توزیع های نمونه ای:

تعریف: اگر $X_1, X_2, \ldots, X_n$ متغیر های تصادفی مستقل و همتوزیع از جامعه $f_X(x)$ باشند، می‌گوییم تشکیل یک نمونه تصادفی از جامعه‌ای نامتناهی را می‌دهند. در این صورت داریم:

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n) = \prod_{i=1}^{n} f(x_i)$$

تعریف: هر تابع حقیقی روی یک نمونه تصادفی مانند $X_1, X_2, \ldots, X_n$ را یک آماره می‌گوییم اگر به پارامتر مجهولی بستگی نداشته باشد. واضح است که یک آماره خود نیز یک متغیر تصادفی است.

به عنوان مثال

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

میانگین نمونه ای و

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

واریانس نمونه ای و

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

انحراف معیار نمونه‌ای هر سه آماره هستند. همچنین اگر مقدار μ مجهول نباشد

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

یک آماره می‌باشد.

تذکر: معمولا اصطلاحات "نمونه تصادفی" یا "آماره"، "میانگین نمونه‌ای"، و "واریانس نمونه‌ای" را در مورد مقادیر متغیرهای تصادفی، به جای خود آنها نیز به کار می‌برند.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

در اینجا x_i، $\bar{x}$، و s^2 مقادیر X_i، $\bar{X}$ و S^2 هستند که پس از مشاهده نمونه به دست آمده‌اند.

توزیع میانگین:

اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه ای نامتناهی با میانگین μ و واریانس σ^2 باشد در این صورت

$$\mu_{\bar{X}} = E[\bar{X}] = \mu \qquad \sigma^2_{\bar{X}} = Var[\bar{X}] = \frac{\sigma^2}{n}$$

اثبات:

$$E[\bar{X}] = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}(n\mu) = \mu$$

$$Var[\bar{X}] = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

نتیجه: با توجه به قضیه بالا اگر حجم نمونه، n، افزایش پیدا کند واریانس $\bar{X}$ کم می‌شود و در صورتی که $n \to \infty$ مقدار واریانس میانگین نمونه ای به سمت صفر میل پیدا کرده و در واقع مقادیر $\bar{X}$ به μ نزدیک و نزدیک تر می‌شود.

قانون اعداد بزرگ: به ازای هر ثابت مثبت c،

$$P\left(\mu - c < \bar{X} < \mu + c\right) \geq 1 - \frac{\sigma^2}{nc^2}$$

قضیه حد مرکزی: اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه ای نامتناهی با میانگین μ و واریانس σ^2 باشد، در این صورت توزیع حدی

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

وقتی $n \to \infty$ ، توزیع نرمال استاندارد است.

طبق قضیه حد مرکزی اگر n به اندازه کافی بزرگ باشد ($n \geq 30$) می توان توزیع میانگین نمونه‌ای $\bar{X}$ را به کمک توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ تقریب زد و این به توزیع جامعه بستگی ندارد.

$$n \geq 30 \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

تذکر: همانطور که قبلا دیدیم، هر ترکیب خطی از متغیر های تصادفی مستقل و نرمال، متغیری نرمال است. بنابراین اگر $\bar{X}$ میانگین یک نمونه تصادفی از جامعه نرمالی با میانگین μ و واریانس σ^2 باشد، بدون توجه به حجم نمونه داریم:

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

مثال: وزن کیک های تولید شده توسط یک کارخانه دارای میانگین 1500 و واریانس 2304 است. یک نمونه تصادفی به حجم ۳۶ از کیک های تولید شده توسط این کارخانه در نظر گرفته ایم، احتمال اینکه میانگین وزن آنها کمتر از ۱۴۹۰ باشد، چقدر است؟

$$\mu = 1500$$
$$\sigma^2 = 2304$$
$$n = 36$$
$$\sigma^2_{\bar{X}} = \frac{2304}{36}$$
$$\bar{X} \sim N(1500, 64)$$
$$P\left(\bar{X} \leq 1490\right) = P\left(\frac{\bar{X} - 1500}{8} \leq \frac{1490 - 1500}{8}\right) = P\left(Z \leq -1.25\right) = 0.1$$

مثال: قد دانشجویان یک دانشگاه از توزیع **نرمال** با میانگین ۱۷۱ و واریانس ۸۰ پیروی می‌کند. از بین این دانشجویان ۵ دانشجو را به تصادف انتخاب کرده و میانگین قد آنها را محاسبه می‌کنیم. احتمال اینکه میانگین قد آنها کمتر از ۱۶۹ باشد، چقدر است؟

$$X_i \sim N(171,80)$$
$$n=5, \sigma^2_{\bar{X}}=16$$
$$\bar{X} \sim N\left(171, 16=\frac{80}{5}\right)$$
$$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$
$$P(\bar{X} \le 169)=P\left(\frac{\bar{X}-171}{4} \le \frac{169-171}{4}\right)=P(Z \le -0.5)=0.3$$

توزیع واریانس نمونه‌ای:

قضیه: اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از جامعه ای نامتناهی با میانگین μ و واریانس σ^2 باشد در این صورت

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

اثبات:

$$E[S^2]=E\left[\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}\right]=\frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i^2-2\bar{X}X_i+\bar{X}^2\right)\right]$$
$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}E[X_i^2]-2E\left[\sum_{i=1}^{n}\bar{X}X_i\right]+\sum_{i=1}^{n}E[\bar{X}^2]\right)$$

حال با توجه به اینکه:

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

داریم:

$$E\left[S^2\right]=\frac{1}{n-1}\left(\sum_{i=1}^{n}E[X_i^2]-2nE\left[\bar{X}^2\right]+\sum_{i=1}^{n}E[\bar{X}^2]\right)$$
$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}E[X_i^2]-nE\left[\bar{X}^2\right]\right)$$

از طرفی:

$$Var[Y] = E[Y^2] - E[Y]^2$$

$$E\left[\bar{X}^2\right] = Var\left[\bar{X}\right] + E\left[\bar{X}\right]^2 = \frac{\sigma^2}{n} + \mu^2$$

$$E\left[X_i^2\right] = Var\left[X_i\right] + E\left[X_i\right]^2 = \sigma^2 + \mu^2$$

پس:

$$E\left[S^2\right] = \frac{1}{n-1}\left(n\left(\mu^2 + \sigma^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

قضیه: اگر Z یک متغیر تصادفی نرمال استاندارد باشد در این صورت متغیر تصادفی Z^2 یک متغیر تصادفی خی دو با یک درجه آزادی است.