشتری نسبت به جامعه انجام می‌شود.

۴ـ به‌روز بودن: نمونه اغلب اطلاعات به‌روز و به‌هنگام نسبت به سرشماری به‌دست می‌دهد. زیرا داده‌های کمتری جمع‌آوری و تجزیه و تحلیل می‌شوند. این جنبه از نمونه‌گیری به‌خصوص زمانی که اطلاعات برای تصمیم‌گیری سریع مورد نیاز است، اهمیت بیشتری پیدا می‌کند.

۵ـ آزمون تخریب‌کننده: وقتی آزمونی موجب خراب شدن و از بین رفتن کالا می‌شود، باید نمونه‌گیری به‌کار رود. زیرا امکان خراب نمودن تمام کالاها وجود ندارد.

روش‌های نمونه‌گیری

شامل موارد زیر می‌شود:

۱ـ نمونه‌گیری تصادفی ساده

در نمونه‌گیری تصادفی ساده هریک از عناصر جامعه موردنظر برای انتخاب شدن شانس برابر دارند. این نوع نمونه‌گیری به دو دسته تقسیم می‌شود.

الف) نمونه‌گیری تصادفی ساده بدون جای‌گذاری:

در این روش اگر واحدی به تصادف از جامعه انتخاب شد، پس از مشاهده صفت مورد بررسی، مجدداً به جمع واحدهای جامعه برگشت داده نمی‌شود.

ب) نمونه‌گیری تصادفی ساده با جای‌گذاری: در این روش اگر عضو یا واحدی از جامعه به تصادف انتخاب شد، پس از مشاهده صفت مورد نظر، مجدداً به جمع واحدهای جامعه برگشت داده می‌شود. بنابراین، این عضو شانس انتخاب برای بار دوم، سوم و ... و n ام را دارد.

۲ـ نمونه‌گیری منظم (سیستماتیک)

در این روش اعضا و عناصر نمونه از فهرست افراد یا اعضاء جامعه آماری که به همین منظور آماده شده است، انتخاب می‌شوند. به عنوان مثال، فرض کنید جامعه‌ای شامل $N = ۳۰۰۰$ عضو است. می‌خواهیم یک نمونه تصادفی به حجم $n = ۱۵۰$ به روش نمونه‌گیری منظم (سیستماتیک) انتخاب نماییم. برای این کار با تقسیم $\frac{N}{n} = \frac{۳۰۰۰}{۱۵۰} = ۲۰$، نقطهٔ شروع نمونه‌گیری از عضوی با شماره کمتر و یا مساوی ۲۰ آغاز و ۲۰ نفر به ۲۰ نفر از روی فهرست اعضاء انتخاب می‌کنیم.

$$x_۱ = ۲۰ ,\ x_i = x_{i-۱} + ۲۰ \quad i = ۲,...,۱۵۰$$

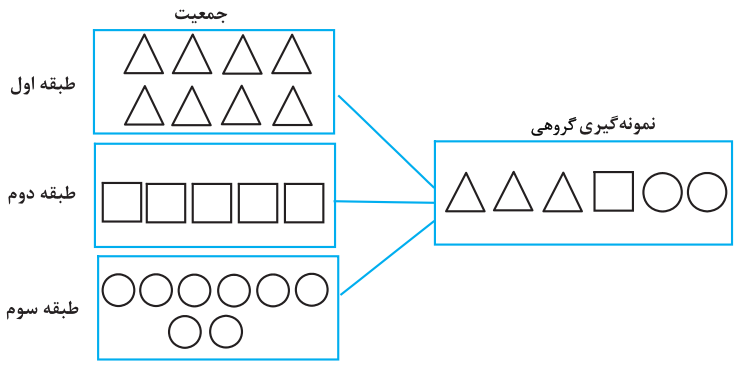
این روش برای جوامعی به کار می‌رود که از قبل اعضاء کدبندی و فهرست شده‌اند. مانند لیست شماره دانشجویان، شماره پرسنلی و یا پلاک منازل و پلاک خودروها.

۳ـ نمونه‌گیری گروهی (طبقه‌بندی)

این روش وقتی به‌کار می‌رود که صفت مورد مطالعه اعضاء جامعه ناهمگن باشد. در این روش عناصر و اعضاء جامعه به چندین گروه و طبقه به گونه‌ای تقسیم و طبقه‌بندی می‌شوند که اعضاء و عناصر هر گروه و طبقه دارای خصوصیات و ویژگی‌های مشابه و یکسان بوده و متجانس باشند. اما ویژگی‌های گروه‌ها و طبقات اختلاف زیادی با یکدیگر دارند. سپس از هر گروه و طبقه، به روش نمونه‌گیری تصادفی ساده و یا سیستماتیک نمونه‌گیری انجام می‌شود.

به عنوان مثال دانش‌آموزان یک مدرسه ابتدایی را در نظر بگیرید که در پایه‌های مختلف کلاس‌بندی شده‌اند و افراد هر پایه دارای خصوصیات و ویژگی‌های یکسان از قبیل سن، قد، وزن و موارد مشابه هستند. لذا به منظور انتخاب یک نمونه تصادفی به اندازه‌ی $n = ۳۰$ از بین دانش‌آموزان این مدرسه، ابتدا به روش نمونه‌گیری تصادفی ساده یا سیستماتیک از هر طبقه و پایه به نسبت افراد آن طبقه نمونه‌گیری می‌شود طوری که حاصل‌جمع نمونه‌های هر طبقه و پایه برابر $n = ۳۰$ باشد.

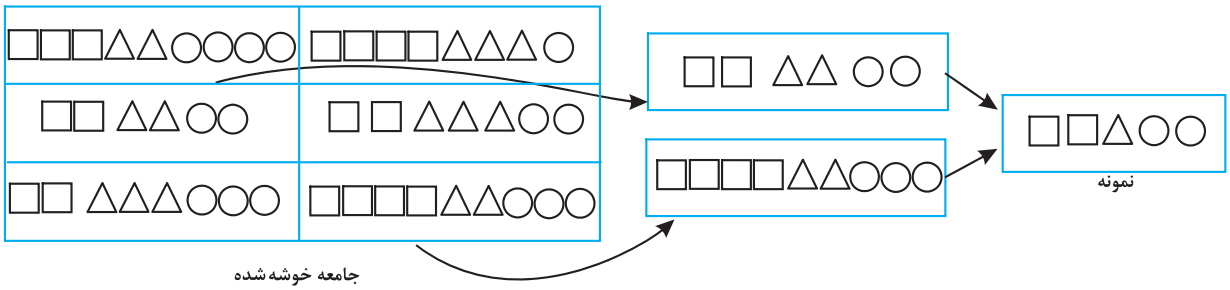
برای درک بهتر به شکل زیر توجه کنید.



۴ـ نمونه‌گیری خوشه‌ای

این روش وقتی به‌کار می‌رود که حجم جامعه مورد بررسی خیلی زیاد و گسترده بوده و اعضای جامعه فهرست نشده باشند. این شیوه نمونه‌گیری معمولاً بر اساس بخش‌های مجزایی که توسط نواحی جغرافیایی تعیین می‌شوند، به کار می‌رود. به عنوان مثال اگر میزان درآمد خانوار مورد بحث باشد و بخواهیم ۱۰۰۰ خانوار از شهرهای مختلف کشور انتخاب کنیم، لذا در انتخاب آن‌ها در روش نمونه‌گیری تصادفی ساده، این احتمال وجود دارد که بیشتر اعضای این نمونه به علت تراکم خانوار در استان تهران، محدود به این استان شوند و سهم استان‌های دیگر در برآورد مجموع درآمد خانوار کاهش یابد. در این گونه مسائل که حجم جامعه بزرگ است، بهتر است برای بالا بردن دقت برآورد از روش نمونه‌گیری خوشه‌ای استفاده شود.

برای چنین حالتی ابتدا از بین استان‌های کشور به‌طور تصادفی ۱۰استان انتخاب می‌شود، از بین هر استان نیز ۱۰ شهر به روش تصادفی ساده انتخاب می‌شود و در نهایت از هر شهر نیز ۱۰ خانوار به روش تصادفی ساده انتخاب می‌شود. در نتیجه یک نمونه ۱۰۰۰تایی از خانوارها داریم که می‌توانیم پرسشنامه مربوطه به درآمد را برایشان تکمیل کنیم. برای درک بهتر به شکل مقابل توجه کنید.



مثال۱: اگر جامعه آماری را بتوان به گروه‌های همگن تقسیم کرد، مناسب‌ترین روش نمونه‌گیری کدام است؟ (مدیریت - دکتری۹۶)

۱) هدفمند ۲) خوشه‌ای ۳) گروهی ۴) گلوله برفی

☑ **پاسخ:** گزینه «۳» در روش نمونه‌گیری گروهی، اعضاء جامعه به گروه‌های همگن تقسیم می شود.

مثال۲: مناسب‌ترین روش نمونه‌گیری از اتومبیل‌های سواری که وارد یک بزرگراه می‌شوند، کدام است؟ (علوم اقتصادی- سراسری ۸۶)

۱) سیستماتیک ۲) ساده ۳) خوشه‌ای ۴) طبقه‌بندی شده

☑ **پاسخ:** گزینه «۱» با توجه به فهرست نمودن پلاک اتومبیل‌های سواری که وارد یک بزرگراه می‌شوند، لذا روش نمونه‌گیری سیستماتیک می‌تواند مناسب‌ترین روش نمونه‌گیری باشد.

مثال۳: به‌طور معمول روزانه حدود ۳۰۰۰ نفر از دانشجویان یک دانشگاه در رستوران آن دانشگاه غذا می‌خورند. روش مناسب نمونه‌گیری برای انتخاب یک نمونه تصادفی ۱۰۰تایی از آنان در یک روز خاص کدام است؟ (علوم اقتصادی- سراسری ۹۸)

۱) سیستماتیک ۲) خوشه‌ای ۳) طبقه‌بندی شده ۴) ساده

☑ **پاسخ:** گزینه «۱» با توجه به اینکه دانشجویان دارای کد دانشجویی هستند، لذا انتخاب نمونه از فهرست کد دانشجویی به روش سیستماتیک انجام می‌شود.

توزیع میانگین جامعه نامتناهی

اگر $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از جامعه نامتناهی را تشکیل دهد که میانگین آن μ و واریانس آن σ^2 است، آنگاه:

$$۱)\ E(\bar{X}) = \mu_{\bar{X}} = \mu \qquad ۲)\ Var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

❖ **تعریف:** انحراف معیار نمونه‌ای را خطای معیار میانگین می‌گویند و به‌صورت مقابل تعریف می‌شود: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

مثال ۴: **توزیع میانگین نمونه تصادفی n تایی دارای انحراف معیار ۲ است. اگر انحراف معیار جامعه آماری ۱۲ باشد، مقدار نمونه (n) چقدر است؟**

(مدیریت و حسابداری ـ دکتری ۹۲)

۱) ۶ ۲) ۳۶ ۳) ۷۲ ۴) ۱۴۴

☑ **پاسخ:** گزینه «۲» طبق فرض، انحراف معیار نمونه برابر $\sigma_{\bar{X}} = ۲$ و انحراف معیار جامعه برابر $\sigma = ۱۲$ است. لذا حجم نمونه (n) با دستور زیر حساب می‌شود:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \Rightarrow ۲ = \frac{۱۲}{\sqrt{n}} = \sqrt{n} = ۶ \Rightarrow n = ۳۶$$

مثال ۵: **واریانس نمرات رضایت شغلی در جامعه ۲۲۵ است. درصورتی‌که بخواهیم انحراف معیار میانگین نمونه‌های تصادفی حداکثر ۳ باشد، حداقل حجم نمونه چقدر است؟**

(مدیریت ـ دکتری ۹۵)

۱) ۵۰ ۲) ۷۵ ۳) ۲۵ ۴) ۱۵

☑ **پاسخ:** گزینه «۳» در اینجا واریانس جامعه برابر $\sigma^2 = ۲۲۵$ و انحراف معیار میانگین نمونه حداکثر برابر $\sigma_{\bar{X}} = ۳$ است. بنابراین حداقل حجم نمونه به‌صورت مقابل حساب می‌شود:

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \Rightarrow ۹ = \frac{۲۲۵}{n} \Rightarrow n = \frac{۲۲۵}{۹} = ۲۵$$

مثال ۶: **میانگین توزیع نمونه‌گیری واریانس نمونه‌های ۳ تایی از جامعه‌ای نرمال با ۴ عضو برابر ۱۶ است. انحراف معیار توزیع نمونه‌گیری میانگین نمونه‌های ۳ تایی از این جامعه چقدر است؟**

(مدیریت و حسابداری ـ سراسری ۱۴۰۱)

۱) $\sqrt{۱۲}$ ۲) ۱۲ ۳) ۴ ۴) ۲

☑ **پاسخ:** گزینه «۴» طبق اطلاعات صورت مسئله داریم: $E(S^2) = ۱۶ = \sigma^2 \ ,\ N = ۴ \ ,\ n = ۳$

در نتیجه انحراف معیار میانگین نمونه‌ای برابر است با:

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \frac{۱۶}{۴} = ۴ \Rightarrow \sigma_{\bar{X}} = ۲$$

توزیع میانگین جامعه‌ی متناهی:

اگر $\bar{X}$ میانگین یک نمونه‌ی تصادفی به اندازه‌ی n از جامعه‌ای متناهی به حجم N با میانگین μ و واریانس σ^2 باشد، آنگاه:

$$۱)\ E(\bar{X}) = \mu \qquad ۲)\ Var(\bar{X}) = \frac{N-n}{N-۱}\left(\frac{\sigma^2}{n}\right)$$

مثال ۷: **درصورتی‌که از یک جامعه ۲۶ عضوی با واریانس ۴، نمونه‌ای ۱۶ عضوی گرفته شود، واریانس توزیع نمونه‌گیری میانگین نمونه کدام است؟**

(مدیریت ـ دکتری ۹۹)

۱) ۰/۱ ۲) ۱ ۳) ۰/۵ ۴) ۰/۲۵

☑ **پاسخ:** گزینه «۱» برای محاسبه واریانس میانگین نمونه‌ای، چون جامعه متناهی است، از رابطه $Var(\bar{X}) = \frac{N-n}{N-۱}\left(\frac{\sigma^2}{n}\right)$ استفاده می‌کنیم:

$$N = ۲۶ \quad \sigma^2 = ۴ \ ,\ n = ۱۶$$

$$Var(\bar{X}) = \left(\frac{N-n}{N-۱}\right)\left(\frac{\sigma^2}{n}\right) = \left(\frac{۲۶-۱۶}{۲۶-۱}\right)\left(\frac{۴}{۱۶}\right) = \frac{۱۰}{۲۵} \times \frac{۱}{۴} = ۰/۱$$

مثال۸: در یک جامعه ۳۶۰۱ عضوی، واریانس نمونه ۲۵ تایی برابر ۱۴۹ می‌باشد. انحراف معیار تصحیح شده آن کدام است؟

(مدیریت و حسابداری ـ سراسری ۹۲)

۱) $\sqrt{۶}$ ۲) $\frac{\sqrt{۱۴۹}}{۵}$ ۳) $\frac{۱۴۹\sqrt{۶}}{۱۵۰}$ ۴) $\frac{۴۹\sqrt{۶}}{۵۰}$

پاسخ: گزینه «۳» طبق اطلاعات صورت مسئله $N=۳۶۰۱$ ، $n=۲۵$ و $\sigma^۲=۱۴۹$ می‌باشد. بنابراین برای محاسبه انحراف معیار تصحیح شده، ابتدا واریانس نمونه‌ای به شکل زیر حساب می‌شود:

$$\sigma_{\bar{X}}^۲=\frac{N-n}{N-۱}(\frac{\sigma^۲}{n})=\frac{۳۶۰۱-۲۵}{۳۶۰۰}\times\frac{۱۴۹}{۲۵}=\frac{۱۴۹\times۲۴}{۳۶۰۰}\times\frac{۱۴۹}{۲۵}\Rightarrow\sigma_{\bar{X}}=\sqrt{\frac{۱۴۹\times۱۴۹\times۲۴}{۳۶۰۰\times۲۵}}=\frac{۱۴۹\sqrt{۲۴}}{۵\times۶۰}=\frac{۱۴۹\times۲\sqrt{۶}}{۵\times۶۰}=\frac{۱۴۹\sqrt{۶}}{۱۵۰}$$

مثال۹: عناصر یک جامعه آماری عبارتند از: ۴، ۷، ۱۰، ۶ و ۳، واریانس توزیع نمونه‌گیری میانگین نمونه‌های ۴تایی از این جامعه به روش با جایگذاری کدام است؟

(مدیریت ـ دکتری ۱۴۰۱)

۱) ۱/۶ ۲) ۰/۸ ۳) ۰/۴ ۴) ۰/۲

پاسخ: «هیچ‌کدام از گزینه‌ها صحیح نیست». در اینجا چون نمونه‌گیری به روش با جایگذاری است، لذا برای محاسبه واریانس میانگین نمونه‌ای از دستور $\text{Var}(\bar{X})=\frac{\sigma^۲}{n}$ استفاده می‌کنیم. داده‌های جامعه عبارتند از: ۳ , ۶ , ۱۰ , ۷ , ۴

$$\mu=\frac{\sum X_i}{N}=\frac{۳+۶+۱۰+۷+۴}{۵}=\frac{۳۰}{۵}=۶$$

$$\sigma^۲=\frac{\sum(X_i-\mu)^۲}{N}=\frac{(۳-۶)^۲+(۶-۶)^۲+(۱۰-۶)^۲+(۷-۶)^۲+(۴-۶)^۲}{۵}=\frac{۹+۰+۱۶+۱+۴}{۵}=\frac{۳۰}{۵}=۶$$

درنتیجه به ازای $n=۴$ داریم: $\text{Var}(\bar{X})=\frac{۶}{۴}=۱/۵$

سازمان سنجش گزینه (۲) را به عنوان گزینه صحیح انتخاب کرده است که اشتباه است.

مثال۱۰: یک نمونه ۲۴ تایی از یک جامعه ۱۴۵ عضوی با واریانس ۶ ، انتخاب می‌شود. انحراف معیار توزیع میانگین نمونه کدام است؟

(مدیریت و حسابداری ـ سراسری ۹۱)

۱) $\frac{۷}{۱۲}$ ۲) $\frac{۱۱}{۱۲}$ ۳) $\frac{۱۲}{۳۵}$ ۴) $\frac{۱۱}{۲۴}$

پاسخ: گزینه «۴» چون $n=۲۴$ ، $N=۱۴۵$ و $\sigma^۲=۶$ است، بنابراین واریانس نمونه برابر است با:

$$\sigma_{\bar{X}}^۲=\frac{N-n}{N-۱}(\frac{\sigma^۲}{n})=\frac{۱۴۵-۲۴}{۱۴۴}\times\frac{۶}{۲۴}=\frac{۱۲۱}{۱۴۴}\times\frac{۱}{۴}\Rightarrow\sigma_{\bar{X}}=\sqrt{\frac{۱۲۱}{۴\times۱۴۴}}=\frac{۱۱}{۲\times۱۲}=\frac{۱۱}{۲۴}$$

مثال۱۱: میانگین توزیع نمونه‌گیری واریانس نمونه‌های تصادفی ۳ تایی از جامعه‌ای نرمال برابر ۲۰ و انحراف معیار توزیع نمونه‌گیری میانگین نمونه‌های ۴تایی از این جامعه ۲ است. تعداد اعضای این جامعه، کدام است؟

(مدیریت و حسابداری ـ سراسری ۱۴۰۲)

۱) ۴ ۲) ۵ ۳) ۱۶ ۴) ۲۵

پاسخ: گزینه «۳» طبق فرض مسئله داریم:

$$E(S^۲)=۲۰\xrightarrow{E(S^۲)=\sigma^۲}\sigma^۲=۲۰$$

$$\sigma_{\bar{x}}=۲, n=۴, \text{Var}(\bar{x})=\frac{N-n}{N-۱}(\frac{\sigma^۲}{n})\xrightarrow{\text{Var}(\bar{x})=۴}\frac{N-۴}{N-۱}(\frac{۲۰}{۴})=۴\Rightarrow ۵N-۲۰=۴N-۴\Rightarrow N=۱۶$$

و چون برای نمونه‌های ۴تایی می‌باشد بنابراین $n=۴$ است، پس خواهیم داشت: $\frac{N-۴}{N-۱}\frac{\sigma^۲}{۴}=۴$

از طرفی میانگین واریانس نمونه برابر ۲۰ می‌باشد بنابراین برآورد $\sigma^۲$ برابر ۲۰ است و خواهیم داشت:

$$\frac{N-۴}{N-۱}\times\frac{۲۰}{۴}=۴\Rightarrow\frac{N-۴}{N-۱}=\frac{۴}{۵}\Rightarrow ۴N-۴=۵N-۲۰\Rightarrow N=۱۶$$

مدرسان شریف

فصل ششم

«نظریه برآورد»

مقدمه

کلیه مسائل مربوط به آمار استنباطی عموماً به مسئله برآورد و یا آزمون فرض برای پارامترهای جامعه منتهی می‌گردد. مسئله برآورد در این فصل و آزمون فرض در فصل بعد ارائه شده است. درخصوص موضوع برآورد یا تخمین، در بسیاری از مسائل علمی و کاربردی پارامترهای جامعه از قبیل میانگین، واریانس و نسبت جامعه نامعلوم است و محاسبه آن‌ها به‌طور مستقیم به دلایل مختلفی که در فصل قبل گفتیم دشوار و یا غیرممکن است. لذا به منظور برآورد آن‌ها با استفاده از یک آماره مناسب به دست آمده براساس یک نمونه انتخابی از جامعه موردنظر، برآورد یا تخمینی از پارامترهای نامعلوم جامعه ارائه می‌گردد. به عنوان مثال تخمین متوسط عمر یک نوع لامپ الکتریکی تولیدی، براساس یک نمونه انتخابی به اندازه $n = ۳۰$ انجام می‌شود و امکان آزمایش بر روی تمام محصولات تولیدی به منظور محاسبه پارامتر موردنظر وجود ندارد.

این فصل شامل سه درسنامه است. درسنامه اول برآورد نقطه‌ای، درسنامه دوم برآورد فاصله‌ای و درسنامه سوم برآورد تعداد نمونه می‌باشد. برآوردگر نقطه‌ای و فاصله‌ای به‌صورت زیر تعریف می‌شود.

الف) برآوردکننده نقطه‌ای: برآوردگری است که تنها یک مقدار عددی را به عنوان تقریبی از پارامتر جامعه ارائه می‌کند.

ب) برآوردکننده فاصله‌ای: برآوردکننده فاصله‌ای تخمین می‌زند که پارامتر نامعلوم جامعه با ضریب اطمینان معلوم در چه فاصله عددی قرار دارد.

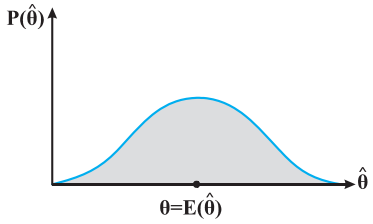
درسنامه (۱): برآورد نقطه‌ای

معمولاً برای هر پارامتر جامعه، برآوردکننده‌های متعددی وجود دارد و مقدار هر برآوردگر براساس یک نمونه انتخابی از جامعه موردنظر تعیین می‌شود و چون نمونه به صورت تصادفی انتخاب می‌شود، پس برآوردگرها خود متغیر تصادفی و دارای توزیع احتمال هستند. به عنوان مثال وقتی میانگین واقعی جامعه یعنی μ براساس میانگین یک نمونه تصادفی یعنی $\bar{x}$ برآورد می‌شود، چون با انتخاب نمونه‌های مختلف مقادیر متفاوتی برای میانگین نمونه به‌دست می‌آید، پس $\bar{X}$ متغیر تصادفی و دارای میانگین و واریانس می‌باشد. در اینجا نکته اساسی و مهم که باید به آن توجه داشت، این است که به ندرت مقدار $\bar{x}$ (میانگین نمونه) با μ (میانگین جامعه) برابر می‌شود اما انتظار اینکه مقدار متوسط $\bar{X}$ یعنی $E(\bar{X})$ برابر μ شود، راضی‌کننده است. موضوع مهم دیگری که به آن اشاره شد، وجود برآوردکننده‌های مختلف برای پارامتر نامعلوم جامعه است. به عبارت دیگر برای پارامتر نامعلوم جامعه، ممکن است بیش از یک برآوردگر وجود داشته باشد. به عنوان مثال $\bar{X}$ (میانگین نمونه) و $\tilde{X}$ (میانه نمونه) هر دو برآوردکننده نقطه‌ای برای μ (میانگین جامعه) می‌باشد. لذا انتخاب یک برآوردگر با کمترین واریانس ممکن، سبب کاهش خطای تقریب و افزایش دقت تخمین می‌گردد. در برآورد نقطه‌ای، معمولاً پارامتر مجهول جامعه را با نماد θ و برآوردکننده آن را با نماد $\hat{\theta}$ نشان می‌دهند. به عنوان مثال $\hat{\theta} = S^۲$ (واریانس نمونه) برآوردگر برای پارامتر $\theta = \sigma^۲$ (واریانس جامعه) است یا $\hat{\theta} = \bar{X}$ (میانگین نمونه) برآوردگر برای پارامتر $\theta = \mu$ (میانگین جامعه) است و یا $\hat{\theta} = \bar{P}$ (نسبت نمونه‌ای) برآوردگر برای پارامتر $\theta = P$ (نسبت جامعه) می‌باشد. اکنون با توجه به ویژگی‌های گوناگون برآوردکننده‌ها، برآوردگری را انتخاب می‌کنیم که قابل اعتمادتر بوده و نسبت به بقیه در شرایط بهتری باشد. لذا به منظور انتخاب یک برآوردگر مناسب، مفاهیم ناریب، کمترین واریانس، کارایی، سازگاری و بسندگی برآوردگرها مطرح می‌شود.

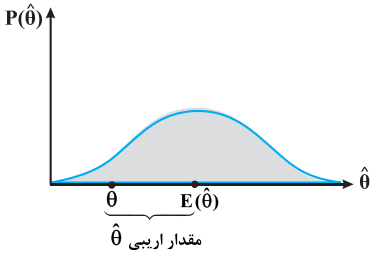
برآوردکننده ناریب

همان‌طورکه گفته شد، برای هر پارامتر مجهول جامعه، متناظر با نمونه‌های انتخابی مختلف، برآوردکننده‌های مختلفی به‌دست می‌آید که لزوماً برابر پارامتر جامعه نمی‌شود اما انتظار داریم که مقدار متوسط برآوردکننده با پارامتر جامعه برابر گردد. از آنجا که برآوردکننده خود متغیر تصادفی است، بنابراین دارای میانگین است.

بنابراین اگر میانگین برآوردگر $\hat{\theta}$ همان‌طور که در شکل مقابل نشان داده شده است، به‌طور دقیق بر θ (پارامتر واقعی) منطبق باشد، $\hat{\theta}$ برآوردکننده‌ی ناریب برای θ خواهد بود.



چنانچه $E(\hat{\theta})$ با θ متفاوت باشد، در این صورت $\hat{\theta}$ برآوردگر اریب برای θ خواهد بود. در شکل مقابل اریب نشان داده شده است.



در یک جامعه نرمال، میانگین نمونه و میانه نمونه هر دو برآوردگر ناریب μ_X هستند. اینکه عملکرد کدام بهتر است، لازم است ویژگی‌های دیگر آنها بررسی گردد.

❖ **تعریف:** برآوردکننده $\hat{\theta}$ به عنوان برآوردگر ناریب برای پارامتر θ جامعه است اگر و تنها اگر $E(\hat{\theta})=\theta$ باشد و چنانچه $E(\hat{\theta})\neq\theta$، $\hat{\theta}$ برآوردگر اریب برای θ است در این حالت مقدار اریبی برابر است با: $\text{مقدار اریبی}=E(\hat{\theta})-\theta$

نکته۱: فرض کنید $X_1,X_2,\ldots,X_n$ یک نمونه تصادفی n تایی از جامعه‌ای با میانگین μ باشد. در این صورت به ازای مقادیر ثابت $a_1,a_2,\ldots a_n$ عبارت $T=a_1X_1+a_2X_2+\ldots+a_nX_n$ وقتی برآوردگر ناریب برای μ است که داشته باشیم: $a_1+a_2+\cdots+a_n=1$

مثال۱: اگر X_1، X_2 و X_3 سه متغیر تصادفی ناریب برای برآورد θ باشند، آنگاه α در رابطه زیر چقدر باشد تا $\hat{\theta}$ برآوردکننده ناریب θ باشد؟

(مدیریت و حسابداری ـ دکتری ۱۴۰۰)

$$\hat{\theta}=\frac{1}{3}X_1+X_2+\alpha X_3$$

۱) $-\frac{1}{3}$ ۲) $-\frac{2}{3}$

۳) $\frac{1}{3}$ ۴) $\frac{2}{3}$

☑ **پاسخ:** گزینه «۱» شرط اینکه $\hat{\theta}=\frac{1}{3}X_1+X_2+\alpha X_3$ برآوردگر ناریب برای پارامتر θ باشد این است که جمع ضرایب برابر ۱ باشد؛ یعنی:

$$\frac{1}{3}+1+\alpha=1\Rightarrow\alpha=-\frac{1}{3}$$

مثال۲: فرض کنید X_1,X_2,X_3 یک نمونه تصادفی از جامعه‌ای با میانگین μ باشد. از برآوردکننده‌های زیر کدام‌یک برای پارامتر $\theta=\mu$ ناریب است؟

$$T_1=\frac{X_1-X_2+2X_3}{3}\qquad T_2=\frac{X_1+X_2+X_3}{3}\qquad T_3=\frac{X_1+2X_2+3X_3}{6}$$

۱) T_1 ۲) T_1,T_2

۳) T_2,T_3 ۴) T_1,T_3

☑ **پاسخ:** گزینه «۳» عبارت‌های کسری T_1 و T_2 و T_3 را تفکیک می‌کنیم.

$$T_1=\frac{1}{3}X_1-\frac{1}{3}X_2+\frac{2}{3}X_3$$

در T_1 جمع ضرایب برابر است با: $\frac{1}{3}-\frac{1}{3}+\frac{2}{3}=\frac{2}{3}$ در نتیجه T_1 برآوردگر ناریب برای $\mu=\theta$ نیست.

$$T_2=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3$$

در T_2 جمع ضرایب برابر است با: $\frac{1}{3}+\frac{1}{3}+\frac{1}{3}=1$ در نتیجه T_2 برآوردگر ناریب برای $\mu=\theta$ است.

$$T_3=\frac{1}{6}X_1+\frac{2}{6}X_2+\frac{3}{6}X_3$$

همچنین در عبارت T_3 جمع ضرایب برابر است با: $\frac{1}{6}+\frac{2}{6}+\frac{3}{6}=1$ در نتیجه T_3 نیز برآوردگر ناریب برای $\mu=\theta$ است.

مثال ۳: فرض کنید نمونه‌ای به حجم ۸ از توزیع پواسون با پارامتر λ انتخاب کرده‌ایم که λ پارامتر نامعلوم است. کدام عبارت برآوردگری ناریب برای λ است؟ (علوم اقتصادی ـ سراسری ۹۴)

۱) $\frac{1}{2}X_1+\frac{3}{4}X_2-\frac{1}{4}X_3$ ۲) $\frac{1}{n-1}\sum_{i=1}^{n}X_i$ ۳) $\frac{1}{2}X_1+\frac{1}{n-1}\sum_{i=1}^{n}X_i$ ۴) $e^{\bar{X}}$

پاسخ: گزینه «۱» مجموع ضرایب در عبارت $\frac{1}{2}X_1+\frac{3}{4}X_2-\frac{1}{4}X_3$ برابر $\frac{1}{2}+\frac{3}{4}-\frac{1}{4}=1$ است. بنابراین عبارت $T=\frac{1}{2}X_1+\frac{3}{4}X_2-\frac{1}{4}X_3$ برآوردگر ناریب برای پارامتر λ (میانگین توزیع پواسون) می‌باشد.

مثال ۴: فرض کنید X_1 و X_2 یک نمونه تصادفی از توزیعی با تابع چگالی احتمال زیر باشد. آماره $C(X_1+2X_2)$ به‌ازای چه مقداری از C یک برآوردگر ناریب $\frac{1}{\theta}$ است؟ $(f_\theta(x)=2\theta^2x;\ 0<x<\frac{1}{\theta})$ (علوم اقتصادی ـ سراسری ۹۷)

۱) ۲ ۲) $\frac{1}{3}$ ۳) ۳ ۴) $\frac{1}{2}$

پاسخ: گزینه «۴» طبق تعریف $T=C(X_1+2X_2)$ وقتی برآوردگر ناریب برای $\frac{1}{\theta}$ است که داشته باشیم: $E[C(X_1+2X_2)]=\frac{1}{\theta}$

طبق ویژگی‌های امید ریاضی می‌توان نوشت: $E[C(X_1+2X_2)]=C(E(X_1)+2E(X_2))=\frac{1}{\theta}$

اکنون $E(X_1)$ و $E(X_2)$ را به‌صورت مقابل حساب می‌کنیم.

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{\frac{1}{\theta}}2\theta^2xxdx=2\theta^2\int_{0}^{\frac{1}{\theta}}x^2dx=\frac{2}{3}\theta^2(x^3)\Big|_{0}^{\frac{1}{\theta}}=\frac{2}{3}\theta^2(\frac{1}{\theta})^3=\frac{2}{3\theta}\Rightarrow E(X_1)=E(X_2)=\frac{2}{3\theta}$$

در نتیجه از تساوی $C(E(X_1)+2E(X_2))=\frac{1}{\theta}$ داریم:

$$C(\frac{2}{3\theta}+\frac{4}{3\theta})=\frac{1}{\theta}\Rightarrow C(\frac{6}{3\theta})=\frac{1}{\theta}\Rightarrow 2C=1\Rightarrow C=\frac{1}{2}$$

مثال ۵: در جامعه‌ای ۳۶% کالاهای مصرفی وارداتی است. یک نمونه تصادفی ۱۶۹ تایی از کالاهای مصرفی انتخاب شده است. انحراف معیار برآورد نسبت کالاهای وارداتی کدام است؟ (علوم اقتصادی ـ سراسری ۱۴۰۲)

۱) $\frac{12}{325}$ ۲) $\frac{12}{352}$ ۳) $\frac{13}{48}$ ۴) $\frac{352}{12}$

پاسخ: گزینه «۱» برای پاسخ به این سؤال از توزیع دوجمله‌ای با پارامتر $n=169$، $p=0/36$ و $q=0/64$ استفاده می‌کنیم.

$$\sigma_{\hat{p}}=\sqrt{\frac{pq}{n}}=\sqrt{\frac{0/36\times0/64}{169}}=\frac{0/6\times0/8}{13}=\frac{0/48}{13}=\frac{48}{1300}=\frac{12}{325}$$

مثال ۶: به ازای چه مقدار از k برآوردکننده $\hat{\theta}=\frac{X}{k}$ برآوردکننده بدون تورشی (ناریبی) از پارامتر θ جامعه‌ای است که دارای تابع احتمال گسسته $\begin{cases}f(x)=\theta^x(1-\theta)^{1-x} & ;\ x=0,1\\ f(x)=0 & ;\ \text{سایر جاها}\end{cases}$ است؟ (علوم اقتصادی ـ سراسری ۸۵)

۱) $\frac{1}{2}$ ۲) ۱ ۳) ۲ ۴) ۵

پاسخ: گزینه «۲» طبق تعریف، $\hat{\theta}$ وقتی برآوردگر ناریب برای پارامتر θ است که داشته باشیم: $E(\hat{\theta})=\theta$. طبق فرض مسئله $\hat{\theta}=\frac{X}{k}$ و X دارای توزیع برنولی با پارامتر θ است. بنابراین از تساوی $E(\hat{\theta})=\theta$ به‌ازای $\hat{\theta}=\frac{X}{k}$ داریم:

$$E(\frac{X}{k})=\theta\Rightarrow\frac{1}{k}E(X)=\theta$$

در توزیع برنولی $E(X)=\theta$ است، بنابراین داریم:

$$\frac{1}{k}(\theta)=\theta\Rightarrow\frac{1}{k}=1\Rightarrow k=1$$

مثال ۷: تابع چگالی احتمال X عبارت است از $f(x)=\begin{cases}\frac{1}{\theta} & ;\ 0<x<\theta \\ 0 & ;\ \text{سایر جاها}\end{cases}$ به ازای کدام مقداری از k تخمین‌زن $\hat{\theta}=kX$ می‌تواند تخمین‌زنی ناتورش (ناریب) برای پارامتر θ در جامعه باشد؟ (علوم اقتصادی ـ سراسری ۸۴)

۱) $\frac{1}{2}$ ۲) ۱ ۳) ۲ ۴) ۳

☑ **پاسخ:** گزینه «۳» طبق تعریف، $\hat{\theta}$ وقتی برآوردگر ناریب برای θ است که داشته باشیم $E(\hat{\theta})=\theta$. طبق فرض مسئله $\hat{\theta}=kX$ بنابراین داریم:

$$E(kX)=\theta \Rightarrow kE(X)=\theta$$

X در اینجا دارای توزیع یکنواخت پیوسته با $a=0$ و $b=\theta$ است. در توزیع یکنواخت پیوسته $E(X)=\frac{a+b}{2}=\frac{0+\theta}{2}=\frac{\theta}{2}$ است.

اکنون از تساوی $kE(X)=\theta$ نتیجه می‌شود:

$$k(\frac{\theta}{2})=\theta \Rightarrow \frac{1}{2}k=1 \Rightarrow k=2$$

مثال ۸: نمونه‌ای تصادفی به اندازه n از جامعه‌ای با توزیع نرمال و با میانگین و واریانس به‌ترتیب $(1+\mu)$ و ۱ انتخاب شده است. برآوردگر $T=\frac{\Sigma X^2}{n}-2X$ یک برآوردگر ناأریب برای کدام پارامتر است؟ (علوم اقتصادی ـ سراسری ۱۴۰۱)

۱) μ^2 ۲) μ ۳) (μ^2-1) ۴) $(\mu-1)^2$

☑ **پاسخ:** گزینه «۱» طبق اطلاعات صورت مسئله داریم:

$$E(X)=1+\mu \ ,\ \sigma^2=1$$

$$T=\frac{1}{n}\sum X^2-2X \Rightarrow E(T)=E[(\frac{1}{n}\sum X^2-2X)]$$

$$=\frac{1}{n}\sum E(X^2)-2E(X)=\frac{1}{n}(nE(X^2))-2E(X)=E(X^2)-2E(X)$$

از دستور واریانس متغیر تصادفی X داریم:

$$E(X^2)=\sigma^2+(E(X))^2=1+(1+\mu)^2=1+1+2\mu+\mu^2=\mu^2+2\mu+2$$

بنابراین:

$$E(T)=(\mu^2+2\mu+2)-2(1+\mu)=\mu^2$$

چون $E(T)=\mu^2$، بنابر تعریف برآوردگر ناریب نتیجه می‌شود، T برآوردگر ناریب برای μ^2 می‌باشد.

مثال ۹: اگر $\hat{\theta}$ برآوردکننده پارامتر θ با اریب (تورش) $k\theta+5$ باشد، کدام برآوردکننده زیر ناریب (بدون تورش) است؟ (علوم اقتصادی ـ سراسری ۸۶)

۱) $\frac{\hat{\theta}-5}{k}$ ۲) $\frac{\hat{\theta}-5}{k+1}$ ۳) $\frac{\hat{\theta}}{k}-\frac{5}{k+1}$ ۴) $(k+1)\hat{\theta}+\frac{5}{k+1}$

☑ **پاسخ:** گزینه «۲» برآوردگر T را طوری می‌یابیم که داشته باشیم $E(T)=\theta$. طبق فرض مسئله $k\theta+5$ برآوردگر اریب برای پارامتر θ است. یعنی:

$$\text{مقدار اریبی}=E(\hat{\theta})-\theta=k\theta+5$$

$$\Rightarrow E(\hat{\theta})=k\theta+\theta+5 \Rightarrow E(\hat{\theta})=(k+1)\theta+5 \Rightarrow E(\hat{\theta})-5=(k+1)\ \theta \Rightarrow E(\hat{\theta}-5)=(k+1)\theta \Rightarrow \frac{1}{k+1}E(\hat{\theta}-5)=\theta$$

اکنون با توجه به ویژگی امیدریاضی، می‌توان نوشت:

$$E(\frac{\hat{\theta}-5}{k+1})=\theta \Rightarrow T=\frac{\hat{\theta}-5}{k+1}$$

یک برآوردگر ناریب برای σ^2 واریانس جامعه

فرض کنید $X_1,X_2,...,X_n$ یک نمونه تصادفی به اندازه n از جامعه‌ای با میانگین معلوم μ باشد. در این صورت S^2 (واریانس نمونه‌ای) با دستور $S^2=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$ برآوردگر ناریب برای σ^2 (واریانس جامعه) است. یعنی:

$$E(S^2)=\sigma^2$$

نکته ۲: اگر میانگین جامعه $\mu=0$ باشد، در این صورت برای نمونه تصادفی $X_1,X_2,...,X_n$ عبارت $S^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ (واریانس نمونه‌ای) برآوردگر ناریب برای σ^2 (واریانس جامعه) است.

نکته۳: بـه ازای نمونـه تصـادفی $X_1, X_2, ..., X_n$ از جامعـه نامتنـاهی، اگـر μ نـامعلوم باشـد در ایـن صـورت، واریـانس نمونـه‌ای بـا دستور $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$، برآوردگر ناریب برای σ^2 است. یعنی: $E(S^2) = \sigma^2$

توجه نمایید که اگرچه S^2 برآوردکننده ناریب برای واریـانس جامعـه نامتنـاهی اسـت ولـی برآوردکننـده ناریـب یـک جامعـه متنـاهی نیسـت و لزومـاً S برآوردکننده ناریب σ نیست.

مثال۱۰: با فرض معلوم بودن μ میانگین جامعه، کدام‌یک از آزمون‌های زیر برآوردگر ناریبی از واریانس جامعه (σ^2) است؟ (علوم اقتصادی ـ سراسری ۹۳)

۱) $(\bar{X}-\mu)^2$ ۲) $\frac{\sum_{i}(X_i-\mu)^2}{n-1}$ ۳) $\frac{\sum_{i}(X_i-\mu)^2}{n}$ ۴) $\frac{(X_i-\mu)^2}{n}$

پاسخ: گزینه «۳» در حالتی‌که μ (میانگین جامعه) معلوم باشد، واریانس نمونه‌ای (S^2) با رابطه $S^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$ برآوردگـر ناریـب بـرای σ^2 (واریانس جامعه) است.

مثال۱۱: با توجه به اینکه می‌دانیم $S^2 = \frac{\sum(X_i-\bar{X})^2}{n-1}$ یک تخمین‌زننده بدون تورش برای پارامتر σ^2 است. درخصوص تخمین‌زننده S برای σ چه می‌توان گفت؟ (علوم اقتصادی ـ دکتری ۱۴۰۱)

۱) S یک تخمین‌زننده تورش‌دار (اریب‌دار) برای σ است.

۲) S یک تخمین‌زننده بدون تورش (ناریب) برای σ است.

۳) S یک تخمین‌زننده بدون تورش (ناریب) برای σ فقط در جامعه با توزیع نرمال است.

۴) S یک تخمین‌زننده بدون تورش (ناریب) برای σ فقط در جامعه با توزیع متقارن است.

پاسخ: گزینه «۱» در هیچ حالتی s (انحراف معیار نمونه‌ای) برآوردکننده ناریب برای σ (انحراف معیار جامعه) نمی‌باشد.

مثال۱۲: n متغیر تصادفی $X_1, ..., X_n$ با میانگین μ و واریانس σ^2 را درنظر بگیرید. اریبی و واریانس تخمین‌زن $\hat{\mu} = \frac{1}{n-10}\sum_{i=11}^{n} X_i$ به ترتیب از راست به چپ برابر کدام است؟ (علوم اقتصادی ـ سراسری ۹۳)

۱) صفر و $\frac{\sigma^2}{n-10}$ ۲) $\frac{-\mu}{n-10}, \frac{\sigma^2}{n}$ ۳) $\frac{-\mu}{n-10}, \frac{\sigma^2}{n-11}$ ۴) صفر و $\frac{\sigma^2}{n-11}$

پاسخ: گزینه «۱» طبق تعریف مقدار اریبی با رابطه زیر به‌دست می‌آید.

مقدار اریبی $= E(\hat{\mu}) - \mu = E[\frac{1}{n-10}\sum_{i=11}^{n} X_i] - \mu = \frac{1}{n-10}\sum_{i=11}^{n} E(X_i) - \mu = \frac{1}{n-10}\sum_{i=11}^{n}(\mu) - \mu = (\frac{n-10}{n-10}\mu) - \mu = \mu - \mu = 0$

همچنین حاصل $Var(\hat{\mu})$ با استفاده از ویژگی $Var(aX+b) = a^2 Var(X)$ برابر است با:

$$Var(\hat{\mu}) = Var(\frac{1}{n-10}\sum_{i=11}^{n} X_i) = \frac{1}{(n-10)^2}\sum_{i=11}^{n} Var(X_i) = \frac{1}{(n-10)^2}\sum_{i=11}^{n}\sigma^2 = \frac{(n-10)}{(n-10)^2}\sigma^2 = \frac{\sigma^2}{(n-10)}$$

کارایی برآوردگرها

همان‌طور که قبلاً اشاره شد، برای یک پارامتر نامعلوم معمولاً بیش از یک برآوردگر ناریب وجـود دارد. در ایـن حالـت بـرای بـه‌دسـت آوردن تقریـب بهتـر، برآوردگری انتخاب می‌شود که کمترین واریانس را داشته باشد. لذا با توجه به توضیحات بالا، اگر $\hat{\theta}_1$ و $\hat{\theta}_2$ دو برآوردگر ناریب برای پـارامتر θ باشـد در ایـن صورت $\hat{\theta}_1$ وقتی کاراتر از $\hat{\theta}_2$ است که داشته باشیم: $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

همچنین اندازه کارایی نسبی $\hat{\theta}_2$ نسبت به $\hat{\theta}_1$ برابر $\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$ می‌باشد.

نکته۴: در بین برآوردگرها، $\bar{X}$ (میانگین نمونه‌ای) برآوردگر ناریب با کمترین واریانس برای μ است.

مدرسان شریف

فصل هفتم

«آزمون فرض»

مقدمه

در آزمون فرض پژوهشگر به دنبال این است که براساس یافته‌ها و مشاهدات حدس و یا ادعایی را بپذیرد یا رد کند. به عنوان مثال یک کارشناس قصد دارد براساس یک نمونه انتخابی از بین محصولات یک شرکت تولیدی که یک نوع لامپ الکتریکی تولید می‌کند، این موضوع را بررسی کند، «آیا متوسط عمر هریک از محصولات این شرکت حداقل ۱۰۰ساعت یا خیر؟» یا یک تولیدکننده موادغذایی می‌خواهد بداند، «آیا احتمال اینکه مصرف‌کننده‌ای نوع جدید از بسته‌بندی را به بسته‌بندی قبلی ترجیح می‌دهد، واقعاً ۸۰ درصد است یا خیر؟» تمام این مسائل را می‌توان به زبان آزمون فرض‌های آماری بیان نمود و در مورد ادعاهای مطرح شده توسط محقق نتیجه‌گیری کرد. از آنجا که پایه و اساس آمار استنباطی بر مبنای احتمال و شانس است، بنابراین نتایجی که از این طریق حاصل می‌شود، با عدم قطعیت همراه است و نمی‌توان با قاطعیت فرضیه پژوهشی را قبول یا رد کرد. بنابراین آزمون فرض در یک سطح معنادار به اندازه α انجام می‌شود. $(0 \le \alpha \le 1)$

در آزمون فرض، همواره یک ادعا یا حدس به نام فرض H_0 (فرض صفر) در مقابل ادعای متقابل به نام فرض H_1 (فرض یک) در سطح معنادار به اندازه α مورد آزمون قرار می‌گیرد. به عنوان مثال فرضیه‌ای به شکل زیر مطرح شده است.

«متوسط عمر یک نوع محصول تولیدی ۱۲ ماه است.» به منظور پذیرش یا رد این ادعا در سطح خطای α، فرضیه پژوهش شامل فرض H_0 به صورت $H_0 : \mu = 12$ در مقابل فرض متقابل H_1 به شکل $H_1 : \mu \neq 12$ مدلسازی می‌شود تا با اطمینان $(1-\alpha)100$درصد «فرض H_0 را بپذیریم و H_1 را رد کنیم» یا «H_0 را رد کنیم و H_1 را بپذیریم.»

درسنامه (۱): تعاریف پایه‌ای آزمون فرض

ایده کلی آزمون فرض این است که محقق براساس یک نمونه انتخابی به اندازه n از جامعه موردنظر، می‌خواهد فرض $H_0 : \theta = \theta_0$ را در مقابل فرض $H_1 : \theta \neq \theta_0$ در سطح معنادار α آزمون کند. θ پارامتر نامعلوم جامعه و θ_0 عدد ثابت است. برای انجام این کار براساس نمونه انتخابی n تایی، آماره آزمون که با $\hat{\theta}$ نشان داده می‌شود، محاسبه می‌گردد. $\hat{\theta}$ به عنوان برآوردگر برای پارامتر نامعلوم θ است. اگر مقدار $\hat{\theta}$ تقریباً با θ_0 برابر باشد، دراین‌صورت فرض H_0 پذیرفته شده و H_1 را رد می‌کنیم و چنانچه اختلاف $\hat{\theta}$ با θ_0 زیاد باشد، در این صورت فرض H_0 را رد کرده و H_1 را می‌پذیریم.

نکته۱: فرض H_0 تساوی را دربرمی‌گیرد. به‌عبارت دیگر فرض صفر به‌صورت $=$ یا $\le$ یا $\ge$ می‌باشد و لذا فرض متقابل H_1 به‌صورت $\neq$ یا $<$ یا $>$ است.

مثال۱: یک تولیدکننده ادعا می‌کند که «حداقل ۸۰ درصد مصرف‌کننده‌ها از محصول این شرکت رضایت دارند.» فرضیه‌های آماری برای آزمون این ادعا کدام است؟

۱) $\begin{cases} H_0 : P > 0/8 \\ H_1 : P \le 0/8 \end{cases}$ ۲) $\begin{cases} H_0 : P \ge 0/8 \\ H_1 : P < 0/8 \end{cases}$ ۳) $\begin{cases} H_0 : P \le 0/8 \\ H_1 : P > 0/8 \end{cases}$ ۴) $\begin{cases} H_0 : P < 0/8 \\ H_1 : P \ge 0/8 \end{cases}$

☑ **پاسخ:** گزینه «۲» اگر نسبت افرادی که از محصول این شرکت راضی هستند، P باشد در این صورت طبق فرض مسئله $P \ge 0/8$. بنابراین ادعای متقابل به‌صورت $P < 0/8$ است. از آنجا که فرض صفر شامل تساوی است، لذا فرضیه آماری به‌صورت مقابل خواهد بود. $\begin{cases} H_0 : P \ge 0/8 \\ H_1 : P < 0/8 \end{cases}$

مثال ۲: **ادعا شده است که نرخ بیکاری جوانان کشور کمتر از ۱۴ درصد است، فرضیه آماری H_0 کدام است؟** (مدیریت و حسابداری ـ سراسری ۱۴۰۰)

۱) $P > 0/14$ ۲) $P \leq 0/14$ ۳) $P < 0/14$ ۴) $P \geq 0/14$

☑ **پاسخ:** گزینه «۴» ادعا شده است که نرخ بیکاری کمتر از ۱۴/۰ است. بنابراین $P < 0/14$. به این ترتیب، فرضیه متقابل به صورت $P \geq 0/14$ خواهد بود و چون فرضیه H_0 شامل تساوی است، نتیجه می‌گیریم:

$$H_0 : P \geq 0/14$$

مثال ۳: **تحقیقی در خصوص سبک رهبری (رابطه‌مداری ـ وظیفه‌مداری) در دست برنامه‌ریزی است. ادعا شده است که اکثر مدیران سازمان دارای سبک رابطه‌مداری هستند. H_0 کدام است؟** (مدیریت ـ دکتری ۹۶)

۱) $P < 0/50$ ۲) $P = 0/5$ ۳) $P \leq 0/5$ ۴) $P \geq 0/5$

☑ **پاسخ:** گزینه «۳» اگر P برابر نسبت مدیران سازمان با سبک رابطه‌مداری باشد، در این صورت وقتی اکثر مدیران سازمان دارای سبک رابطه‌مداری هستند که بیش از ۵۰ درصد آن‌ها با سبک رابطه‌مدار می‌باشند. یعنی $P > 0/50$ و لذا فرضیه متقابل به شکل $P \leq 0/50$ می‌شود و چون H_0 شامل تساوی است پس:

$$H_0 : P \leq 0/5$$

مثال ۴: **ادعا شده است، نسبت ضایعات در کارخانه (۱) کمتر از نسبت ضایعات در کارخانه (۲) است. فرضیه H_0 کدام است؟**

(مدیریت و حسابداری ـ سراسری ۱۴۰۱)

۱) $H_0 : p_1 \geq p_2$ ۲) $H_0 : \mu_1 \geq \mu_2$ ۳) $H_0 : p_1 \leq p_2$ ۴) $H_0 : \mu_1 \leq \mu_2$

☑ **پاسخ:** گزینه «۱» از ادعای صورت مسئله داریم: $P_1 < P_2$.

لذا ادعای متقابل به‌صورت $p_1 \geq p_2$ خواهد بود و چون فرض H_0 لازم است شامل تساوی باشد، بنابراین نتیجه می‌شود:

$$H_0 : p_1 \geq p_2$$

مثال ۵: **فردی ادعا کرده است که میانگین عمر لامپ‌های رشته‌ای کمتر از ۱۰۰۰۰ ساعت می‌باشد. اگر ما به‌دنبال شواهد قوی برای تأیید این ادعا باشیم، فرضیه‌های معتبر کدام است؟** (علوم اقتصادی ـ سراسری ۹۵)

۱) $\begin{cases} H_0 : \mu > 10000 \\ H_1 : \mu \leq 10000 \end{cases}$ ۲) $\begin{cases} H_0 : \mu = 10000 \\ H_1 : \mu > 10000 \end{cases}$ ۳) $\begin{cases} H_0 : \mu < 10000 \\ H_1 : \mu \geq 10000 \end{cases}$ ۴) $\begin{cases} H_0 : \mu = 10000 \\ H_1 : \mu < 10000 \end{cases}$

☑ **پاسخ:** گزینه «۴» گزینه‌های (۱) و (۳) رد است. زیرا H_0 باید شامل تساوی باشد. طبق فرض مسئله ادعا بر این است که $\mu < 10000$. فرض مقابل به شکل $\mu \geq 10000$ است. و چون H_0 باید شامل تساوی باشد پس فرضیه آزمون به شکل $\begin{cases} H_0 : \mu \geq 10000 \\ H_1 : \mu < 10000 \end{cases}$ می‌شود. توجه نمایید که هر تساوی به‌شکل $\mu = 10000$ را می‌توان به‌صورت $\mu \geq 10000$ نوشت.

خطای نوع اول و دوم

در آزمون فرض آماری، تصمیم‌گیری برای رد یا پذیرش فرضیه متکی بر احتمالات است. این نشان می‌دهد نوعی عدم قطعیت در تصمیم‌گیری برای رد یا پذیرش یک فرضیه وجود دارد و ممکن است فرضیه‌ای را رد کنیم درحالی‌که به واقع درست باشد و یا فرضیه‌ای را قبول کنیم درحالی‌که در واقعیت جامعه نادرست باشد. بنابراین رد فرضیه درست و قبول فرضیه نادرست دو نوع خطایی است که در آزمون فرض با آن مواجه هستیم. به‌عنوان مثال، یک شرکت سازنده دارویی جدید، می‌خواهد این فرض را که حداقل ۹۰ درصد بیماران مصرف‌کننده این دارو بهبود یافته‌اند، مورد آزمون قرار دهد. درصورتی‌که به خطا این فرض صفر را رد کنند که حداقل ۹۰ درصد بیماران با استفاده از داروی جدید بهبود یافته‌اند، مرتکب خطای اول می‌شوند و اگر به خطا این فرض صفر را قبول کنند که حداقل ۹۰ درصد بیماران با استفاده از این دارو بهبود یافته‌اند، مرتکب خطای نوع دوم خواهند شد. اکنون برای آزمون فرض $H_0 : \theta = \theta_0$ در مقابل فرض $H_1 : \theta \neq \theta_1$ دو نوع خطا به نام خطای نوع اول و دوم به شکل زیر تعریف می‌شود.

الف ـ خطای نوع اول: وقتی فرض H_0 درست باشد ولی آن را رد کنیم، خطای نوع اول رخ داده است. مقدار خطای نوع اول را با α نشان می‌دهیم و به‌صورت زیر تعریف می‌شود.

$$\alpha = P(H_0 \text{ رد} \mid H_0 \text{ درست است})$$

بنابر تعریف α، خطای نوع اول برابر است با «احتمال رد فرضیه درست H_0». یعنی:

$$\alpha = P(\text{رد فرضیه درست } H_0)$$

ب ـ خطای نوع دوم: وقتی فرض H_0 در واقعیت نادرست باشد ولی آن را بپذیریم، در این صورت خطای نوع دوم رخ داده است. مقدار خطای نوع دوم را با β نشان می‌دهیم و به‌صورت زیر تعریف می‌شود:

$$\beta = P(H_0 \text{ نادرست است} \mid \text{قبول } H_0)$$

با توجه به تعریف β، احتمال خطای نوع دوم برابر است با «احتمال قبول فرضیه نادرست H_0». یعنی:

$$\beta = P(\text{قبول فرضیه نادرست})$$

نکته ۲: از تعریف α, β داریم:

$$1-\alpha = 1-P(H_0 \text{ رد} \mid H_0 \text{ درست است}) = 1-P(H_0 \text{ درست} \mid \text{رد فرضیه}) = P(H_0 \text{ درست} \mid \text{قبول فرضیه})$$

همچنین:

$$1-\beta = 1-P(H_0 \text{ قبول} \mid H_0 \text{ نادرست است}) = 1-P(H_0 \text{ نادرست} \mid \text{قبول فرضیه نادرست}) = P(H_0 \text{ نادرست} \mid \text{رد فرضیه})$$

توان آزمون: در آزمون فرض $H_0: \theta = \theta_0$ در برابر $H_1: \theta \neq \theta_0$، به $1-\beta$ توان آزمون گفته می‌شود.

$$\text{توان آزمون} = \beta^* = 1-\beta$$

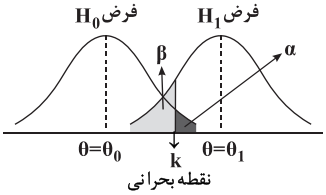
با توجه به تعریف توان آزمون داریم:

$$\text{توان آزمون} = \beta^* = 1-P(H_0 \text{ قبول} \mid H_0 \text{ رد فرض} \ H_1 \text{ درست باشد}) = P(H_0 \text{ رد فرض} \mid H_1 \text{ درست باشد})$$

ناحیه بحرانی آزمون

به ناحیه‌ای که در آن فرض H_0 در سطح معنی‌دار α رد می‌شود، ناحیه بحرانی (ناحیه رد H_0) به اندازه α گفته می‌شود.

برای آزمون فرض $H_0: \theta = \theta_0$ در مقابل $H_1: \theta = \theta_1$ ناحیه بحرانی و همچنین ارتباط بین α و β در شکل مقابل نشان داده شده است. همچنین آزمونی توانمند است که با توجه به مقدار مشخص α دارای خطای نوع دوم (β) کمتری باشد.



همان‌طور که در شکل ملاحظه می‌کنید α, β با یکدیگر رابطه معکوس داشته و داریم $\alpha + \beta \leq 1$. مطالب بالا را می‌توان به‌طور خلاصه در جدول زیر ارائه نمود.

تصمیم محقق / واقعیت جامعه	قبول H_0	رد H_0
H_0 درست است	$1-\alpha$ تصمیم‌گیری درست	α خطای نوع اول
H_0 نادرست است	β خطای نوع دوم	$1-\beta$ تصمیم‌گیری درست

مثال ۶: **کدام مورد بیانگر مفهوم توان آزمون در آزمون فرضیه آماری است؟** (مدیریت و حسابداری ـ دکتری ۱۴۰۱)

۱) احتمال رد H_0 درحالی‌که فرض H_0 درست است.
۲) احتمال رد H_0 درحالی‌که فرض H_1 درست است.
۳) احتمال قبول H_0 درحالی‌که فرض H_0 درست نیست.
۴) احتمال قبول H_0 درحالی‌که فرض H_1 درست نیست.

☑ **پاسخ:** گزینه «۲» توان آزمون به‌صورت زیر تعریف می‌شود:

$$\beta^* = 1-\beta = 1-P(H_0 \text{ قبول} \mid H_1 \text{ قبول است}) = P(H_0 \text{ رد فرضیه در صورتی‌که } H_1 \text{ درست است})$$

مثال ۷: **در ارتباط با خطای نوع اول α و خطای نوع دوم β کدام‌یک از موارد زیر درست است؟** (علوم اقتصادی ـ سراسری ۹۳)

۱) $\alpha + \beta = 1$
۲) $\alpha + \beta$ ممکن است برابر، کوچک‌تر یا بزرگ‌تر از واحد باشد.
۳) $\alpha + \beta < 1$
۴) $\alpha + \beta > 1$

☑ **پاسخ:** «هیچ‌یک از گزینه‌ها صحیح نیست.» زیرا: $\alpha + \beta \leq 1$

مثال ۸: **کدام مورد، نادرست است؟** (مدیریت و حسابداری ـ دکتری ۱۴۰۲)

۱) احتمال پذیرش H_0 درحالی‌که H_1 درست است، بیانگر توان آزمون است.

۲) احتمال رد فرض H_0 درحالی‌که H_0 درست است، بیانگر خطای نوع اول است.

۳) احتمال پذیرش فرض H_0 درحالی‌که H_0 نادرست است، بیانگر خطای نوع دوم است.

۴) احتمال پذیرش H_0 درحالی‌که H_1 نادرست است، بیانگر فاصله اطمینان است.

☑ **پاسخ:** گزینه «۱» خطای نوع اول، دوم و توان آزمون به‌صورت مقابل تعریف می‌شود:

$$\alpha = P(H_0 \text{ رد} \mid H_0 \text{ درست است})$$

$$\beta = P(H_0 \text{ قبول} \mid H_0 \text{ رد است})$$

$$\beta^* = 1-\beta = 1-P(H_0 \text{ قبول} \mid H_0 \text{ رد است}) = P(H_0 \text{ رد فرضیه نادرست})$$

مثال ۹: **در جدول زیر، که در مورد خطای نوع اول (α) و دوم (β) است، کدام گزینه درست است؟** (مدیریت ـ دکتری ۹۵)

	تصمیم	
واقعیت	رد کردن H_0	پذیرفتن H_0
H_0 درست است	a	b
H_0 نادرست است	c	d

۱) a: خطای نوع دوم، c: خطای نوع اول

۲) b: خطای نوع اول، d: خطای نوع دوم

۳) a: خطای نوع اول، d: خطای نوع دوم

۴) a: خطای نوع دوم، d: خطای نوع اول

☑ **پاسخ:** گزینه «۳» از جدول داده شده با توجه به واقعیت جامعه و تصمیم محقق داریم:

احتمال خطای نوع اول: $a = P(H_0 \text{رد} \mid H_0 \text{درست است}) = \alpha$

احتمال قبول فرضیه درست H_0: $b = P(H_0 \text{قبول} \mid H_0 \text{درست است}) = 1-\alpha$

احتمال رد فرضیه نادرست H_0: $c = P(H_0 \text{ رد} \mid H_0 \text{نادرست است}) = 1-\beta$

احتمال خطای نوع دوم: $d = P(H_0 \text{قبول} \mid H_0 \text{نادرست است}) = \beta$

مثال ۱۰: **کارخانه‌ای ادعا کرده است که میانگین محصولاتش، بیش از ۴۰ کیلوگرم است. معنی خطای نوع اول، کدام است؟** (مدیریت و حسابداری ـ سراسری ۱۴۰۲)

۱) میانگین محصولات کارخانه بیشتر از ۴۰ کیلوگرم است، درحالی‌که چنین نیست.

۲) میانگین محصولات کارخانه حداکثر ۴۰ کیلوگرم است، درحالی‌که چنین است.

۳) میانگین محصولات کارخانه بیشتر از ۴۰ کیلوگرم است، درحالی‌که چنین است.

۴) میانگین محصولات کارخانه حداکثر ۴۰ کیلوگرم است، درحالی‌که چنین نیست.

☑ **پاسخ:** گزینه «۱» طبق ادعای صورت مسئله $\mu > 40$ لذا ادعای متقابل به‌صورت $\mu \le 40$ خواهد بود و چون فرض H_0 شامل تساوی است پس فرضیه آزمون به شکل مقابل می‌باشد.

$$\begin{cases} H_0 : \mu \le 40 \\ H_1 : \mu > 40 \end{cases}$$

خطای نوع اول به‌صورت مقابل تعریف می‌شود: $\alpha = P(H_0 \text{ رد} \mid H_0 \text{ درست است}) = P(H_0 \text{ رد فرضیه درست})$

این معادل است با اینکه پژوهشگر فرضیه $\mu > 40$ را قبول کرده اما در واقعیت چنین نیست.

رابطه بین α, β، توان آزمون، حجم نمونه (n)، خطای برآورد (e):

۱ـ α, β رابطه معکوس دارند. با افزایش یکی دیگری کم و با کاهش یکی، دیگری افزایش می‌یابد.

۲ـ با افزایش حجم نمونه (n) مقدار α, β کاهش می‌یابد.

۳ـ $\alpha + \beta \le 1$

۴ـ احتمال خطای نوع اول (α) با توان آزمون $(\beta^* = 1-\beta)$ رابطه مستقیم دارد.

۵ـ احتمال خطای نوع دوم (β) با توجه به رابطه $\beta^* = 1-\beta$، با توان آزمون رابطه معکوس دارد.

۶ـ احتمال خطای نوع اول (α) با خطای برآورد (e) رابطه معکوس دارد.

۷ـ احتمال خطای نوع دوم (β) با خطای برآورد (دقت برآورد، e) رابطه مستقیم دارد.

۸ـ با افزایش حجم نمونه، توان آزمون $(1-\beta)$ افزایش می‌یابد.

مثال ۱۱: با افزایش سطح معنا (α) چه تغییری در توان آزمون ($1-\beta$) به‌وجود خواهد آمد؟ (علوم اقتصادی ـ سراسری ۹۱)

۱) توان آزمون کم می‌شود.
۲) توان آزمون تغییر نمی‌کند.
۳) توان آزمون زیاد می‌شود.
۴) بستگی به حجم نمونه دارد.

☑ **پاسخ:** گزینه «۳» α (سطح معنادار یا احتمال خطای نوع اول) با توان آزمون رابطه مستقیم دارد. بنابراین با افزایش α، مقدار توان آزمون $(1-\beta)$ نیز افزایش پیدا می‌کند.

مثال ۱۲: کدام مورد منجر به افزایش قدرت یا توان آزمون ($1-\beta$) برای فرضیه $H_0: \mu = \mu_0$ در مقابل $H_1: \mu \neq \mu_0$ می‌شود؟ (علوم اقتصادی ـ ارشد ۱۴۰۱)

۱) افزایش حجم نمونه
۲) افزایش واریانس جامعه
۳) کاهش α (خطای نوع اول)
۴) کاهش فاصله میانگین تحت فرضیه H_0 و میانگین واقعی یا $|\mu_0 - \mu|$

☑ **پاسخ:** گزینه «۱» توان آزمون $(1-\beta)$ با حجم نمونه (n) رابطه مستقیم دارد. بنابراین افزایش حجم نمونه (n) باعث افزایش $1-\beta$ می‌گردد.
توجه داشته باشید که $1-\beta$ با واریانس جامعه رابطه معکوس دارد با α (خطای نوع اول) رابطه مستقیم و با خطای برآورد $(e = |\mu_0 - \mu|)$ رابطه مستقیم دارد.

مثال ۱۳: کدام مورد درباره افزایش حجم نمونه، صحیح نیست؟ (مدیریت و حسابداری ـ دکتری ۹۳)

۱) خطای نوع دوم را کاهش می‌دهد.
۲) صحت برآورد را افزایش می‌دهد.
۳) باعث افزایش دقت برآورد می‌شود.
۴) خطای نوع اول را افزایش می‌دهد.

☑ **پاسخ:** گزینه «۴» با افزایش حجم نمونه خطای نوع اول و دوم کاهش و دقت برآورد و صحت برآورد افزایش می‌یابد.

مثال ۱۴: آزمون فرضیه مقابل را درنظر بگیرید: $H_0: \mu = \mu_0, H_1: \mu = \mu_A$ که در آن μ_0 و μ_A به ترتیب میانگین تحت فرضیه و میانگین جامعه می‌باشند ..... حجم نمونه و ..... فاصله $|\mu_0 - \mu_A|$، باعث افزایش توان آزمون می‌شود. (علوم اقتصادی ـ دکتری ۹۲)

۱) کاهش ـ افزایش
۲) کاهش ـ کاهش
۳) افزایش ـ کاهش
۴) افزایش ـ افزایش

☑ **پاسخ:** گزینه «۳» در اینجا μ_0 میانگین تحت فرضیه و μ_A میانگین جامعه می‌باشد بنابراین $|\mu_0 - \mu_A|$ برابر دقت برآورد (e) می‌باشد.
دقت برآورد (e) با توان آزمون رابطه معکوس دارد لذا توان آزمون وقتی افزایش می‌یابد که خطای برآورد $e = |\mu_0 - \mu_A|$ کاهش یابد. همچنین با افزایش حجم نمونه توان آزمون $(1-\beta)$ افزایش می‌یابد. درنتیجه افزایش حجم نمونه و کاهش فاصله $|\mu_0 - \mu_A|$ سبب افزایش توان آزمون می‌شود.

مثال ۱۵: این فرضیه را درنظر بگیرید: $H_0: \mu = \mu_0$ / $H_1: \mu \neq \mu_0$ چنانچه میانگین واقعی جامعه μ_a باشد، آنگاه با افزایش $|\mu_0 - \mu_a|$ خطای ..... .

(علوم اقتصادی ـ دکتری ۹۳)

۱) نوع دوم کاهش می‌یابد.
۲) نوع اول افزایش و خطای نوع دوم کاهش می‌یابد.
۳) نوع دوم افزایش می‌یابد.
۴) نوع اول کاهش و خطای نوع دوم افزایش می‌یابد.

☑ **پاسخ:** گزینه «۴» در اینجا μ_a مقدار واقعی میانگین جامعه و μ_0 تقریبی از آن می‌باشد. لذا اختلاف آن‌ها برابر خطای برآورد می‌باشد. یعنی $e = |\mu_0 - \mu_a|$. خطای برآورد با خطای نوع اول (α) رابطه معکوس و با خطای نوع دوم (β) رابطه مستقیم دارد.
اکنون با افزایش e (خطای برآورد) نتیجه می‌شود، خطای نوع اول کاهش و خطای نوع دوم افزایش می‌یابد.

مدرسان شریف

فصل هشتم

«رگرسیون و همبستگی»

مقدمه

یکی از هدف‌های اصلی بسیاری از پژوهش‌های آماری، پیش‌بینی یک یا چند متغیر برحسب سایر متغیرها یا ایجاد وابستگی‌ها بین آن‌ها است. به عنوان مثال مطالعاتی انجام می‌شود تا سطح فروش یک کالا برحسب قیمت آن یا وزن کودکان برحسب سن آن‌ها یا میزان فروش یک نوع محصول برحسب هزینه تبلیغ و موارد مشابه را پیش‌بینی نمایند.

درسنامه: بهترین تابع پیش‌بینی‌کننده

برای متغیرهای تصادفی X و Y با توزیع توأم $f(x,y)$ مسئله اصلی رگرسیون دو متغیره عبارت از تعیین متوسط مقدار Y به‌ازای مقدار مفروض $X=x$ می‌باشد. برای این منظور بهترین تابع پیش‌بینی‌کننده Y به شرط آنکه $X=x$ باشد برابر $E(Y|X=x)$ است که در حالت‌های گسسته و پیوسته به شکل زیر محاسبه می‌شود:

$$E(Y|X=x)=\mu_{Y|X=x}=\begin{cases}\sum_{y} yP(Y=y|X=x) & ;\ \text{اگر } x \text{ و } y \text{ گسسته باشند}\\ \int_{-\infty}^{+\infty} y\,f(y|x)\,dy & ;\ \text{اگر } x \text{ و } y \text{ پیوسته باشند}\end{cases}$$

مشابهاً بهترین تابع پیش‌بینی‌کننده X به شرط آنکه $Y=y$ باشد برابر $E(X|Y=y)$ است و در حالت‌های گسسته و پیوسته به‌صورت زیر محاسبه می‌شود:

$$E(X|Y=y)=\mu_{X|Y=y}=\begin{cases}\sum_{x} xP(X=x|Y=y) & ;\ \text{اگر } x \text{ و } y \text{ گسسته باشند}\\ \int_{-\infty}^{+\infty} x\,f(x|y)\,dx & ;\ \text{اگر } x \text{ و } y \text{ پیوسته باشند}\end{cases}$$

مثال ۱: متغیرهای تصادفی X و Y با تابع احتمال توأم زیر را در نظر بگیرید. بهترین تابع پیش‌بینی‌کننده Y به ازای $x=1$ کدام است؟

y \ x	-1	0	1
0	$0/2$	$0/05$	$0/1$
1	$0/15$	$0/1$	$0/2$
2	0	$0/1$	$0/1$

۱) $0/2$

۲) $0/4$

۳) $\frac{1}{4}$

۴) 1

☑ **پاسخ:** گزینه «۴» برای محاسبه بهترین تابع پیش‌بینی‌کننده Y به‌ازای $x=1$، مقدار $E(Y|x=1)$ را با استفاده از دستور زیر حساب می‌کنیم. چون متغیرهای تصادفی X و Y گسسته‌اند، لذا داریم:

$$E(Y|X=1)=\sum_{y=0}^{2} yp(Y=y|X=1)=0\times P(Y=0|X=1)+1\times P(Y=1|X=1)+2\times P(Y=2|X=1)$$

$$=0+\frac{P(X=1,Y=1)}{f_X(1)}+2\left(\frac{P(X=1,Y=2)}{f_X(1)}\right)=\frac{0/2}{0/1+0/2+0/1}+2\left(\frac{0/1}{0/1+0/2+0/1}\right)=\frac{0/2}{0/4}+\frac{0/2}{0/4}=1$$

مثال ۲: برای متغیرهای تصادفی X و Y با تابع چگالی مشترک زیر بهترین تابع پیش‌بینی‌کننده Y به ازای $X=\frac{۱}{۲}$ کدام است؟

$$f(x,y)=\begin{cases}x+y & ; \circ<x<۱ \quad \circ<y<۱\\ \circ & ; \quad \text{سایر جاها}\end{cases}$$

۱) $\frac{۱}{۳}$ ۲) $\frac{۷}{۱۲}$ ۳) $\frac{۱}{۴}$ ۴) $\frac{۱}{۲}$

پاسخ: گزینه «۲» برای محاسبه بهترین تابع پیش‌بینی‌کننده Y به‌ازای $x=\frac{۱}{۲}$، حاصل $E(Y|X=\frac{۱}{۲})$ را حساب می‌کنیم. چون متغیرهای تصادفی X و Y پیوسته‌اند، بنابراین حاصل $E(Y|X=\frac{۱}{۲})$ با دستور مقابل حساب می‌شود:

$$E(Y|X=\tfrac{۱}{۲})=\int_{-\infty}^{\infty} y f(y|\tfrac{۱}{۲})dy$$

ابتدا احتمال شرطی $f(y|\frac{۱}{۲})$ را محاسبه می‌کنیم. برای این منظور داریم:

$$f(y|\tfrac{۱}{۲})=\frac{f(\frac{۱}{۲},y)}{f_X(\frac{۱}{۲})}$$

در اینجا حاصل $f(\frac{۱}{۲},y)$ از روی تابع چگالی توأم داده شده برابر است با:

$$f(x,y)=x+y \Rightarrow f(\tfrac{۱}{۲},y)=\tfrac{۱}{۲}+y \quad ; \quad \circ<y<۱$$

همچنین برای محاسبه $f_X(\frac{۱}{۲})$، تابع چگالی حاشیه‌ای X را به‌صورت زیر حساب می‌کنیم:

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_{\circ}^{۱}(x+y)dy=(xy+\frac{y^۲}{۲})|_{\circ}^{۱}=(x+\tfrac{۱}{۲}) \Rightarrow f_X(\tfrac{۱}{۲})=\tfrac{۱}{۲}+\tfrac{۱}{۲}=۱$$

با محاسبه $f_X(\frac{۱}{۲})$ و $f(\frac{۱}{۲},y)$، حاصل $f(y|\frac{۱}{۲})$ برابر می‌شود با:

$$f(y|\tfrac{۱}{۲})=\frac{y+\frac{۱}{۲}}{۱}=y+\tfrac{۱}{۲} \quad , \quad \circ<y<۱$$

در نتیجه مقدار $E(Y|X=\frac{۱}{۲})$ برابر است با:

$$E(Y|X=\tfrac{۱}{۲})=\int_{\circ}^{۱} y(y+\tfrac{۱}{۲})dy=\int_{\circ}^{۱}(y^۲+\tfrac{۱}{۲}y)dy=(\frac{y^۳}{۳}+\frac{۱}{۴}y^۲)|_{\circ}^{۱}=\frac{۱}{۳}+\frac{۱}{۴}=\frac{۷}{۱۲}$$

رگرسیون خطی

بنابر دلایلی، معادلات رگرسیون خطی که به‌صورت $\mu_{Y|X}=\beta x+\alpha$ تعریف می‌شوند، مورد توجه خاص هستند. زیرا اغلب آن‌ها تقریب‌های خوبی برای معادلات رگرسیون پیچیده‌تر محسوب می‌شوند. در حالت توزیع نرمال دو متغیره، معادلات رگرسیون در واقع خطی هستند.

اکنون به منظور آسان نمودن مطالعه معادلات رگرسیون خطی، ضرایب رگرسیون α و β را برحسب بعضی گشتاورهای مرتبه پایین‌تر توزیع توأم X و Y یعنی برحسب $E(X)=\mu_۱$ ، $E(Y)=\mu_۲$ ، $Var(X)=\sigma_۱^۲$ و $Var(Y)=\sigma_۲^۲$ و $Cov(X,Y)=\sigma_{۱۲}$ بیان کرده و نتایج زیر حاصل می‌شود.

۱) اگر رگرسیون Y روی X خطی باشد، آنگاه:

$$\mu_{Y|x}=E(Y|x)=\mu_۲+\rho\frac{\sigma_۲}{\sigma_۱}(x-\mu_۱)$$

۲) اگر رگرسیون X روی Y خطی باشد، آنگاه:

$$\mu_{X|y}=E(X|y)=\mu_۱+\rho\frac{\sigma_۱}{\sigma_۲}(y-\mu_۲)$$

که در آن ρ (ضریب همبستگی X و Y) با رابطه $\rho=\frac{\sigma_{۱۲}}{\sigma_۱\sigma_۲}$ حساب می‌شود.

اکنون اگر ضریب همبستگی X و Y، برابر $\rho=\circ$ باشد نتیجه می‌شود:

$$E(Y|x)=\mu_۲=E(Y) \ , \ E(X|y)=\mu_۱=E(X)$$

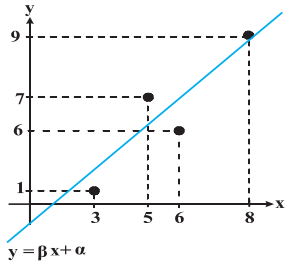
این نشان می‌دهد $E(Y|x)$ به x بستگی نداشته و لذا X و Y مستقل‌اند.

روش کمترین مربعات

در بخش قبل در حالتی که متغیرهای تصادفی دارای توزیع توأم بودند، پیش‌بینی یک متغیر برحسب متغیر دیگر مورد بررسی قرار گرفت. اکنون حالتی را در نظر بگیرید که در آن مجموعه‌ای از داده‌های زوج‌شده به شکل (x_i,y_i) در دسترس است و توزیع توأم آن‌ها را نمی‌دانیم. به عنوان مثال فرض کنید داده‌های زوج‌شده (x,y) مطابق جدول زیر و مربوط به میزان هزینه تبلیغ و درآمد ناشی از فروش یک نوع محصول باشد.

هزینه تبلیغ x	۵	۳	۸	۶
درآمد y	۷	۱	۹	۶

و مدیر علاقه‌مند به پیش‌بینی مقدار درآمد به ازای هزینه تبلیغ $x = ۳۰$ باشد. در اینجا تابعی داریم که شامل چهار نقطه به شکل مقابل است:



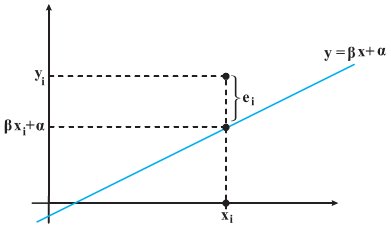
x	$x_۱$	$x_۲$	...	x_n
y	$y_۱$	$y_۲$	...	y_n

به منظور پیش‌بینی مقدار Y به‌ازای $x = ۳۰$ خط با معادله $Y = \beta X + \alpha$ را طوری مشخص می‌کنیم که فاصله این خط از نقاط معلوم (x_i, y_i) حداقل مقدار ممکن باشد. به چنین خطی، خط رگرسیون می‌گویند.

در معادله خط رگرسیون (خط کمترین مربعات) $\hat{Y} = \hat{\beta}X + \hat{\alpha}$، α و β را ضرایب رگرسیون می‌گویند که باید محاسبه شوند. $\hat{\alpha}$ عرض از مبدأ و $\hat{\beta}$ شیب خط $Y = \hat{\beta}X + \hat{\alpha}$ می‌باشد.

در حالت کلی فرض کنید نقاط زوج‌شده با مختصات $(x_۱, y_۱), (x_۲, y_۲), \ldots, (x_n, y_n)$ موجود باشد.

هدف تعیین ضرائب $\hat{\alpha}$ و $\hat{\beta}$ مربوط به خط رگرسیون $\hat{Y} = \hat{\beta}X + \hat{\alpha}$ است طوری‌که فاصله خط از نقاط (x_i, y_i) حداقل مقدار ممکن باشد.



در اینجا مقدار انحراف برابر $e_i = y_i - (\hat{\beta}x_i + \hat{\alpha})$ می‌باشد که لازم است مقدار e_i مینیمم گردد. $\hat{\alpha}$ و $\hat{\beta}$ مجهول است و مقادیر آن‌ها به منظور حداقل نمودن e_i به شکل مقابل به دست می‌آید.

$$F = \sum_{i=۱}^{n} e_i^۲ = \sum_{i=۱}^{n} (y_i - (\hat{\beta}x_i + \hat{\alpha}))^۲$$

به منظور مینیمم نمودن عبارت F برحسب ضرایب رگرسیون $\hat{\alpha}$ و $\hat{\beta}$، مشتقات جزیی را حساب کرده و مساوی صفر قرار می‌دهیم:

$$\frac{\partial F}{\partial \hat{\alpha}} = \sum_{i=۱}^{n} (-۲)(y_i - (\hat{\beta}x_i + \hat{\alpha})) = ۰$$

$$\frac{\partial F}{\partial \hat{\beta}} = \sum_{i=۱}^{n} (-۲)x_i(y_i - (\hat{\beta}x_i + \hat{\alpha})) = ۰$$

از تساوی‌های بالا دستگاه دو معادله و دو مجهول برحسب $\hat{\alpha}$ و $\hat{\beta}$ به شکل مقابل به دست می‌آید:

$$\begin{cases} n\hat{\alpha} + \hat{\beta}\sum_{i=۱}^{n} x_i = \sum_{i=۱}^{n} y_i \\ \hat{\alpha}\sum_{i=۱}^{n} x_i + \hat{\beta}\sum_{i=۱}^{n} x_i^۲ = \sum_{i=۱}^{n} x_i y_i \end{cases}$$

با حل دستگاه دو معادله، دو مجهول فوق نسبت به $\hat{\alpha}$ و $\hat{\beta}$ برآورد کمترین مربعات برای α و β به‌صورت زیر به دست می‌آید.

$$\hat{\beta} = \frac{n(\sum_{i=۱}^{n} x_i y_i) - (\sum_{i=۱}^{n} x_i)(\sum_{i=۱}^{n} y_i)}{n(\sum_{i=۱}^{n} x_i^۲) - (\sum_{i=۱}^{n} x_i)^۲}, \quad \hat{\alpha} = \frac{\sum_{i=۱}^{n} y_i - \hat{\beta}\sum_{i=۱}^{n} x_i}{n}$$

برای آسان کردن فرمول مربوط به $\hat{\beta}$، نمادهای مقابل را معرفی می‌کنیم.

$$S_{xx} = \sum_{i=۱}^{n} (x_i - \bar{x})^۲ = \sum_{i=۱}^{n} (x_i^۲ + \bar{x}^۲ - ۲\bar{x}x_i) = \sum_{i=۱}^{n} x_i^۲ - n(\bar{x})^۲$$

$$S_{xy} = \sum_{i=۱}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=۱}^{n} (x_i y_i - \bar{y}x_i - \bar{x}y_i + \bar{x}\,\bar{y}) = \sum_{i=۱}^{n} x_i y_i - n\bar{x}\,\bar{y}$$

$$S_{yy} = \sum_{i=۱}^{n} (y_i - \bar{y})^۲ = \sum_{i=۱}^{n} (y_i^۲ + \bar{y}^۲ - ۲\bar{y}y_i) = \sum_{i=۱}^{n} y_i^۲ - n(\bar{y})^۲$$

با توجه به روابط بالا، ضرایب خط کمترین مربعات $\hat{Y} = \hat{\beta}X + \hat{\alpha}$ عبارتند از:

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

که در آن:

$$\bar{y} = \frac{\sum_{i=۱}^{n} y_i}{n} \text{ و } \bar{x} = \frac{\sum_{i=۱}^{n} x_i}{n}$$

توجه نمایید که S_{xx} و S_{yy} را با علایم مقابل نیز نشان می‌دهند.

$$S_{xx} = SS_x = S_x^2 \Rightarrow S_x = \sqrt{S_x^2}$$
$$S_{yy} = SS_y = S_y^2 \Rightarrow S_y = \sqrt{S_y^2}$$

ویژگی‌های خط رگرسیون

۱) نقطه $(\bar{x}, \bar{y})$ در معادله خط رگرسیون صدق می‌کند، یعنی: $\bar{y} = \hat{\beta}\bar{x} + \hat{\alpha}$

۲) میانگین انحرافات برابر صفر است. یعنی: $e_i = \hat{y}_i - (\hat{\beta}x_i + \hat{\alpha}) \quad ; \quad i = 1, \ldots, n \Rightarrow \bar{e} = \frac{\sum_{i=1}^{n} e_i}{n} = 0$

۳) $\sum_{i=1}^{n} e_i y_i = 0$

۴) $\sum_{i=1}^{n} e_i x_i = 0$

۵) در خط رگرسیون جامعه با رابطه $Y = \alpha + \beta X$، توزیع‌های Y به ازای مقادیر مختلف X، همه نرمال با انحراف معیار یکسان σ هستند.

۶) e_iها مقادیر n متغیر تصادفی نرمال با میانگین صفر و واریانس مشترک σ^2 هستند.

مثال ۳: **کدام جمله در مورد مقادیر خطا (e_i) در یک مدل رگرسیونی نادرست است؟** (مدیریت ـ دکتری ۹۹)

۱) متغیر تصادفی (e_i) دارای توزیع نرمال باشد.
۲) مقادیر خطاها (e_i) مستقل از یکدیگر باشند.
۳) واریانس مقادیر خطا (e_i) به ازای مقادیر x متغیر باشد.
۴) مقدار مورد انتظار مقادیر خطا (e_i) برابر با صفر باشد.

☑ **پاسخ:** گزینه «۳» در معادله خط رگرسیون $y_i = \beta_i x + \alpha + e_i$، e_iها دارای توزیع نرمال با میانگین صفر و واریانس مشترک هستند.

مثال ۴: **اطلاعات زیر از بررسی رابطه بین میزان کود مصرفی با میزان محصولات ذرت برای مدت ۱۰ سال به‌دست آمده است. معادله خط رگرسیون کدام است؟**

$$\sum (x - \bar{x})^2 = 100 \,,\ \sum (x - \bar{x})(y - \bar{y}) = 160 \,,\ \sum x = 200 \,,\ \sum y = 350$$

(علوم اقتصادی ـ سراسری ۱۴۰۰)

۱) $y = 8 - 1/6x$
۲) $y = 3 + 1/6x$
۳) $y = 12 + 0/8x$
۴) $y = 9 + 0/75x$

☑ **پاسخ:** گزینه «۲» معادله خط رگرسیون به‌صورت $Y = \beta x + \alpha$ است که ضرایب رگرسیون β, α به‌صورت زیر حساب می‌شوند:

$$\beta = \frac{S_{xy}}{Ss_x} \quad , \quad \alpha = \bar{y} - \beta\bar{x}$$

در اینجا $n = 10$ است. لذا داریم:

$$\left.\begin{aligned} S_{xy} &= \sum (x_i - \bar{x})(y_i - \bar{y}) = 160 \\ Ss_x &= \sum (x_i - \bar{x})^2 = 100 \end{aligned}\right\} \Rightarrow \beta = \frac{160}{100} = 1/6$$

با توجه به گزینه‌ها نیازی به محاسبه α نمی‌باشد.

$$y = 1/6x + 3$$

مثال ۵: **قیمت (X) و تقاضای کالای (Y) در ۵ بازار مختلف جمع‌آوری شده است. شیب خط رگرسیون کدام است؟** (مدیریت و حسابداری ـ سراسری ۹۸)

x	۱۲	۱۰	۱۳	۱۶	۹
y	۲۰	۱۵	۱۱	۷	۲۲

۱) $-1/7$
۲) $-1/6$
۳) $-1/9$
۴) $-1/4$

☑ **پاسخ:** گزینه «۳» برای محاسبهٔ شیب خط رگرسیون $(\hat{\beta})$، ابتدا مقادیر S_{xx} و S_{xy} را حساب می‌کنیم. برای این کار سطر $(x_i-\bar{x})^2$ و سطر $(x_i-\bar{x})(y_i-\bar{y})$ را تشکیل می‌دهیم. جمع سطر $(x_i-\bar{x})^2$ برابر S_{xx} و جمع سطر $(x_i-\bar{x})(y_i-\bar{y})$ برابر S_{xy} است. تعداد نقاط $n=5$ می‌باشد.

					جمع
x	۱۲	۱۰	۱۳	۱۶	۹	۶۰
y	۲۰	۱۵	۱۱	۷	۲۲	۷۵
$(x_i-\bar{x})^2$	۰	۴	۱	۱۶	۹	$S_{xx}=0+4+1+16+9=30$
$(x_i-\bar{x})(y_i-\bar{y})$	۰	۰	−۴	−۳۲	−۲۱	$S_{xy}=0+0-4-32-21=-57$

$$\bar{x}=\frac{\sum x_i}{n}=\frac{60}{5}=12$$
$$\bar{y}=\frac{\sum y_i}{n}=\frac{75}{5}=15$$

در نتیجه حاصل $\hat{\beta}$ برابر است با:

$$\hat{\beta}=\frac{S_{xy}}{S_{xx}}=\frac{-57}{30}=-1/9$$

مثال ۶: اگر خط رگرسیون ساده برازش‌شده به صورت $\hat{y}=20+0/75x$ باشد. مقدار باقیمانده به ازای $x=100$ و $y=90$ کدام است؟

(علوم اقتصادی ـ دکتری ۱۴۰۲)

۱) ۱۵ ۲) ۵ ۳) صفر ۴) -۵

☑ **پاسخ:** گزینه «۴» مقدار باقیمانده به ازای $x=100$ به‌صورت زیر محاسبه می‌شود:

$$\hat{y}(100)=20+0/75(100)=95 \Rightarrow e(100)=y-\hat{y}(100)=90-95=-5$$

مثال ۷: معادله خط رگرسیون بین دو متغیر X و Y از جدول زیر کدام است؟ (مدیریت و حسابداری ـ سراسری ۹۵)

x	۱۱	۱۳	۱۴	۱۶	۱۷	۱۹
y	۱۰	۱۲	۱۵	۱۴	۱۱	۱۶

۱) $\hat{y}=0/8x+3$
۲) $\hat{y}=0/4x+7$
۳) $\hat{y}=0/6x+4$
۴) $\hat{y}=0/5x+5/5$

☑ **پاسخ:** گزینه «۴» معادله خط رگرسیون بین دو متغیر X و Y به‌صورت $Y=\hat{\beta}X+\hat{\alpha}$ است که در آن مقدار $\hat{\alpha}$ و $\hat{\beta}$ با دستورهای $\hat{\beta}=\frac{S_{xy}}{S_{xx}}$ و $\hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}$ حساب می‌شود. به منظور محاسبه $\hat{\alpha}$ و $\hat{\beta}$، ابتدا $\bar{x}$ (میانگین x_iها) و $\bar{y}$ (میانگین y_iها) را حساب کرده، سپس سطرهای $(x_i-\bar{x})^2$ و $(x_i-\bar{x})(y_i-\bar{y})$ را تشکیل می‌دهیم. جمع سطر $(x_i-\bar{x})^2$ برابر S_{xx} و جمع سطر $(x_i-\bar{x})(y_i-\bar{y})$ برابر S_{xy} است.

							جمع
x	۱۱	۱۳	۱۴	۱۶	۱۷	۱۹	۹۰
y	۱۰	۱۲	۱۵	۱۴	۱۱	۱۶	۷۸
$(x_i-\bar{x})^2$	۱۶	۴	۱	۱	۴	۱۶	$S_{xx}=42$
$(x_i-\bar{x})(y_i-\bar{y})$	۱۲	۲	−۲	۱	−۴	۱۲	$S_{xy}=21$

$$\bar{x}=\frac{\sum x_i}{n}=\frac{90}{6}=15$$
$$\bar{y}=\frac{\sum y_i}{n}=\frac{78}{6}=13$$

$$\hat{\beta}=\frac{S_{xy}}{S_{xx}}=\frac{21}{42}=0/5 \qquad \hat{\alpha}=\bar{y}-\beta\bar{x}=13-0/5\,(15)=5/5$$

با محاسبه $\hat{\alpha}$ و $\hat{\beta}$، معادله خط رگرسیون برابر می‌شود با:

$$\hat{y}=\hat{\beta}x+\hat{\alpha} \Rightarrow \hat{y}=0/5x+5/5$$

مثال ۸: با استفاده از خط رگرسیون بین دو صفت X و Y در جدول زیر، مقدار برآورد y برای $x=22$ کدام است؟ (حسابداری ـ دکتری ۹۵)

x	۱۲	۱۵	۱۶	۱۸	۱۹
y	۸	۷	۱۱	۱۳	۱۶

۱) ۱۸/۵
۲) ۱۷/۵
۳) ۱۷
۴) ۱۸