

توزیع های نمونه ای

فصل ششم

جامعه آماری و متغیرهای مورد بررسی

عبداله جلیلیان، گروه آمار، دانشگاه رازی

جامعه (جمعیت) آماری: همهی واحدها (افراد، اشیاء و ...) که برای یک هدف پژوهشی معین باید مورد ارزیابی قرار گیرند

هدف پژوهش: آگاهی از وضعیت مشخصه (هایی) از جامعه یا بررسی فرض (هایی) در مورد جامعه

متغیر(های) مورد بررسی: ویژگی یا کمیت‌هایی که در راستای هدف تحقیق باید اندازه‌گیری و ثبت شوند

مثال:	جامعه آماری	متغیرهای مورد بررسی
	واحدهای مسکن مهر در سراسر کشور	میزان استحکام لرزه‌ای، میزان بهره‌وری انرژی
	خودروهای تولید داخل از سال ۱۳۹۷ تا کنون	میزان آلایندگی، میزان مصرف سوخت
	لامپ‌های تولید شده توسط یک کارخانه‌ی خاص	طول عمر
	میلگرد A4 با قطر ۱۲ میلیمتر یک تولیدکننده خاص	مقاومت کششی، مقاومت در مقابل خوردگی

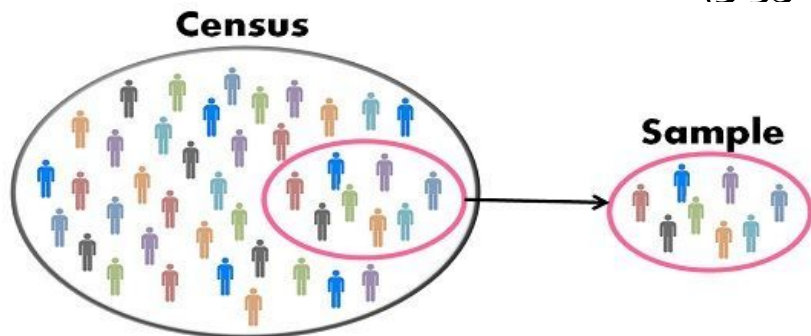
سرشماری و نمونه‌گیری

سرشماری: اندازه‌گیری و ثبت متغیر(های) مورد نظر برای تمامی اعضای جامعه

- دسترسی به تمامی اعضای جامعه ممکن است امکان‌پذیر نباشد
- زمان‌بر و هزینه‌بر است
- اندازه‌گیری متغیر(های) مورد بررسی ممکن است به تخریب یا صدمه به اعضای جامعه منجر شود
- اندازه‌گیری متغیر(های) مورد بررسی برای تمامی اعضای جامعه ممکن است با دقت کمی صورت گیرد

نمونه‌گیری: اندازه‌گیری و ثبت متغیر(های) مورد نظر برای یک زیرمجموعه از اعضای جامعه

نمونه مطلوب: باید معرف و نماینده‌ی کل جامعه باشد (مشت نمونه‌ی خروار)



طرح نمونه‌گیری: سازوکار انتخاب و قرارگیری اعضای جامعه در نمونه

حجم نمونه: تعداد اعضای جامعه که در نمونه حضور دارند

نمونه تصادفی

نمونه تصادفی: اعضای جامعه به صورت تصادفی و با احتمال‌های معین انتخاب و در نمونه قرار می‌گیرند

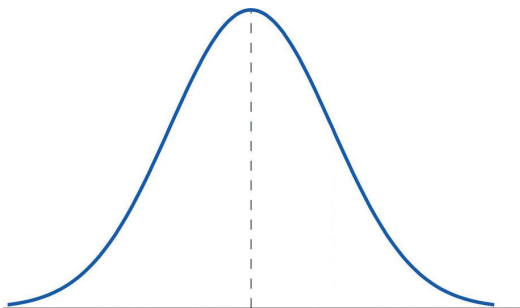
نمونه تصادفی ساده: همگی اعضای جامعه شانس یکسانی برای قرارگیری در نمونه دارند. برای این منظور باید فهرست همگی اعضای جامعه در اختیار باشد و به تصادف (قرعه‌کشی) اعضای نمونه از این فهرست انتخاب شوند

$$X_1, \dots, X_n \stackrel{iid}{\sim} F_X$$

نمونه تصادفی ساده: متغیرهای تصادفی مستقل و هم‌توزیع

توزیع جامعه: $F_X(x)$ تابع توزیع جامعه است که اغلب نامعلوم است

جامعه نرمال: توزیع جامعه نرمال با میانگین μ و انحراف معیار σ است که اغلب مجهول هستند



توزیع نمونه‌ای

عبداله جلیلیان، گروه آمار، دانشگاه رازی

$$T(X_1, \dots, X_n)$$

آماره: هر تابعی از نمونه‌ی تصادفی یک

مقدار آماره از نمونه‌ای به نمونه‌ی دیگر تغییر می‌کند و با توجه به تصادفی بودن نمونه، پیش از مشاهده‌ی مقادیر نمونه تصادفی است

توزیع نمونه‌ای آماره: متغیر تصادفی $T(X_1, \dots, X_n)$ دارای چه توزیعی است؟

$$F_T(t) = P\{T(X_1, \dots, X_n) \leq t\}$$

آماره‌های پرکاربرد

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

• میانگین نمونه

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

• واریانس نمونه

توزیع نمونه‌ای میانگین نمونه

امید ریاضی و واریانس میانگین نمونه

$$E[\bar{X}] = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n}$$

توزیع نمونه‌ای میانگین نمونه

• در حالت جامعه نرمال: برای هر حجم نمونه‌ی دلخواهی دقیقاً نرمال است

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

• در حالت سایر توزیع‌ها: بنابر قضیه‌ی حد مرکزی، با افزایش حجم نمونه به توزیع نرمال نزدیک و نزدیک‌تر می‌شود

$$n > 30 : \quad \bar{X} \overset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

توزیع نمونه‌ای واریانس نمونه

امید ریاضی واریانس نمونه

$$E[S^2] = \sigma^2$$

توزیع نمونه‌ای واریانس نمونه در حالت جامعه نرمال

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

تابع چگالی احتمال توزیع $\chi^2(m)$: توزیع کای-دو (خی-دو) با درجه‌ی آزادی $m = n - 1$

$$f_Y(y) = \frac{1}{\Gamma(m)2^{\frac{m}{2}}} y^{\frac{m}{2}-1} e^{-\frac{y}{2}}$$

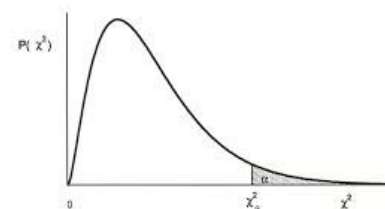


Figure J.1: The χ^2 distribution

احتمال‌های مربوط به توزیع کای-دو با استفاده از جدول‌های انتهای کتاب یا نرم‌افزارها قابل محاسبه است

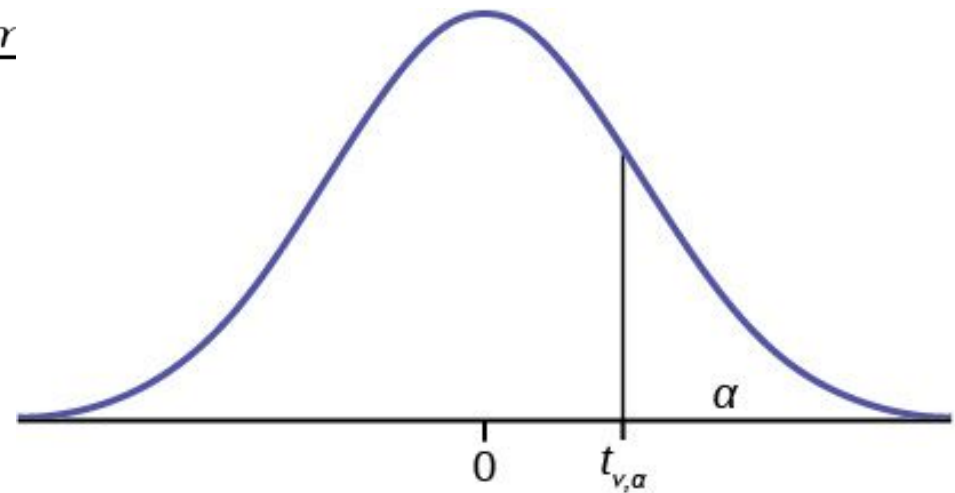
توزیع نمونه‌ای آماره‌ی t

توزیع نمونه‌ای آماره‌ی t در حالت جامعه نرمال

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n - 1)$$

تابع چگالی احتمال توزیع t ی استودنت با درجه‌ی آزادی $m = n - 1$

$$f_T(t) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})\sqrt{\pi m}} \left(1 + \frac{t^2}{m}\right)^{-\frac{n}{2}}$$



احتمال‌های مربوط به توزیع t ی استودنت با استفاده از جدول‌های انتهای کتاب یا نرم‌افزارها قابل محاسبه است

توزیع‌های نمونه‌ای برای دو نمونه مستقل

عبداله جلیلیان، گروه آمار، دانشگاه رازی

فرض کنید $X_{11}, X_{12}, \dots, X_{1n_1}$ و $X_{21}, X_{22}, \dots, X_{2n_2}$ دو نمونه‌ی تصادفی با حجم‌های n_1 و n_2 از دو جامعه‌ی مستقل با میانگین‌های μ_1 و μ_2 و واریانس‌های σ_1^2 و σ_2^2 باشند. اگر \bar{X}_1 و \bar{X}_2 میانگین‌های این دو نمونه و S_1^2 و S_2^2 واریانس‌های این دو نمونه باشند، آن‌گاه در حالت جامعه‌های نرمال برای اختلاف میانگین دو نمونه داریم

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

و برای نسبت واریانس‌های دو نمونه نیز

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1 - 1, n_2 - 1)$$

احتمال‌های مربوط به توزیع F با استفاده از جدول‌های انتهای کتاب یا نرم‌افزارها قابل محاسبه است

توزیع های نمونه ای برای دو نمونه مستقل

عبداله جلیلیان، گروه آمار، دانشگاه رازی

با فرض برابری واریانس های دو جامعه

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$S_p^2 = \frac{1}{n_1 + n_2 - 2} \left((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 \right)$$

بدون فرض برابری واریانس های دو جامعه

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx t(\min(n_1 - 1, n_2 - 1))$$