فرمول‌های آماری و محاسبه احتمال

مفاهیم اولیه

جامعه و نمونه (Population - Sample)

جامعه آماری: به مجموعه اشیا یا افرادی گفته می‌شود که به منظور بررسی یک یا چند ویژگی کمی یا کیفی مورد مطالعه قرار می‌گیرند.

داده‌های کمی: به کمک اندازه‌گیری یا شمارش ویژگی‌ها به دست می‌آیند.

داده‌های کیفی: مقادیری هستند که از طریق کدگذاری روی ویژگی‌های کیفی حاصل می‌شوند.

جامعه آماری متناهی: به جامعه‌ای گفته می‌شود که اعضای آن قابل‌شمارش و باپایان باشد.

جامعه آماری نامتناهی: اعضای این جامعه، قابل‌شمارش ولی بی‌پایان یا غیرقابل شمارش هستند.

نمونه: قسمتی از جامعه آماری است که طبق قانون خاصی به منظور بررسی و استنباط در مورد پارامترهای آن انتخاب می‌شود.

پارامتر: به مقادیری گفته می‌شود که بیانگر خصوصیات جامعه آماری باشند.

آماره: مقادیری هستند که براساس نمونه محاسبه شده و در آن پارامتر جامعه وجود ندارد.

معیارهای تمرکز (Central Tendency)

مد – نما (M): مقداری که بیشترین فراوانی را دارد.

میانه (m): مقداری که ۵۰٪ داده‌ها از آن کوچکتر یا بزرگتر هستند.

$$m = \begin{cases} X_{\left(\frac{n}{2}\right)} & n \text{ فرد} \\ \dfrac{\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right)}{2} & n \text{ زوج} \end{cases}$$

میانگین: مرکز ثقل داده‌ها است. (μ میانگین جامعه و $\overline{X}$ میانگین نمونه)

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

نکته: مجموع فاصله داده‌ها از میانگین برابر با صفر است.

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

انواع میانگین (Mean)

میانگین حسابی: از رابطه زیر به دست می‌آید.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

میانگین براساس جدول فراوانی: اگر f_i فراوانی رده iام باشد، میانگین حسابی براساس جدول فراوانی برابر است با

$$\overline{X}_f = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i}$$

میانگین وزنی: اگر $1 \le w_i \le 0$ وزن مربوط به مشاهده iام باشد، میانگین وزنی برابر است با

$$\overline{X}_w = \sum_{i=1}^{n} w_i x_i$$

میانگین هندسی: اگر مقدارهای x همگی مثبت باشند، میانگین هندسی برابر است با

$$G = \sqrt[n]{x_1 x_2 x_3 \cdots x_n}$$

میانگین هندسی براساس جدول فراوانی: برابر است با

$$G = \sqrt[n]{{x_1}^{f_1}{x_2}^{f_2}\cdots{x_n}^{f_n}}, \sum f_i = n$$

میانگین توافقی (همساز): اگر مقدارهای x همگی مثبت باشند، میانگین توافقی برابر است با

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

میانگین توافقی (همساز) براساس جدول فراوانی: برابر است با

$$H = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}}$$

میانگین آمیخته: اگر میانگین m مقدار برابر با $\overline{X}_m$ و میانگین n مقدار دیگر برابر با $\overline{X}_n$ باشد، میانگین کل این دو دسته برابر است با

$$\overline{X}_{n+m} = \frac{n\overline{X}_m + m\overline{X}_n}{n+m}$$

رابطه بین میانگین، میانه و نما: به صورت زیر است.

$$\overline{x} - M = 3(\overline{x} - m)$$

رابطه بین میانگین حسابی، میانگین هندسی و میانگین توافقی (همساز):

$$H \le G \le \overline{X}$$

معیارهای پراکندگی (Dispersion Measures)

دامنه تغییرات (R): تفاوت بین بزرگترین و کوچکترین مقدار را نشان می‌دهد.

$$R = \max(x) - \min(x)$$

واریانس جامعه (σ^2): میانگین مربعات تفاضل مقدارها از میانگین است.

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

واریانس نمونه (S^2): از رابطه زیر به دست می‌آید.

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

انحراف معیار (s, σ): برابر با جذر واریانس است.

$$\sigma = \sqrt{\sigma^2}, \quad S = \sqrt{S^2}$$

قدرمطلق انحراف از میانگین: معدل فاصله مقادیر از میانگین آن‌ها محسوب و از رابطه زیر محاسبه می‌شود.

$$AD = \frac{\sum_{i=1}^{n}|x_i - \mu|}{n}$$

چندک‌ها (Quantiles)

چارک: مقدارهایی که داده‌ها را به چهار قسمت مساوی تقسیم می‌کنند. (چارک اول p=25%، دوم p=50%، سوم p=75%)

دهک: مقدارهایی که داده‌ها را به ده قسمت مساوی تقسیم می‌کنند. (دهک اول p=10%، دوم p=20%، ...، نهم p=90%)

صدک: مقدارهایی که داده‌ها را به صد قسمت مساوی تقسیم می‌کنند. (صدک اول p=1%، دوم p=2%، ...، نود و نهم p=99%)

$$Q_p = X_{(np)}$$

محاسبه چندک pام از جدول فراوانی:

$$Q_p = L_p + \frac{(pn - F_p)}{f_p} w$$

$\boldsymbol{L_p}$: کران پایین رده چندک pام

$\boldsymbol{F_p}$: فراوانی تجمعی رده قبل از چندک pام

$\boldsymbol{f_p}$: فراوانی رده چندک pام

$\boldsymbol{w}$: طول رده

نکته: میانه، چارک دوم، دهک پنجم و صدک پنجاهم است.

ضریب چولگی (Coefficient of Skewness)

ضریب چولگی: میزان انحراف افقی منحنی توزیع داده‌ها محسوب می‌شود.

$$S_{skew1} = \frac{3(\overline{X} - m)}{s}, \quad S_{2skew} = \frac{\overline{X} - M}{s}$$

چوله به چپ: حالتی را گویند که ضریب چولگی منفی باشد.
چوله به راست: حالتی را گویند که ضریب چولگی مثبت باشد.

ضریب تغییرات (Coefficient of Variation)

اگر s انحراف استاندارد و $\overline{X}$ میانگین باشد، ضریب تغییرات برابر است با

$$V = \frac{S}{\overline{X}}$$

نکته: هرچه مقدار ضریب تغییرات کوچکتر باشد، پراکندگی نسبت به میانگین کمتر است.

داده‌های استاندارد شده (Z-Score)

اگر هر داده را از میانگین کم کرده و حاصل را بر انحراف استاندارد تقسیم کنیم، نتیجه را مقدار استاندارد یا Z-score می‌نامند.

$$Z_{score}(x) = \frac{x - \overline{x}}{s}$$

نکته: معمولا داده‌های استاندارد شده در بازه ۳- تا ۳ قرار دارند. میانگین داده‌های استاندارد شده برابر با صفر و واریانس آن‌ها برابر با ۱ است.

همبستگی (Correlation)

کوواریانس: میزان ارتباط بین دو متغیر X و Y را نشان می‌دهد و از رابطه زیر محاسبه می‌شود.

$$Cov(X,Y) = \overline{XY} - \overline{X}\,\overline{Y}$$

ضریب همبستگی: میزان ارتباط خطی بین دو متغیر را نشان می‌دهد و مقداری بین ۱- و ۱ است.

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{S_X^2 S_Y^2}} = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{S_X S_Y}$$

ترکیبیات (Combinatorial)

اصول شمارش (Counting Rules)

اصل جمع: اگر کار A به n طریق و کار B به m طریق امکان‌پذیر باشند، کار A یا B به m+n طریق امکان‌پذیر است.

اصل ضرب: اگر کار A به n طریق و کار B به m طریق امکان‌پذیر باشد، کار A و B بطور متوالی به n×m طریق امکان‌پذیر هستند.

صورت کلی اصل ضرب: اگر کار یک، دو الی کار m به ترتیب با n_1، n_2 و ... n_m طریق امکان‌پذیر باشند، این کارها به صورت متوالی به $n_1 \times n_2 \times \cdots \times n_m$ طریق قابل انجام خواهند بود.

جایگشت (Permutation)

تعداد حالات قرارگیری n شی متمایز (بدون تکرار) در کنار هم به ترتیب برابر با !n است.

تعداد حالاتی که n شی متمایز (با تکرار) در کنار یکدیگر قرار گیرند، n^n است.

ترتیب k شی از n شی

انتخاب k شی از n شی با در نظر گرفتن ترتیب و بدون تکرار برابر است با

$$P_k^n = \frac{n!}{(n-k)!}$$

انتخاب k شی از n شی با در نظر گرفتن ترتیب و بدون تکرار برابر است با

$$n^k$$

ترکیب k شی از n شی

انتخاب k شی از n شی متمایز (بدون تکرار) که در آن ترتیب قرارگیری مهم نیست به صورت زیر محاسبه می‌شود.

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

رابطه زیر برای محاسبه انتخاب k شی از n شی متمایز (با تکرار) که در آن ترتیب قرارگیری مهم نیست و همچنین تعداد ریشه‌های نامنفی و صحیح n مجهولی $X_1 + X_2 + \cdots + X_n = k$، مورد استفاده قرار می‌گیرد.

$$\frac{(n+k-1)!}{k!\,(n-1)!}$$

احتمال (Probability)

آزمایش تصادفی، فضای نمونه، پیشامد

آزمایش تصادفی: آزمایشی است که نتایج آن با تکرار آزمایش در شرایط یکسان، متفاوت و برمبنای تصادف حاصل می‌شوند.

فضای نمونه Ω: به مجموعه نتایج آزمایش تصادفی گفته می‌شود.

پیشامد: هر زیرمجموعه‌ای از فضای نمونه یک پیشامد است.

پیشامد متمم *A*: اعضایی از فضای نمونه که در A نیستند پیشامد متمم A نامیده می‌شود و به صورت A'=Ω-A نیز نمایش داده می‌شود.

پیشامدهای ناسازگار: پیشامد A و B ناسازگار هستند، اگر $A \cap B = \emptyset$ باشد.

پیشامد حتمی: به Ω یا همان فضای نمونه گفته می‌شود.

پیشامد غیرممکن: به مجموعه تهی یا $\emptyset$ گفته می‌شود.

اصول احتمال (Probability Axioms)

اصل اول: احتمال برای هر پیشامد، نامنفی است. بنابراین اگر A یک پیشامد باشد، داریم $P(A) \geq 0$.

اصل دوم: مقدار احتمال برای فضای نمونه برابر با یک است، $P(\Omega) = 1$.

اصل سوم: احتمال اجتماع هر دنباله نامتناهی از پیشامدهای دو به دو ناسازگار، برابر با مجموع احتمال پیشامدهای دنباله است.

$$P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1} P(A_i)$$

قضیه‌های احتمال

- احتمال پیشامد تهی برابر است با صفر. $P(\emptyset) = 0$.
- احتمال اجتماع متناهی از پیشامدهای دو به دو ناسازگار برابر است با جمع احتمالات آن‌ها $P(A_1 \cup A_2) = P(A_1) + P(A_2)$
- احتمال برای مکمل پیشامد A برابر است با $P(A') = 1 - P(A)$
- اگر پیشامد A زیر مجموعه پیشامد B باشد، مقدار احتمال نیز برای پیشامد A از B کمتر است، $P(A) \leq P(B)$.
- اگر C پیشامدی باشد که از اجتماع دو پیشامد A و B ایجاد شده، احتمال پیشامد C برابر است با $P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

احتمال شرطی (Conditional Probability)

احتمال رخداد پیشامد A به شرط اطلاع از رخداد B برابر است با

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

دو پیشامد مستقل (Independent Events)

پیشامد A را نسبت به B مستقل گویند اگر

$$P(A|B) = P(A) \text{ یا } P(A \cap B) = P(A)P(B)$$

نکته: اگر پیشامد A از B مستقل باشد، پیشامد B نیز از A مستقل است.

افزار فضای نمونه و قضیه احتمال کل

$B_1, B_2, \cdots, B_n$ را یک افراز روی فضای نمونه Ω گویند، اگر

$$\Omega = B_1 \cup B_2 \cup \cdots \cup B_n$$

به طوری که پیشامدهای B_i، دو به دو ناسازگار باشند، یعنی

$$B_i \cap B_j = \emptyset, i \neq j$$

قضیه احتمال کل: اگر فضای نمونه (Ω) به مجموعه‌های مجزا افراز شود، داریم

$$\mathrm{P}(\mathrm{A}) = \mathrm{P}(A|\mathrm{B}_1)P(\mathrm{B}_1) + P(A|B_2)P(B_2) + \cdots + P(\mathrm{A}|B_n)P(B_n)$$

قضیه بیز (Bayes' theorem)

اگر فضای نمونه به $B_1, B_2, \cdots, B_n$ افراز شده باشد، بطوری که $P(\mathrm{B_i}) \neq 0$، برای هر پیشامد A می‌توان نوشت

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum P(A|B_j)P(B_j)}$$

برای مشاهده دیگر «تقلب‌نامه‌های» مجله فرادرس، به این لینک مراجعه فرمایید.

جهت آگاهی از آخرین تقلب‌نامه‌های منتشر شده، در کانال تلگرام مجله فرادرس عضو شوید.

تهیه و تنظیم: مجله فرادرس