

فرمول‌ها و درجات آزادی مهم آمار

همواره و در هر توزیعی مجموع انحراف داده‌ها از میانگین برابر صفر است: $\sum (x - \bar{x}) = 0$

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots}{n_1 + n_2 + \dots}$$

میانگین وزنی

$$Q = \frac{Q_r - Q_l}{2}$$

انحراف چارکی

$$S^2 = \frac{\sum (x - \bar{x})^2}{n}$$

واریانس

$$S = \sqrt{S^2}$$

انحراف استاندارد (معیار)

$$CV = \frac{S}{\bar{x}}$$

ضریب تغییرات

$$Z = \frac{x - \bar{x}}{S} \rightarrow \begin{cases} \bar{X}_Z = \bar{Z} = 0 \\ S_Z = 1 \end{cases}$$

نمره استاندارد

ضریب همبستگی گشتاوری پیرسون

$$r_{xy} = \frac{Cov_{xy}}{S_x S_y}$$

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

به کمک آزمون t می‌توان معناداری همبستگی پیرسون را بررسی کرد:

$$r_{xy}^2 = \frac{Cov_{xy}^2}{S_x^2 S_y^2} = \frac{\text{واریانس مشترک}}{\text{واریانس کل}} = \frac{SS_{reg}}{SS_{total}}$$

نحوه تشکیل معادله رگرسیون:

$$\left. \begin{aligned} b &= r_{xy} \frac{S_y}{S_x} = \frac{Cov_{xy}}{S_x^2} \\ a &= \bar{y} - b\bar{x} \end{aligned} \right\} \hat{y} = bx + a$$

$$(S_{xy}) \text{ خطای استاندارد پیش‌بینی} = S_y \sqrt{1 - r^2}$$

واریانس میانگین‌ها (واریانس نمونه‌گیری) $S_{\bar{x}} = \frac{s}{\sqrt{n}} \rightarrow S_{\bar{x}}^2 = \frac{s^2}{n}$ انحراف معیار میانگین‌ها (واریانس نمونه‌گیری)

$$\mu = \bar{X} \pm Z S_{\bar{x}} \text{ فاصله اطمینان جامعه}$$

$$n = \frac{z^2 s^2}{d^2} = \left(\frac{zs}{d}\right)^2 \text{ محاسبه حجم نمونه}$$

d: نصف طول فاصله اطمینان - میزان خطای حاشیه‌ای - تفاوت بین میانگین‌ها - حداکثر خطای قابل قبول - دقت برآورد - اندازه یا حجم اثر

محاسبه ضرایب همبستگی به کمک خی دو دومتغیره

آزمون خی دو $\chi^2 = \sum \frac{(O - E)^2}{E}$ تنها معناداری همبستگی را بررسی می‌کند و قادر به محاسبه شدت همبستگی نیست؛

برای محاسبه شدت رابطه (اندازه اثر) می‌توان از ضرایب زیر استفاده کرد:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \text{ توافقی}$$

$$\phi = \sqrt{\frac{\chi^2}{n}} \text{ فی}$$

$$V = \sqrt{\frac{\chi^2}{n(a-1)}} \text{ کرامر}$$

a: حداقل تعداد سطر یا ستون

آزمونهای مقایسه‌ای

$$Z = \frac{\bar{x} - \mu}{s_{\bar{x}}} \quad \text{یا } t \text{ تک گروهه}$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}} = \frac{\text{تفاضل دو میانگین}}{\text{خطای استاندارد تفاضل دو میانگین}} \quad \text{یا } t \text{ مستقل}$$

$$t = \frac{\bar{D}}{S_{\bar{D}}} \quad \text{یا } z \text{ وابسته}$$

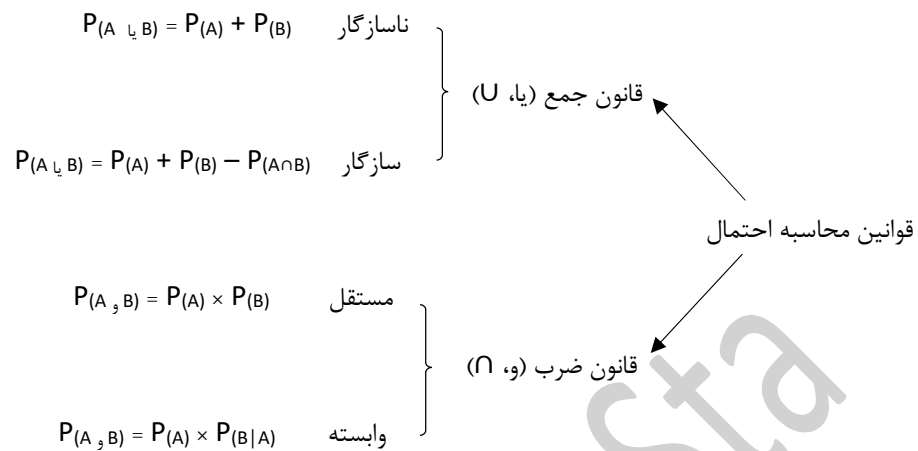
$$F = \frac{\text{تفاوت بین گروهها}}{\text{تفاوت درون گروهها}} = \frac{\text{پراکندگی بین گروهی}}{\text{پراکندگی درون گروهی}} = \frac{\text{واریانس بین گروهی}}{\text{واریانس درون گروهی}} = \frac{\text{میانگین مجذورات بین گروهی}}{\text{میانگین مجذورات درون گروهی}} = \frac{\text{Mean Squares of Between}}{\text{Mean Squares of Within}}$$

$$F = \frac{S_b^2}{S_w^2} = \frac{MS_b}{MS_w} = \frac{MS_{tr}}{MS_e} = \frac{\text{واریانس آزمایشی}}{\text{واریانس خطا}} = \frac{\text{واریانس نظام‌دار (سیستماتیک)}}{\text{واریانس مزاحم}}$$

$$F = \frac{\frac{\text{مجموع مجذورات بین گروهی}}{\text{درجه آزادی بین گروهی}}}{\frac{\text{مجموع مجذورات درون گروهی}}{\text{درجه آزادی درون گروهی}}} = \frac{\frac{SS_b}{df_b}}{\frac{SS_w}{df_w}}$$

اندازه اثر

$$\eta^2 = \frac{SS_b}{SS_{total}} \quad \text{یا} \quad 1 - \frac{SS_w}{SS_{total}}$$



اصل شمارش

اگر رویدادی به n طریق و رویداد دیگری به m طریق ممکن باشد، تعداد حالات وقوع هم‌زمان این دو رویداد $n \times m$ خواهد بود.

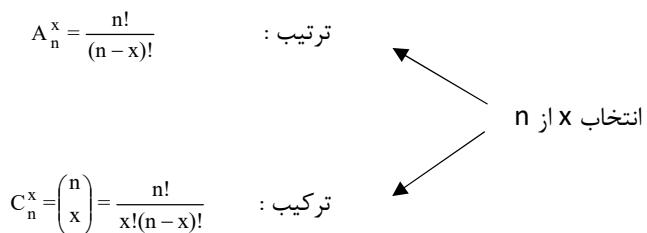
اصل جایگشت

تعداد حالاتی که می‌توان n شیء، شخص یا ... را در کنار هم قرار داد برابر است با:

$$n! = n \times (n - 1) \times (n - 2) \times \dots \times 1$$

انتخاب

تعداد حالاتی که می‌توان x شیء یا شخص را از میان n تای آن انتخاب کرد، برابر است با ترکیب یا ترتیب x از n .



توزیع دوجمله‌ای (برنولی)

به کمک این توزیع می‌توان احتمال وقوع X بار رویداد مورد نظر را از میان وقوع n بار رویدادی دوحالتی، محاسبه کرد
(احتمال موفقیت در n بار وقوع یک رویداد دو وضعیتی):

$$(P + q)^n = \binom{n}{x} p^x q^{n-x}$$

P : احتمال موفقیت در یک بار وقوع رویداد و $q = 1 - p$

محاسبه شاخص‌های مرکزی و پراکندگی توزیع دوجمله‌ای

میانگین و واریانس توزیع دوجمله‌ای برابر است با:

$$\bar{X} = np$$

$$S^2 = npq$$

$$S = \sqrt{npq}$$

درجه آزادی χ^2 دو تک متغیره: $df = K - 1$

درجه آزادی χ^2 دو دومتغیره (درجه آزادی جداول توافقی $r \times c$): $df = (r - 1)(c - 1)$

و درجه آزادی t تک نمونه‌ای برابر است با: $df = n - 1$

$n_1 - 2$ یا $n_2 - 2$ $df = n_1 + n_2 - 2$ درجه آزادی t مستقل

و درجه آزادی t وابسته نیز همانند t تک گروه برابر است با: $df = n - 1$

درجه آزادی معناداری یک ضریب همبستگی: $df = n - 2$

درجه آزادی های تحلیل واریانس تک عاملی برای مقایسه گروهها

درجه آزادی بین گروهی یا آزمایشی (صورت): $df_b = K - 1$

درجه آزادی درون گروهی یا خطا (مخرج): $df_w = n_{\text{کل}} - K$

K : تعداد گروهها یا مقولهها و n : تعداد کل آزمودنیها ($n_1 + n_2 + \dots$)

درجه آزادی های تحلیل واریانس با اندازه گیری مکرر

تغییر عمل آزمایشی تغییر مانده (خطا)

$$df = j - 1 \quad df = (j - 1)(n - 1)$$

j : تعداد مراحل اندازه گیری

محاسبه درجات آزادی دو عاملی

۱) درجه آزادی مربوط به متغیرهای A و B : برابر است با تعداد سطوح هر متغیر منهای ۱:

$$df_A = (a - 1) \quad \text{و} \quad df_B = (b - 1)$$

۲) درجه آزادی مربوط به تعامل AB : برابر است با حاصل ضرب درجات آزادی A در B :

$$df_{AB} = (a - 1)(b - 1)$$

۳) درجه آزادی مربوط به عامل خطا (درون گروهها): برابر است با حاصل ضرب تعداد سطوح دو متغیر A و B در $(a.b)$ در تعداد کل مشاهدات (n) منهای یک.

$$df_{\text{error}} = ab(n - 1)$$

درجه آزادی های آزمون F برای بررسی معناداری R^2 (مورد استفاده در تحلیل رگرسیون)

$$df_{\text{reg}} = k$$

$$df_{\text{res}} = n - k - 1$$

(k : تعداد متغیرهای x یا پیش بین)