
بسم الله الرحمن الرحيم

سرشناسه	: غفاری، مختار، ۱۳۵۹-
عنوان و نام پدیدآور	: شرح جامع طرح آزمایش‌های کشاورزی / مولفان مختار غفاری – مهدی رنجبری.
مشخصات ظاهری	: [۲۰۶] ص.:مصور، نمودار.
شابک	:تومان ۹۴۰.۰۰۰ ۹۷۸-۶۰۰-۱۱۵-۷۰۹-۷
شمارگان	: ۵۰ ن
وضعیت فهرست نویسی	: فیپا
یادداشت	: بالای عنوان: مجموعه آزمون‌های ده سال کنکور ارشد.
موضوع	: کشاورزی تجربی – راهنمای آموزشی (عالی)
موضوع	: کشاورزی تجربی -- آزمون‌ها و تمرین‌ها (عالی)
موضوع	: دانشگاه‌ها و مدارس عالی -- ایران -- آزمون‌ها
موضوع	: آزمون دوره‌های تحصیلات تکمیلی -- ایران
شناسه افزوده	: رنجبری، مهدی
رده‌بندی کنگره	: SB۴۰/آ۲غ۷
رده بندی دیویی	: ۶۳۱/۰۷۶
شماره کتابشناسی ملی	: ۲۱۳۳۵۶۴

انتشارات ارشد

نام کتاب : شرح جامع طرح آزمایش‌های کشاورزی

نام مؤلفان : مختار غفاری – مهدی رنجبری

حروفچینی و صفحه‌آرایی : واحد تایپ و صفحه‌آرایی انتشارات ارشد

نوبت چاپ :دوم، ۱۴۰۵

شماره استاندارد بین‌المللی کتاب (شابک): ۷ - ۷۰۹-۱۱۵-۶۰۰-۹۷۸

International Standard Book Number(ISBN): 978-600-115-709-7

مرکز پخش: میدان انقلاب – خیابان وحیدنظری- بین دوازده فروردین و منیری جاوید – پلاک ۱۳۰ – طبقه اول – واحد ۲ ۶۶۹۷۹۲۲۷-۶۶۹۷۹۲۳۴

www.arshadbook.com

هرگونه الگوبرداری از سبک و سیاق و یا تکثیر کل یا قسمتی از کتاب‌های انتشارات ارشد ممنوع بوده و متخلف براساس قانون دفاع از حقوق مولفین و مصنفین تحت پیگرد شدید قانونی قرار خواهد گرفت. واحد حقوقی انتشارات ارشد: ۶۶۹۷۹۲۲۷-۶۶۹۷۹۲۳۴

پیشگفتار ناشر

خداوند بزرگ را شاکریم که فرصتی عنایت فرمود تا بتوانیم قدمی در جهت گسترش علم و فرهنگ کشور عزیزمان برداریم و مسیری هموار را در اختیار مخاطبان و داوطلبان علاقمند به ادامه تحصیل قرار دهیم. از آنجائی که درحوزه کتاب‌های درسی و کمک درسی کارشناسی ارشد کتاب‌های زیادی به رشته تحریر در آمده است با این حال وجود یک مجموعه جامع، دقیق، کامل و دارای رفرنس(منابع) مستند احساس می‌شود.

همیشه دغدغه داشتیم که ای کاش می‌شد تمامی منابع پراکنده و گاهاً دست نایافتنی کنکور در یک مجموعه به گونه‌ای جذاب و شیوا آماده می‌شد و همچنین سوالات دوره‌های قبل کنکور نیز به گونه‌ای تشریحی پاسخ داده می شدند، که خدشه‌ای بر آنها وارد نشود و خیال داوطلبان عزیز آنقدر آسوده باشد که اطلاعات مندرج در پاسخنامه‌ها را برای بهره‌گیری در آزمون خود بکار گیرند. این کار با یاری خداوند بزرگ و همکاری نخبه‌ترین دانشجویان و اساتید کشور در انتشارات ارشد هماهنگ گردید و در تمامی مقاطع تحصیلی از کاردانی به کارشناسی، کارشناسی ارشد و دکتری در حوزهٔ آزمون‌های سراسری – آزاد و پیام نور بسته‌های آموزشی شامل کتاب‌های بانک سوالات و کتاب‌های شرح جامع درس آماده گردید. در تهیه مجموعه‌های فوق‌الذکر از فارغ‌التحصیلان کارشناسی ارشد و دکتری که حائز رتبه‌های تک رقمی کنکور می‌باشند، استفاده کرده‌ایم، تا شما عزیزان بتوانید در راستای افزایش توان علمی خود در جهت نیل به مقصود خویش نایل آیید. بدون تردید می‌توان گفت که برای اولین بار در نظام آموزشی کشور مجموعه‌ای به این سبک عرضه می‌شود که در کیفیت و کمیت بی‌نظیر بوده و تمامی نیازهای دانشجویان و داوطلبان عزیز را پوشش می‌دهد. امید است که استفاده از این مجموعه عزیزان را در جهت موفقیت در آزمون یاری رسانده باشد. در پایان لازم به ذکر است که بی نقصی محض از آن خداوند متعال می‌باشد، بنا براین از تمامی خوانندگان و داوطلبان عزیز خواهشمندیم نظرات سازنده خود را در مورد این کتاب به سایت انتشارات ارشد اعلام نمایند تا ما را در جهت هر چه بهتر آماده کردن این محصولات یاری رسانند.

با تشکر

ناشـــــر

www.arshadbook. ir

فهرست مطالب

عنوان **صفحه**

فصل اول : چکیده‌ای از مفاهیم آمار

فصل دوم: مقدمه طرح آزمایش‌های کشاورزی

فصل سوم: طرح کامل تصادفی(COMPLETELY RAMDOMIZED DESIGNS- CRD)

تستهای آخر فصل ۹۳
پاسخ تست‌های آخر فصل ۹۴

فصل چهارم: طرح بلوک‌های کامل تصادفی (RONDOMIZED COMPLETE BLOCK DESIGN - RCBD)

تستهای تألیفی آخر فصل..... ۱۱۰
پاسخ تستهای تألیفی آخر فصل..... ۱۱۱

فصل پنجم: طرح مربع لاتین (LATIN SQUARE DESIGN)

تستهای تألیفی آخر فصل..... ۱۲۴
پاسخ تستهای تألیفی آخر فصل..... ۱۲۵

فصل ششم: آزمایش‌های فاکتوریل (FACTORIAD EXPERIMENTS)

تستهای آخر فصل ۱۴۸
پاسخ تستهای آخر فصل ۱۴۹

فصل هفتم: اختلاط (CONFOUNDING)

تستهای آخر فصل ۱۶۴
پاسخ تستهای آخر فصل ۱۶۵

فصل هشتم: کرت‌های خردشده (اسپلیت پلات): SPLIT-POLT DESIGN

تستهای آخر فصل ۱۷۹
پاسخ تستهای آخر فصل ۱۸۱

فصل اول

چکیده‌ای از مفاهیم آمار

عناوین اصلی:

- تعریف آمار
- شاخص‌های آماری
- جمعیت و نمونه
- داده‌ها (مشاهدات)
- متغیر
- انواع مقیاس اندازه‌گیری
- توزیع فراوانی
- احتمال
- قوانین احتمال
- توزیع احتمال
- ویژگیهای متغیر تصادفی
- مقایسه‌ی میانگین دو نمونه برآورد حدود اطمینان میانگین‌ها
- رابطه بین متغیرها
- رگرسیون
- همبستگی
- تبدیل داده‌ها

فصل اول

چکیده‌ای از مفاهیم آمار

تعریف آمار

اصول آمار و روش‌های آن از پایه‌های ضروری تحقیق علمی به شمار می‌روند. برای آمار تعریف‌های مختلفی ارائه شده است ولی تعریفی که از هر جهت و از نظر کاربردی جامع باشد، چنین است: آمار علمی است که در مورد جمع‌آوری، سازمان‌دادن و تجزیه و تحلیل داده‌ها و تفسیر و تعمیم نتایج بحث می‌کند.

در بسیاری از رشته‌های کاربردی مثل کشاورزی، علوم انسانی و ... تجزیه تحلیل‌های آماری نقش بسیار مهمی در استنتاجهای علمی دارند، بطوریکه بدون استفاده از آنها، امکان نتیجه‌گیری صحیح از داده‌های جمع‌آوری شده، ناممکن یا بسیار دشوار است.

لذا در این راستا آمار به عنوان ابزاری برای دو هدف عمده زیر در علوم مختلف بکارگرفته می‌شود.

الف) بررسی فرضیه‌های مختلف

ب) تخیمن پارامترهای یک یا چند جامعه

شاخص‌های آماری

شاخص‌های آماری عبارتند از مقادیر، کمیت‌ها و معیارهائی که برای نشان‌دادن نتیجه یک بررسی آماری بکار برده می‌شوند. به شاخص‌های آماری در جامعه پارامتر گفته می‌شود.

شاخص‌های تمایل مرکزی: در بسیاری از تحقیقات لازم می‌شود که کمیت یا شاخصی را به عنوان معرف و یا نماینده کلی داده‌ها و مشاهدات ارائه نمود و از آن در بررسی نتایج تحقیق استفاده کرد. مثلاً فرض کنید مقایسه دو نژاد گاو شیرده از نظر میزان شیردهی موردنظر باشد. برای این منظور آزمایشی در شرایط یکسان روی دو نژاد انجام می‌شود و طی آن مقدار شیر چند گاو از هر نژاد اندازه‌گیری می‌گردد.. و برای مقایسه دقیق‌تر شیردهی دو نژاد، بایستی برای هر نژاد مقدار یا شاخصی را به عنوان بهترین نماینده یا معرف شیردهی آن ارائه داد. به عبارت دیگر بایستی عددی را بدست آورد که اگر به جای میزان شیردهی تک تک گاوها به حساب آوریم، مقدار تفاوتی که به این ترتیب پدید می‌آید، حداقل باشد. بدیهی است که این عدد یا شاخص نه در حد بالا و نه در حد پائین بلکه در حد متوسط دامنه مشاهدات (میزان شیردهی) واقع شده است در آمار، مهمترین شاخص‌های مرکزی شامل نما یا مد، میانه، میانگین‌های حسابی، هندسی و همساز می‌باشند.

نما(مُد): مشاهده‌ای است که دارای بیشترین فراوانی باشد و معمولاً در دامنه میانی قرار دارد.

میانه: سلسله مرتبّ شده مشاهدات، میانه مشاهده‌ای است که این رشته را از نظر تعداد به دو قسمت مساوی تقسیم می‌کند. بطوریکه نیمی از اعداد کمتر و نیمی دیگر بیشتر از آن می‌باشند.

میانگین حسابی، عبارت است از حاصل جمع داده‌ها تقسیم بر تعداد آنها است:

$$\mu = \frac{\Sigma x_i}{N}$$

μ: میانگین

Σx_i: مجموع کل داده‌ها و N تعداد کل مشاهدات

میانگین هندسی، چنانچه مشاهدات حاصل از یک بررسی آماری برحسب درصد، نسبت و امثال اینها باشد، معمولاً میانگی دیگری بنام میانگین هندسی محاسبه می‌شود. میانگین هندسی N داده آزمایشی برابر است با ریشه N ام حاصل ضرب این داده‌هاو آن را با m_g نمایش می‌دهند.

$$m_g = \sqrt[n]{X_1.X_2.....X_i.....X_n}$$

میانگین همساز: چنانچه مقادیر Xها را معکوس نمائیم و میانگین آنها را محاسبه کرده و سپس این میانگین را معکوس کنیم، میانگین همساز و یا میانگین هارمونیک بدست می‌آید. کاربرد این میانگین زمانی است که شدت و ارزش داده‌ها متفاوت باشد.

$$m_h = \frac{N}{\sum \frac{1}{x_i}}$$

شاخص‌های پراکندگی

شاخص‌های مرکزی، مقدار متوسط اطلاعات بسیار مهمی را در خصوص داده‌های تحقیقاتی بیان می‌کنند، ولی تصویر کاملی از ماهیت داده‌ها را در اختیار قرار نمی‌دهند. فرض کنید تعداد لارو یک حشره که آفت ذرت می‌باشد، روی ۴ بوته در ۳ مزرعه مورد شمارش قرار گرفته است و نتایج زیر حاصل شده‌اند.

مزرعه اول: ۲۰، ۲۲، ۲۸، ۳۰

مزرعه دوم: ۸، ۱۰، ۴۰، ۴۲

مزرعه سوم: ۲، ۵، ۴۵، ۴۸

میانگین تعداد لاروها در هر سه مزرعه مساوی ۲۵ می‌باشد، اما گیاهان این سه مزرعه به هیچ وجه بطور یکنواخت مورد حمله این آفت قرار نگرفته‌اند. در مزرعه اول همه گیاهان مقاومت مشابه و در حد متوسط دارند، در مزرعه دوم گیاهانی با مقاومت کم و همچنین با مقاومت خوب مشاهده می‌شود. در مزرعه سوم گیاهان بسیار حساس و بسیار مقاوم وجود دارد. فرقی که بین گیاهان این سه مزرعه وجود دارد، با چگونگی توزیع داده‌ها در اطراف میانگین مشخص می‌شود این خصوصیت را در آمار پراکندگی یا تنوّع می‌گویند.

مهمترین شاخص‌های پراکندگی عبارتند از: دامنه کلی تغییرات، انحراف متوسط و انحراف معیار.

دامنه کلی تغییرات: اختلاف بین کمترین و بیشترین مشاهده، که برای مزرعه اول در این مثال (۱۰=۲۰-۳۰) و برای مزرعه دوم ۳۴ و برای مزرعه سوم ۴۶ می‌باشد.

انحراف متوسط

$$AD = \frac{\sum |X_i - \mu|}{N}$$

انحراف معیار

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

جمعیت و نمونه

به مجموعهٔ بزرگی از افراد که دست کم دارای یک صفت مشترک باشند جمعیت گفته می‌شود. به طور مثال به افرادی که در یک کشور زندگی می‌کنند جمعیت آن کشور و به مجموعه ژنوتیپ های متعلق به یک گونهٔ گیاهی یا جانوری، جمعیت آن گونه گیاهی یا جانوری گفته می‌شود. اگر شمار افراد یک جمعیت مشخص باشد آن جمعیت را محدود می‌نامند، مانند جمعیت درختان یک گونه که توسط آفت بخصوصی مورد حمله قرار می‌گیرد. در صورتی که تعداد افراد بیشمار باشد آن جمعیت را «نامحدود» می‌گویند. بطور مثال ژنوتیپ‌های موجود در یک تودهٔ بومی ذرت جمعیتی نامحدود را تشکیل می‌دهند. در مواردی که یک جمعیت محدود از مجموعه بسیار بزرگی از افراد تشکیل می‌شود آن را نیز به عنوان جمعیت نامحدود در نظر می‌گیرند. جمعیت های مورد استفاده توسط پژوهشگران غالباً بسیار بزرگ است. نظر به اینکه اندازه گیری کلیه افراد یک جمعیت امکان پذیر نیست، به منظور بررسی خصوصیات جمعیت تعدادی از افراد را به عنوان نمونه، انتخاب و ارزیابی می‌کنند. به مجموعه ای از افرادی که از یک جمعیت انتخاب می‌شوند نمونه می‌گویند. برای رسیدن به یک نتیجهٔ معتبر، نمونه باید نمایانگر جمعیت مورد نظر باشد. معمولاً یکی از روش های مناسب انتخاب نمونه، انتخاب تصادفی است تا کلیهٔ افراد جمعیت، شانس مساوی در تشکیل آن داشته باشند. علاوه بر این، هر چه تعداد افراد در نمونه بیشتر باشد برآورد ویژگی های جمعیت دقیق تر خواهد بود. با وجود این، عموماً از نظر عملی محدودیت هایی در افزایش اندازهٔ نمونه (تعداد افراد در نمونه) وجود دارد. در آمار اگر اندازه نمونه بیشتر از ۳۰ باشد آن را «نمونه ی بزرگ» و اگر کمتر از ۳۰ باشد «نمونه کوچک» می‌گویند.

معیار جمعیت (پارامتر) و معیار نمونه

کمیتی که ویژگی های جمعیت را نشان می‌دهد معیار جمعیت یا پارامتر نامیده می‌شود. معیار نمونه نیز ویژگی های نمونه را نمایان می‌سازد. پارامتر جمعیت کمیتی ثابت ولی ناشناخته است و آن را از طریق نمونه برداری و محاسبه معیار نمونه برآورد می‌کنند. با وجود این، مقدار عددی معیار نمونه قابل نوسان است، چرا که نمونه های برداشت شده از یک جمعیت الزاماً مشابه نیستند. معیار نمونه را به طور معمول با حروف کوچک لاتین و پارامتر جمعیت را با حروف کوچک یونانی نشان می‌دهند. به عنوان مثال در یک جمعیت نرمال برای میانگین نمونه علامت $\bar{x}$ و برای میانگین جمعیت حرف μ را به کار می‌برند.

داده‌ها (مشاهدات)

داده‌ها (مشاهدات) اعداد و ارقامی هستند که از اندازه گیری صفت یا متغیر مورد مطالعه در افراد یا از شانس افراد برای داشتن یک صفت حاصل می‌شوند.

متغیر

متغیر عبارت از عامل یا صفتی است که ارزش آن از فردی به فرد دیگر تغییر می‌کند. اگر مقادیر به طور تصادفی از جمعیت برداشت شوند، آن را متغیر تصادفی می‌نامند، در حالی که اگر یک متغیر به طور غیر تصادفی و توسط پژوهشگر تعیین شود، به آن متغیر ثابت (تثبیت شده) می‌گویند. هنگامی که ارزش متغیر با اعداد و ارقام قابل بیان نباشد و افراد درون جمعیت به چند نوع مجزا تقسیم شوند، در این حالت به آن، متغیر کیفی می‌گویند، همانند رنگ گل، داشتن یا نداشتن ریشک، رنگ دانه (سفید یا سیاه). برعکس در متغیر کمیّ ارزش افراد توسط اعداد بدست آمده مشخص می‌شود. از این جمله می‌توان ارتفاع گیاه، عملکرد دانه، عملکرد علوفه، تعداد دانه در سنبله، تعداد نوزادان در یک زایمان در حیوانات و درصد پروتئین را نام برد. از طرف دیگر متغیرها به دو گروه پیوسته و نا پیوسته نیز تقسیم می‌شوند. به طور معمول متغیر پیوسته مورد اندازه گیری قرار گرفته و افراد می‌توانند هر مقدار ممکن در دامنهٔ تغییرات (اختلاف دو حد بالا و پایین متغیر) موجود را به خود اختصاص دهند. در این نوع متغیرها، اختلاف ارزش های نزدیک بهم کوچک و غالباً قابل تشخیص نیست. به عبارت دیگر ارزش افراد دریک مقیاس پیوسته نوسان می‌کند. به طور مثال اگر گیاهان یک توده بومی گندم از نظر ارتفاع بوته گروه بندی شوند، نمی‌توان آنها را به دو کلاس «پابلند» و «پاکوتاه» تقسیم کرد، بلکه ممکن است انواع افراد با ارتفاع متفاوت در این جمعیت وجود داشته باشند. صفاتی مانند عملکرد دانه، عملکرد علوفه، ارتفاع گیاه، درصد روغن در پنبه دانه، درصد چربی شیر، میزان تولید شیر در دام ها در زمرهٔ متغیرهای پیوسته به حساب می‌آیند. در متغیرهای ناپیوسته، ارزش افراد از طریق شمارش بدست می‌آید و قابل اندازه گیری نیست. نظر به اینکه مقادیر حاصل جدا از هم بوده و در دسته های مجزا گروه بندی می‌شوند، تغییرات در این نوع متغیرها در مقیاس ناپیوسته به چشم می‌خورد از مثال های بارز در این مورد می‌توان تعداد گلبرگ های یک گل، تعداد دانه در سنبلهٔ گندم، تعداد ردیف های دانه در بلال ذرت، تعداد لاروهای موجود در یک گیاه و تعداد نوزادان در یک زایمان در حیوانات را برشمرد. به طوری که ملاحظه می‌شود، متغیرهای کمی می‌توانند به دو صورت پیوسته (مانند عملکرد دانه و ارتفاع گیاه) و ناپیوسته (مانند تعداد دانه در سنبله‌ی غلات، تعداد لارو در یک گیاه) تظاهر یابند. متغیرها را می‌توان به دو گروه وابسته و مستقل نیز متمایز کرد. مقادیر متغیر مستقل معمولاً توسط آزمایش کننده تعیین می‌شود. بنابراین، متغیر مستقل غالباً متغیر ثابت نیز است. متغیرهایی مانند فاصلهٔ ردیف عمق کاشت، مقدار کود ازت و زمان کاشت از نوع متغیرهای مستقل هستند. از طرف دیگر مقادیر متغیر وابسته بستگی به سطح انتخابی متغیر مستقل دارد. نظر به اینکه مقدار متغیر وابسته قبل از آزمایش قابل تشخیص نیست، متغیر وابسته همواره یک متغیر تصادفی به حساب می‌آیند. از متغیرهای وابسته می‌توان عملکرد دانه، درصد پروتئین، میزان تولید شیر و میزان اسید اسکوربیک در گوجه فرنگی را نام برد.

انواع مقیاس اندازه‌گیری

مقیاس‌های اندازه‌گیری عبارتند از اسمی، رتبه‌ای، فاصله‌ای و نسبتی. در مقیاس اسمی اعداد را به عنوان اسم به کار می‌برند. این اعداد مقادیر کمیّ واقعی نیستند و فقط اختلافات بین افراد را نشان می‌دهند. مثال کلاسیک در این مورد، شماره بازیکنان فوتبال است که هیچ مفهومی به غیر از متمایزکردن بازیکنان در زمین بازی ندارد. از مقیاس اسمی غالباً برای گروه

بندی افراد یا اشیاء استفاده می‌کنند. داده‌های کیفی بر اساس مقیاس اسمی اندازه گیری می‌شوند، در حالی که داده های کمیّ بر مبنای سه مقیاس دیگر قابل اندازه‌گیری هستند. مقیاس رتبه‌ای افزون بر اینکه مانند مقیاس اسمی، تفاوت ها را نمایان می‌سازد، افراد و اشیاء را در طول یک مقیاس کمیّ مرتب می‌کند. از این نوع مقیاس می‌توان رتبهٔ ورزشکاران در یک مسابقه و یا رتبهٔ دانشجویان در یک دورهٔ تحصیلی را نام برد. در مقیاس فاصله بین ۱۰ و ۲۰ درجه سانتی گراد مانند فاصله‌ی بین ۲۰ و ۳۰ درجه سانتی گراد است، در صورتی که در مقیاس رتبه ای ممکن است فاصلهٔ دانشجوی رتبه‌ی اول از دوم ۰/۵ نمره ولی فاصله‌ی دانشجوی رتبه‌ی دوم از سوم ۱/۵ نمره باشد. در مقیاس فاصله‌ای نقطهٔ صفر قراردادی است، مانند صفر درجهٔ سانتیگراد که نشانگر نقطهٔ انجماد آب در ارتفاعی برابر سطح دریا است. همچنین در این مقیاس نمی‌توان صحبتی از نسبت به میان آورد. به طور مثال نمی‌توان گفت که دمای ۳۰ درجهٔ سانتی گراد از گرمایی به میزان دو برابر دمای ۱۵ درجهٔ سانتی گراد برخوردار است. زیرا اگر سانتی گراد به فارنهایت تبدیل شود، نسبت دو برابر به دست نمی‌آید. مقیاس نسبتی کلیهٔ ویژگی های سه مقیاس قبلی را دارا است. علاوه بر این، نقطه ی صفر آن واقعی بوده و امکان گزارش داده ها به صورت نسبت وجود دارد. صفانی مانند عملکرد، ارتفاع بوته، وزن دام و درصد پروتئین دارای مقیاس نسبتی هستند. به عنوان مثال ارتفاع یک بوته ۲۰۰ سانتی متری دو برابر بوته‌ی ۱۰۰ سانتی متری است و این نسبت با تبدیل داده ها به سیستم غیر متریک تغییر نمی‌کند. اطلاع از نوع مقیاس اندازه گیری نقش مهمی در انتخاب روش های آماری مناسب برای تجزیه‌ی داده ها ایفا می‌کند. داده‌هایی که از طریق مقیاس رتبه‌ای به دست می‌آیند، فاقد توزیع نرمال هستند. بنابراین، تجزیه‌ی آنها باید توسط روش های آماری غیر پارامتری انجام شود. از طرف دیگر داده های برخوردار از مقیاس فاصله ای و نسبتی می‌توانند دارای توزیع نرمال باشند و برای تجزیه‌ی این داده ها روش های پارامتری مبتنی بر توزیع نرمال، در صورت صادق بودن فرض های مربوطه، به کار می‌روند.

توزیع فراوانی

یکی از مراحل مهم در تجزیه داده‌ها، پس از اندازه‌گیری نمونه از جامعه، تنظیم یا سازماندهی مشاهدات بصورت کمیت‌های آماری و نمودارهای با معنی و قابل قبول است. کمیت‌های قابل استفاده در آمار عبارتند از: فراوانی مطلق، نسبی و تراکمی.

فراوانی نسبی

فراوانی نسبی از نسبت تعداد مشاهدات موردنظر بر تعداد کل افراد در نمونه بدست می‌آید و بصورت درصد نشان داده می شود. مقادیر این فراوانی‌ها از صفر تا یک متغیر بوده و مجموع آنها برابر یک است.

نمودار ستونی

برای توصیف بهتر داده‌ها، به جای جدول فراوانی نسبی، از نمودار فراوانی نسبی استفاده می‌شود. که آن را نمودار میله‌ای ستونی می‌نامند و برای متغیرهای کیفی و ناپیوسته بکار می‌رود. در این نمودارها ارتفاع ستون‌ها فراوانی هر دسته یا گروه را نشان می‌دهد.

فراوانی نسبی تراکمی

فراوانی نسبی تراکمی یک دسته، از مجموع فراوانی نسبی آن دسته و دسته‌های دارای ارزش‌های پایین‌تر حاصل می‌شود.

احتمال

کلمه احتمال برای نشان دادن شانس وقوع برخی از پیشامدها بکار برده می‌شود. بطور کلی هر فرآیند جمع‌آوری داده، یک آزمایش محسوب می‌شود. احتمال در زبان ریاضی بصورت زیر نشان داده می‌شود.

$$p = \frac{m}{n}$$

در این رابطه **m** تعداد آزمایش‌های مورد نظر و یا حالات موفقیت است و **n** تعداد کل آزمایش‌ها و یا تعداد کل حالات است در رابطه با مبحث احتمال ابتدا دو لغت «آزمایش» و «پیشامد» تعریف می‌شوند. آزمایش برای نشان دادن مجموعه ای از وضعیت هایی که یک پیشامد را به وجود می‌آورند بکار می‌رود. آزمایش ممکن است پرتاب یک سکه یا تولید نتایج در یک خانواده را شامل شود و پیشامد آن می‌تواند خط در حالت اول و جنس نر در حالت دوم باشد. چون پیشامد یک آزمایش بطور تصادفی ظاهر شده و قابل پیش بینی نیست، آن را متغیر تصادفی می‌نامند. برای مثال «تعداد خط» در دو پرتاب یک سکه سه مقدار ۰ و ۱ و ۲ را در بر می‌گیرد و به عنوان متغیر تصادفی ناپیوسته به حساب می‌آید. اگر نمونه ای مرکب از ۱۰۰۰ بذر برداشته و تعداد بذور علف های هرز را بشماریم، این تعداد که از صفر تا ۱۰۰۰ قابل تغییرات بعنوان یک متغیر تصادفی ناپیوسته در نظر گرفته می‌شود.چنین پیشامدها را که وقوع یکی مانع وقوع دیگری می‌شود پیشامدهای ناسازگار می‌گویند. سایر متغیرهای تصادفی در کلاس های ناسازگار قرار نمی‌گیرند، بلکه در طول یک مقیاس پیوسته توزیع می‌یابند. امّا مبحث آزمایش و پیشامد باز هم قابل کاربرد است. طول سنبله در یک جمعیت بومی‌گندم را می‌توان به عنوان یک پیشامد مطرح نمود. اگر داده های حاصل از این اندازه گیری را در طول یک مقیاس گروه بندی کنیم، فراوانی نسبی اینگونه متغیر پیوسته همانند متغیرهای ناپیوسته قابل محاسبه خواهد بود. فراوانی نسبی یک پیشامد یا متغیر تصادفی را در جمعیت احتمال آن پیشامد می‌گویند. مقدار احتمال از صفر تا یک متغیر است. برخی از پیشامدها دارای احتمال شناخته شده ای هستند. مثلاً احتمال خط در پرتاب یک سکه $\frac{۱}{۲}$ ، احتمال دختر یا پسر بودن صرفنظر از تأثیر عوامل محیطی $\frac{۱}{۲}$، احتمال در سایر پیشامدها قابل محاسبه نبوده و یک عدد تئوریک است و فقط از طریق فراوانی نسبی دو نمونه ها برآورد می‌شود. اگر صفت مورد مطالعه یا مورد اندازه گیری در یک جمعیت دارای تغییرات پیوسته باشد، برآورد احتمال یک فرد یا ارزش بخصوصی امکان پذیر نیست و اساساً مساوی صفر می‌باشد. امّا در این حالت می‌توان احتمال ارزش های بیش از یک مقدار ثابت،K را برآورد کرد. برای مثال اگر احتمال نوزادان سنگین تر از ۴ کیلو گرم، ۰/۱۵ برآورد شده باشد آن را به ترتیب زیر نشان می‌دهند.

$$P(\text{وزن} > ۴) = ۰/۱۵$$

قوانین احتمال

۱. احتمال وقوع هر پیشامد ساده بین ۰ و ۱ است. $۰ \le P(Ei) \le ۱$

۲. مجموع احتمالات حاصل از تمام پیشامدهای ساده برابر با ۱ است. $\Sigma_i P(E_i) = ۱$

قوانین شمارش:

به یاد دارید که احتمال عبارتست از:

$$P = \frac{\text{تعداد آزمایش‌های مطلوب}}{\text{تعداد کل آزمایش‌ها}} = \frac{m}{n}$$

قانون ضرب

به تعداد K مجموعه با تعداد عضوهای $n_۱, n_۲$ و ... و n_k را در نظر بگیرید. اگر به طور تصادفی ازهر مجموعه یک عضو انتخاب شود تعداد کل نتایج مختلف برابراست با: $n_۱, n_۲, n_۳, ..., n_k$.

مثال

گیاهان موجود در سه کرت به صورت زیر علامت‌گذاری شده‌اند.

کرت ۱: ۱و۲و ۳

کرت ۲: B,A **و** C

کرت ۳: y,x

تعداد گیاهان موجود در هر کرت عبارتند از $n_۳ = ۲$ $n_۲ = ۳$ $n_۱ = ۳$

اگر از هر کرت یک گیاه انتخاب شود ترکیب‌های سه‌تایی ممکن عبارتند از:

$$۱Ax, ۱Ay, ۱Bx, ۱By, ۱Cx, ۱Cy$$

$$۲Ax, ۲Ay, ۲Bx, ۲By, ۲Cx, ۲Cy$$
$$۳Ax, ۳Ay, ۳Bx, ۳By, ۳Cx, ۳Cy$$

تعداد ترکیب‌های سه‌تایی ممکن برابر است با: $۳\times۳\times۳\times۲=۱۸$.

جایگشت: (ترتیب)

در یک مجموعه n عضوی، تعداد روش‌هایی که به وسیله آن می‌توان اعضاء را مرتب نمود (به صورت‌های مختلف در کنار هم قرار داد)را تعداد جایگشت‌های n عضو گویند. و به صورت زیر نمایش می‌دهند:

$$p_n = n!$$

نماد $n!$ (n فاکتوریل)، نشان‌دهنده حاصلضرب متوالی اعداد طبیعی ۱ تا n می‌باشد:

$$n! = (۱)(۲)(۳)...(n)$$

$$۰! = ۱ \quad *$$

مثال

به چند طریق می‌توان سه حیوان x و y و z را مرتب نمود یا به عبارتی چند جایگشت از سه حیوان x و y و z وجود دارد.

تعداد جایگشت‌های سه شیء متمایز عبارتند از: $p(۳) = ۳! = (۱)(۲)(۳) = ۶$

ترکیب‌های سه‌تایی ممکن: $xyz, xzy, yxz, yzx, zxy, zyx$

به طور کلی، می‌توان تعداد جایگشت‌های kتائی n عضو را به صورت زیر بدست آورد:

$$p_{n,k} = \frac{n!}{(n-k)!}$$

مثال

به چند روش می‌توان جفت‌های مختلفی از ۳ حیوان x و y و z را با در نظر گرفتن ترتیب جفت‌ها مرتب نمود (xz با zx تفاوت دارد.)

این ۶ جفت عبارتند از:

$$p_{n,k} = \frac{۳!}{(۳-۲)!} = ۶ \qquad xy, xz, yx, yz, zx, zy$$

ترکیب

در ترکیب مجموعه n عضوی، تعداد راه‌های مختلف برای انتخاب k عضو از n عضو، بدون در نظر گرفتن ترتیب (xz و zx متفاوت نیستند، بعبارت دیگر جایگشت وجود ندارد) برابر است با:

$$\binom{n}{k}=\frac{ni}{k!(n-k)!}=\frac{n(n-۱)...(n-k+۱)}{k!}$$

مثال

به چند طریق می‌توان جفت‌های مختلفی از ۳ حیوان z,y,x بدون در نظر گرفتن ترتیب آرایش داد؟

این سه جفت عبارتند از:

$$\binom{n}{k}=\binom{۳}{۲}=\frac{۳!}{۲!(۳-۲)!}=۳ \qquad xy, xz, yz$$

قانون تفکیک (تقسیم‌بندی)

در یک مجموعه n عضوی، تعداد راه‌های مختلف اختصاصی n عضو به j گروه با اندازه‌های $n_۱,n_۲,n_۳....,n_j$ برابر است با: که در رابطه بالا: $n=n_۱+n_۲+...+n_j$

$$\frac{n!}{n_۱!\,n_۲!...n_j!}$$

مثال

به چند طریق می‌توان مجموعه‌ای از ۵ حیوان را در سه جایگاه $(j=۳)$ جای داد، به طوریکه در جایگاه اول ۲ حیوان $n_۱=۲$ در جایگاه دوّم ۲ حیوان و در جایگاه سوّم ۱ حیوان قرار گیرد.

$$\frac{۵!}{۲!\,۲!\,۱!}=۳۰$$

باید توجه داشت که قانون قبلی در مورد ترکیب، حالتی خاص از این قانون است که در آن یک مجموعهٔ به دو گروه با اندازه‌های k و n-k تفکیک می‌شود.

نمودار درختی

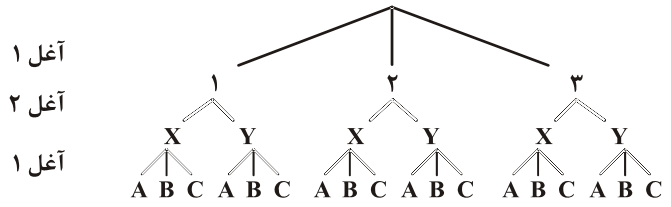
در نمودار درختی تعداد کل نتایج ممکن حاصل از یک آزمایش نشان داده می‌شود. این نمودار برای نشان دادن و بررسی احتمال یک پیشامد خاص مورد استفاده قرار می‌گیرد. به عنوان مثال شکل زیر یک نمودار درختی از ترکیب‌های سه‌تایی را نشان می‌دهد که از هر آغل یک حیوان انتخاب شده است.

آغل ۱: ۱،۲،۳

آغل ۲: x و y

آغل ۳: A، B، C

تعداد کل ترکیب‌های سه‌تایی برابر است با $۳\times۳\times۲=۱۸$ نمودار درختی آن به صورت زیر رسم می‌شود:



اولین ترکیب. سه‌تائی، حیوان ۱ از آغل ۱، حیوان x از آغل ۲، و حیوان A از آغل ۳ را شامل می‌شود. چنانچه احتمال مربوط به هر پیشامد در نمودار درختی گنجانده شود به آن نمودار احتمال درختی گفته می‌شود.

نکته: اگر پیشامد یک آزمایش به وقوع یا عدم وقوع پیشامد دیگر وابسته نباشد آن دو پیشامد را مستقل گویند ولی در برخی از حالات پیشامدها مستقل نیستند و احتمال وقوع یک پیشامد به وقوع پیشامد دیگر وابسته است.

پیشامدهای مرکب

پیشامد مرکب: پیشامدی است که ترکیبی از دو یا چند پیشامد دیگر باشد. دو پیشامد A و B را در نظر بگیرید.
پیشامد مرکبی که شامل وقوع هر دو پیشامد A و B باشد اشتراک پیشامدها نامیده می‌شود و با نماد $A\cap B$ نشان داده می‌شود.
پیشامد مرکبی که شامل وقوع A یا B باشد اجتماع پیشامدها نامیده می‌شود و با نماد $A\cup B$ نشان داده می‌شود.
احتمال اشتراک دو پیشامد A و B را با نماد $P(A\cap B)$ و احتمال اجتماع دو پیشامد A و B را با نماد $P(A\cup B)$ نشان می‌دهند. همچنین:

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

متمم پیشامد A، پیشامدی است که در آن A واقع نمی‌شود و با نماد A^c نشان داده می‌شود. احتمال متمّم A برابر است با:

$$P(A^c)=۱-p(A)$$

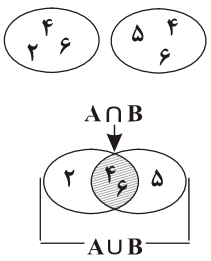
مثال

اگر در آزمایش پرتاب تاس، پیشامد A، مشاهدهٔ یک عدد زوج و پیشامد B مشاهده عدد بزرگتر از ۳ باشد.
پیشامد A مجموعه‌ای است از: $\{۲,۴,۶\}$
پیشامد B مجموعه‌ای است از $\{۴,۵,۶\}$
اشتراک A و B پیشامدی است که نتیجه آن، یک عدد زوج و یک عدد بزرگتر از ۳، باشد (یعنی مشاهدهٔ عددی که هم زوج و هم بزرگتر از ۳ باشد) بنابراین: $(A\cap B)=\{۴,۶\}$
احتمال اشتراک دو پیشامد A و B برابر است با:

$$P(A\cap B)=P(۴)+P(۶)=\frac{۲}{۶}$$

اجتماع پیشامدهای A و B پیشامدی است که نتیجه آن یک عدد زوج یا یک عدد بزرگتر از ۳ باشد
بنابراین $(A\cup B)=\{۲,۴,۵,۶\}$ **احتمال اجتماع دو پیشامد A و B برابر است با:**

$$P(A\cup B)=P(۲)+P(۴)+P(۵)+P(۶)=\frac{۴}{۶}$$



احتمال شرطی: عبارتست از، احتمال وقوع یک پیشامد، به شرط آنکه پیشامد دوم واقع شده باشد. به عبارت دیگر احتمال شرطی، بیانگر این واقعیت که اگر پیشامد A اتفاق افتاده باشد، احتمال وقوع پیشامد B چقدر خواهد بود.
احتمال شرطی B به شرط A با استفاده از فرمول زیر محاسبه می‌شود.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

پیشامدها مستقل یا غیر مستقل هستند، اگر پیشامد A و B مستقل باشند داریم:

$$P(B \mid A) = P(B)$$

$$P(A \mid B) = P(A)$$

اگر پیشامدها مستقل از هم باشند احتمال وقوع پیشامد B مستقل از احتمال A می‌باشد بنابراین احتمال وقوع همزمان دو پیشامد برابر است با حاصلضرب احتمال وقوع هر یک از پیشامدها:

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

اگر پیشامدها غیرمستقل باشند به عنوان مثال احتمال وقوع پیشامد B، منوط به وقوع پیشامد A باشد در این صورت:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

بنابراین احتمال وقوع هر دو پیشامد برابر است با:

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

مثالی از پیشامدهای مستقل: تاس دوبار پرتاب می‌شود احتمال آوردن دو عدد ۶ چقدر است؟

پرتاب اول را به عنوان پیشامد A و پرتاب دوم را به عنوان پیشامد B در نظر می‌گیریم. به دنبال احتمال $P(A \cap B)$ هستیم.

احتمال هر پیشامد برابر است با $P(B) = \frac{۱}{۶}$ و $P(A) = \frac{۱}{۶}$ پیشامدها مستقل هستند بعبارتی

$$P(A \cap B) = P(A)P(B) = (\frac{۱}{۶})(\frac{۱}{۶}) = \frac{۱}{۳۶}$$

بنابراین احتمال اینکه در دوبار پرتاب تاسی دو عدد ۶ حاصل شود $\frac{۱}{۳۶}$ است.

توزیع احتمال

قبلاً ملاحظه شد که احتمال یک متغیر همان فراوانی نسبی آن متغیر در جمعیت می‌باشد. لذا توزیع احتمال یک متغیر تصادفی نیز از توزیع فراوانی های نسبی ارزش های آن متغیر در جمعیت بدست می‌آید.سه نوع توزیع احتمال که کاربرد زیادی در تجزیه ی داده ها دارند عبارتند از، توزیع دو جمله ای، پواسن و نرمال.

توزیع دوجمله‌ای یا باینومیال

این توزیع خاص وقایعی است که دو حالت وقوع و عدم وقوع دارند. چنانچه در یک آزمایش احتمال وقوع پیشامدی مانند A برابر با P باشد و $q = ۱ - P$ احتمال عدم وقوع آن پیشامد باشد، احتمال وقوع X مرتبه از پیشامد مزبور در N آزمایش برابر است با:

$$P(X) = C_N^x P^x q^{n-x} \qquad C_n^x = \frac{n!}{X!(N-X)!}$$

$(n\text{-}x)$ عبارت است از تعداد دفعاتی که پیشامد موردنظر اتفاق نمی‌افتد. چنانچه برای تمام مقادیر X رویدادها تعیین شوند و احتمال وقوع آنها محاسبه گردد، توزیع ناپیوسته احتمال دوجمله‌ای یا باینومیال بدست می‌آید.

امید ریاضی و واریانس یک متغیر دو جمله‌ای عبارتست از:

$$E(y) = \mu = np \qquad Var(y) = \sigma^۲ = npq$$

مثال

در پرتاب یک سکه احتمال وقوع پشت (P) برابر $\frac{۱}{۲}$ و احتمال وقوع رو (B) نیز $\frac{۱}{۲}$ است اگر سکه را دو بار پرتاب کنیم نتیجه یکی از چهار حالت زیر خواهد بود.

نتیجه	سکه دوم	سکه اول
PP	P	P
PR	R	
RP	P	R
RR	R	

احتمال وقوع هر یک از پیشامدهای مزبور برابر با $\left(\frac{۱}{۲}\right)^۲$ یا $\frac{۱}{۴}$ است. و مجموع آنها برابر با یک می‌باشد.

نکته: احتمال کلاسهای موجود در توزیع دوجمله‌ای را از بسط دوجمله‌ای $(p+q)^n$ نیز محاسبه می‌کنند. n تعداد افراد در نمونه است که بسط دوجمله‌ای و سه‌جمله‌ای و nجمله‌ای به ترتیب زیر خواهد بود:

$$(p+q)^۲ = p^۲ + ۲pq + q^۲$$

$$(p+q)^۳ = p^۳ + ۳p^۲q + ۳pq^۲ + q^۳$$

$$\vdots$$

$$(p+q)^n = p^n + C_n^۱ p^{n-۱}q + C_n^۲ p^{n-۲}q^۲ + \cdots + C_n^x p^x q^{n-x} + \cdots + C_n^۱ pq^{n-۱} + q^n$$

ضمناً می‌توان ضریب دوجمله‌ای‌ها را از طریق جدول پاسکال – خیام نیز بدست آورد.

n یا توان دوجمله‌ای	ضریب دوجمله‌ای
۰	۱
۱	۱ ۱
۲	۱ ۲ ۱
۳	۱ ۳ ۳ ۱
۴	۱ ۴ ۶ ۴ ۱
۵	۱ ۵ ۱۰ ۱۰ ۵ ۱
⋮	⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮

جدول ۱- ۱: جدول پاسکال- خیام

مثال

خانواده‌ای دارای چهار فرزنداست، مطلوب است احتمال وجود حداقل یک پسر در این خانواده، احتمال پسر یا دختر بودن نوزادان را برابر با $\frac{۱}{۲}$ فرض نمایید.

$$\left.\begin{aligned} P(\text{۱ پسر}) &= C_{۴}^{۱}\left(\frac{۱}{۲}\right)^{۱}\left(\frac{۱}{۲}\right)^{۳}=\frac{۱}{۴} \\ P(\text{۲ پسر}) &= C_{۴}^{۲}\left(\frac{۱}{۲}\right)^{۲}\left(\frac{۱}{۲}\right)^{۲}=\frac{۳}{۸} \\ P(\text{۳ پسر}) &= C_{۴}^{۳}\left(\frac{۱}{۲}\right)^{۳}\left(\frac{۱}{۲}\right)^{۱}=\frac{۱}{۴} \\ P(\text{۴ پسر}) &= C_{۴}^{۴}\left(\frac{۱}{۲}\right)^{۴}\left(\frac{۱}{۲}\right)^{۵}=\frac{۱}{۱۶} \end{aligned}\right\} P(\text{حداقل یک پسر})=\frac{۱}{۴}+\frac{۳}{۸}+\frac{۱}{۴}+\frac{۱}{۱۶}=\frac{۱۵}{۱۶}$$

روش دیگر:

$$P(\text{حداقل یک پسر})=۱-P(\text{تمام فرزندان دختر})=۱-\left(\frac{۱}{۲}\right)^{۴}=\frac{۱۵}{۱۶}.$$

نکته: توزیع دوجمله‌ای یک الگوی ریاضی است که میانگین و واریانس آن به شرح زیر است:

$$\mu = np \quad , \quad \sigma^{۲} = npq \quad , \quad \sigma = \sqrt{npq}$$

توزیع پواسن

در مواردی که در توزیع دوجمله‌ای n بزرگ باشد. محاسبه احتمالات کاری پیچیده و مشکل می‌گردد. از طرفی توزیع دوجمله‌ای زمانی صادق است که احتمال p و q بزرگ باشد از توزیع پواسن بجای دو جمله‌ای استفاده می‌کنند. و احتمال وقوع پیشامد در n آزمایش بصورت زیر محاسبه می‌گردد:

$$P(x)=\frac{\lambda^{x}e^{-\lambda}}{x!}$$

پایه لگاریتمی طبیعی $= e = ۲/۷۱۸۲۸$

$$\lambda = np$$

نکته: میانگین و واریانس توزیع پواسن برابر با λ است.

از توزیع پواسن به عنوان یک الگو برای بررسی وقایعی که بطور تصادفی و بطور نادر در زمان و مکان توزیع می‌شوند استفاده می‌شود. مثلاً در مورد تعداد موارد گزارش‌شده از یک بیماری خاص، می‌توان نام برد.

مثال

میزان مرگ و میر برای یک نوع بیماری ۰/۰۰۶ است احتمال وقوع ۴ مرگ را در یک گروه ۵۰۰ نفری محاسبه نمائید.

$$p = ۰/۰۰۶ \quad , \quad q = ۰/۹۹۴$$

$$p(x) = C^{۴}_{۵۰۰}(۰/۰۰۶)^{۴}(۰/۹۹۴)^{۴۹۶}$$

مشاهده می‌گردد که بسط و حل این دوجمله‌ای مشکل می‌باشد. ولی از طریق پواسن می‌توان به راحتی به جواب رسید.

$$p(X=۴) = \frac{\lambda^{۴}e^{-\lambda}}{۴!} \quad , \quad \lambda = np = ۵۰۰(۰/۰۰۶) = ۳$$

$$p(X=۴) = \frac{۳^{۴}e^{-۳}}{۴!} = \frac{۳^{۴}(۰/۰۴۹۸)}{۴\times۳\times۲} = ۰/۱۶۸$$

توزیع برنولی: متغیری را در نظر بگیرید که فقط دارای دو مقدار باشد به عنوان مثال: بله یا خیر، یا صفر و یک چنین متغیری متغیر برنولی یا دوتایی (دو رویی) نامیده می‌شود. به عنوان مثال: اگر متغیر y، نشانگر ابتلاء به بیماری باشد در این صورت متغیر y دارای مقادیر زیر می‌باشد.

اگر حیوان بیمار باشد. $y_i = ۰$

اگر حیوان بیمار نباشد. $y_i = ۱$

توزیع احتمال متغیر تصادفی y، توزیع برنولی است:

در اینجا $q = ۱ - p$، بنابراین:

$$P(y) = P^{y}q^{۱-y} \qquad y = ۰,۱$$

$$P(y_i = ۱) = p$$

$$P(y_i = ۰) = q$$

امید ریاضی و واریانس متغیر برنولی برابر است با:

$$\sigma^{۲} = pq$$

$$E(y) = \mu = p$$

توزیع فوق هندسی

یک مجموعه N عنصری که دارای R پیروزی و N-R شکست می‌باشد را در نظر بگیرید. هر آزمایش منفرد فقط دو نتیجه در پی خواهد داشت (پیروزی یا شکست).

متغیر تصادفی y، تعداد موفقیت‌ها در یک نمونهٔ n عنصری است که از مجموعهٔ N عنصر انتخاب شده باشد چنین متغیری دارای توزیع احتمال فوق هندسی می‌باشد.

$$P(y) = \frac{\binom{R}{y}\binom{N-R}{n-y}}{\binom{N}{n}}$$

اجزاء این رابطه عبارتند از:

y= متغیر تصادفی، تعداد آزمایشات موفق در نمونه

n= اندازه نمونه

$n - y$ = تعداد شکست‌ها در نمونه

N= اندازهٔ نمونه (تعداد عناصر مجموعه)

R= تعداد آزمایش‌های موفق در مجموعه N

N-R= تعداد شکست‌ها در مجموعه N

امید ریاضی و واریانس این توزیع برابر است با:

$$Var(y)=\sigma^{۲}=\frac{nR(N-R)}{N^{۲}}(۱-\frac{n-۱}{N-۱})$$

$$E(y)=\mu=\frac{nR}{N}$$

مثال

در یک جعبه ۱۲ توله خوک نر و ۶ توله خوک ماده وجود دارد اگر ۶ توله خوک به صورت تصادفی انتخاب شود (بدون جایگذاری) احتمال برداشتن ۵ نر و ۱ ماده چقدر است؟

$$p(y)=\frac{\binom{R}{y}\binom{N-R}{n-y}}{\binom{N}{n}}=\frac{\binom{۱۲}{۵}\binom{۶}{۱}}{\binom{۱۸}{۶}}=۰/۲۵۵۹$$

بنابراین احتمال انتخاب ۵ توله خوک نر و یک ماده برابر با ۰/۲۵۵۹ می‌باشد.

توزیع چند جمله‌ای

توزیع احتمال چند جمله‌ای، تعمیمی از توزیع دو جمله‌ای است که نتیجه هر آزمایش منفرد فقط شامل بله یا خیر و یا ۰ و ۱ نیست. بلکه هر آزمایش منفرد بیش از دو نتیجه دارد. هر نتیجه دارای یک احتمال می‌باشد.

پس برای هر آزمایش منفرد K نتیجه ممکن با فراوانی‌های P_k و ... و $P_۲$ و $P_۱$ وجود دارد. همانند توزیع دو جمله‌ای، آزمایش‌ها مستقل از هم نمی‌باشند. متغیر تصادفی عبارتند از تعداد نتایج مشخص در هر آزمایش؛ $y_۱$ برای نتیجه ۱؛ $y_۲$ برای نتیجه ۲؛ و y_k برای نتیجه k تابع احتمال توزیع چند جمله‌ای عبارت است از:

$$P(y_۱,y_۲,...,y_k)=\frac{n!}{y_۱!\ y_۲!...y_k!}\ p_۱^{y_۱}\quad p_۲^{y_۲}....P_k^{y_k}$$

همچنین:

$$n=y_۱+y_۲+...+y_k$$

امید ریاضی و واریانس y_i (تعداد وقوع پیشامد I ام) برای نتیجه I برابر است با:

$$Var(y)=\sigma^{۲}=np_۱(۱-p_۱)\quad \text{و}\quad E(y_۱)=\mu_۱=np_i$$

مثال

فرض کنید صفت آسان‌زایی، در سه دسته به صورت ۱ و ۲ و ۳ تعریف شده باشد. چنانچه در هر زایش (تک قلو) احتمالات دسته‌های ۱ و ۲ و ۳ به ترتیب ۰/۶، ۰/۳ و ۰/۱ باشد، احتمال اینکه از ۱۰ گاو ماده، ۸ رأس گاو در دستهٔ اوّل، یک رأس در دستهٔ دوّم و یک رأس در دستهٔ سوّم قرار گیرد چقدر است؟ تعداد مورد انتظار گاوها در هر دسته چند رأس می‌باشد.

$$P_۱=۰/۶\quad ,\quad P_۲=۰/۳\quad ,\quad P_۳=۰/۱$$

$$n=۱۰\quad y_۱=۸\quad ,\ y_۲=۱\quad ,\ y_۳=۱$$

$$P(y_۱,y_۲,y_۳)=\frac{n!}{y_۱!y_۲!\ y_۳!}\ P_۱^{y_۱}\ P_۲^{y_۲}\ P_۳^{y_۳}=P(۸,۱,۱)=\frac{۱۰!}{۸!\ ۱!\ ۱!}(۰/۶)^{۸}\ (۰/۳)^{۱}\ (۰/۱)^{۱}=۰/۴۵$$

تعداد مورد انتظار در هر دسته برابر است با:

$$\mu_۱=np_۱=۱۰(۰/۰۶)=۶\qquad \mu_۲=np_۲=۱۰(۰/۳)=۳\qquad \mu_۳=np_۳=۱۰(۰/۱)=۱$$

توزیع یکنواخت

متغیر یکنواخت y، متغیری است که برای هر مقدار از y_i دربازهٔ $(a\le y\le b)$ احتمالی مشابه باشد تابع چگالی متغیر y به صورت زیر تعریف می‌شود.

$$F(y)=\begin{cases}\frac{1}{b-a} & \text{اگر } (a\le y\le b) \text{ آنگاه}\\ 0 & \text{در غیر اینصورت}\end{cases}$$

امید ریاضی و واریانس توزیع عبارتند از:

$$E(y)=\mu=\frac{a+b}{2} \qquad Var(y)=\sigma^2=\frac{(b-a)^2}{12}$$

توزیع نرمال

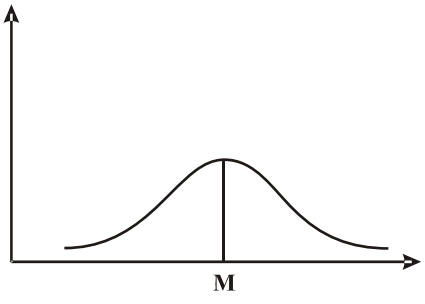
توزیع نرمال مهمترین نوع از توزیع‌های پیوسته در تجزیهٔ داده‌ها می‌باشد. این توزیع برای اولین بار توسط گوس مطالعه شد و به آن توزیع گوس نیز می‌گویند. توزیع هر متغیری که ناشی از تجمع اثرات کوچک افزایشی باشد مثل وزن حیوانات، عملکرد دانه و ارتفاع گیاه می‌تواند با توزیع نرمال توجیه گردد. منحنی توزیع احتمال نرمال توسط فرمول زیر بیان می‌شود:

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$$

شکل توزیع نرمال شبیه زنگوله و متقارن است و نسبت به میانگین جمعیت حالت قرینه دارد و کل سطح زیر منحنی برابر یک می‌باشد (شکل ۵-۱) پارامترهای μ , σ به ترتیب میانگین و انحراف استاندارد این توزیع هستند که مربع σ را یعنی σ^2 واریانس جمعیت نرمال می‌نامند.پس محل و شکل یک منحنی نرمال توسط دو کمیت μ و σ، مشخص می‌گردد.
محاسبه‌ی احتمال های مربوط به یک توزیع نرمال با دسترسی به جداول نرمال استاندارد صورت می‌گیرد. این جداول بر اساس توزیع نرمال استاندارد تنظیم شده اند. متغیر نرمال استاندارد که به «انحراف نرمال استاندارد» نیز معروف است دارای توزیع نرمال با میانگین صفر و واریانس یک می‌باشد. این متغیر را با علامت Z نشان می‌دهند و برابر است با:

$$Z=\frac{X-\mu}{\sigma}$$

جداول منحنی نرمال استاندارد در کتاب ها و در منابع مختلف به شکل متفاوت ارائه شده اند.



شکل (۱- ۱): توزیع نرمال

برای مثال در جدول (ضمیمه شماره ۱) سطح زیر منحنی بین مقادیر ۰ و Z(۳/۹)در فواصل ۰/۰۱ داده شده‌اند چون توزیع نرمال استاندارد قرینه است، دو طرف میانگین سطح احتمالی برابر ۰/۵ دارند. بطور مثال اگر مقدار Z=1/96 باشد سطح زیر منحنی ۰/۴۷۵ خواهد بود (شکل ۲-۱). برای پیدا کردن احتمال مقادیر بین X_1 و X_2 نیز می‌توان مقادیر X را به