متغیر تصادفی (Random variable):

به عنوان مثال:

x: معیوب بودن رایانه

n=10 تعداد کل رایانه ها

مقادیری که x می‌تواند به خود اختصاص دهد:

x=0, 1, 2, … , 10

توزیع بینم $b(x; n, p)$

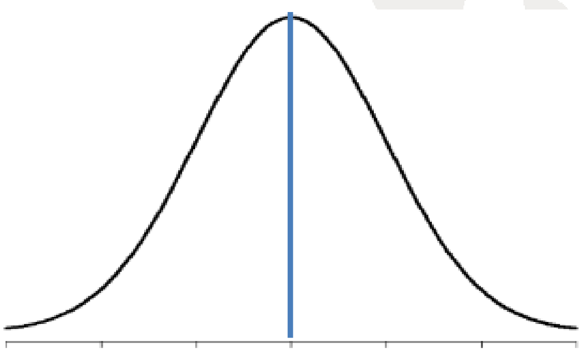
به صورت گسسته بوده که احتمال موفقیت p و تعداد کل حالات n می‌باشد.

توزیع پواسون $p(x;\ \mu)$

به صورت پیوسته می‌باشد و فاقد تعداد کل حالات و جامعه بوده و از مولفه میانگین تعداد حالات اتفاق افتاده در فاصله مورد نظر استفاده می‌کند.

توزیع نرمال $n(x;\ \mu, \sigma)$

میانگین، میانه و مُد در توزیع نرمال در یک نقطه قرار دارند



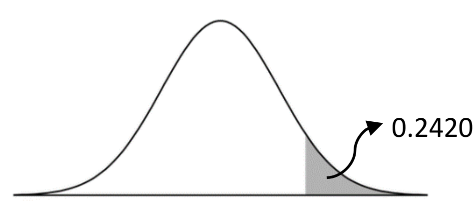
مثال: تولید کننده لاستیک اتومبیل، لاستیک جدید طراحی نموده است. پیش از ارسال در داخل کارخانه مورد تست قرار گرفته و اطلاعات ذیل بدست آمده است.

$$\mu = 36500\ km$$
$$\sigma = 5000\ \ km$$

الف) چند درصد از لاستیک‌ها با عمری بالاتر از ۴۰,۰۰۰ کیلومتر کار خواهند نمود.

حل:

$$Z = \frac{x-\mu}{\sigma} = \frac{40'000 - 36'500}{5'000} = 0.7$$
$$P(0.7) = 0.2420$$



۲۴,۲٪ از لاستیک‌ها عمری بالاتر از ۴۰,۰۰۰ کیلومتر دارند.

ب) به چه میزان گارانتی کند که ۱۰٪ از لاستیک ها برگشت بخورد.

حل:

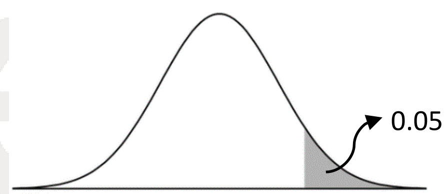
$$P\left(\frac{x-\mu}{\sigma}\right) = 0.1 \quad \Rightarrow \quad x = 5'000 \times (-1.28) + 36'500 = 30'100\ km$$

نکته: واحد انحراف معیار، میانگین و مشاهدات همگی یکسان می‌باشند.

مثال: میانگین نمره آمار دارای توزیع نرمال با مولفه‌های زیر می‌باشد. حداقل نمره‌ای که یک دانشجو می‌تواند بگیرد و جزو ۵٪ نمره بالایی کلاس باشد چقدر است؟

$$\mu = 14$$
$$\sigma = 2$$

$$P(0.05) = 1.64$$
$$1.64 = \frac{x-14}{2} \Rightarrow x = 17.28$$



۱۷,۲۸ حداقل نمره ای است که یک دانشجو می تواند بگیرد تا جزو ۵٪ بالایی کلاس باشد.

توزیع نمونه گیری (Sampling distribution)

مثال: جامعه چهار نفری تایپیست و تعداد خطای آنها بدین شرح می‌باشد.

برای این جامعه داریم

جامعه	تعداد خطا
A	3
B	2
C	1
D	4

$$\mu = 2.5$$
$$\sigma = 1.12$$

نمونه های دو تایی با جایگزاری اخذ می‌شود:

$$AA \quad 3,3 \quad \bar{x}_1 = 3$$
$$AB \quad 3,2 \quad \bar{x}_1 = 2.5$$
$$\dots$$
$$DD \quad 4,4 \quad \bar{x}_1 = 4$$

داریم:

$$\bar{\bar{x}} = 2.5$$

میانگین نمونه برابر با میانگین جامعه می‌باشد.

$$\mu_{\bar{x}} = \mu \quad , \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

قضیه حد مرکزی:

مجموع و میانگین مقادیر یک نمونه تصادفی nتایی از یک جامعه آماری ناشناخته، به یک توزیع نمونه گیری قرینه گرایش دارد.

اگر توزیع جامعه ناشناخته باشد و حجم نمونه نیز کم باشد (کوچکتر از ۳۰) از آماره نا پارامتری استفاده می‌کنیم.

تخمین ها (Estimation)

تخمین کمی – انحراف معیار جامعه موجود باشد: توزیع نرمال

تخمین کمی – انحراف معیار جامعه موجود نیست: توزیع t

تخمین کیفی : توزیع نرمال

مثال: مدیر یک شرکت انتظار دارد که میانگین طول کاغذهای تولیدی "11 باشد. یک نمونه از 100 برگ کاغذ با میانگین "10.998 و انحراف معیار "0.02 را نشان می‌دهد. با 95% اطمینان میانگین جامعه را تخمین بزنید.

$n = 100$
$\bar{x} = 10.998"$
$\sigma = 0.02"$
$\alpha = 95\%$

$$\bar{x} \pm Z\alpha_{/2}\frac{\sigma}{\sqrt{n}}$$

$$10.998 \pm 1.96\frac{0.02}{\sqrt{100}} \Rightarrow 10.99408 \le \mu \le 11.00192$$

انتظار مدیر برآورده می‌گردد.

مثال فوق با 99% اطمینان:

$Z\alpha_{/2} = 2.57$
$10.99284 \le \mu \le 11.00316$

تعیین فاصله اطمینان زمانی که انحراف معیار جامعه مشخص نباشد:

$$\bar{x} \pm t\alpha_{/2}\frac{S}{\sqrt{n}}$$

مثال: یک نمونه ۱۰۰تایی از حقوق هفتگی کارگران، میانگین $ 110.27 و انحراف معیار $ 28.95 را نشان می‌دهد. با ضریب اطمینان 95% میانگین جامعه را تخمین بزنید.

$$n = 100$$
$$\bar{x} = 110.27\ \$$$
$$s = 28.95\ \$$$
$$\alpha = 95\%$$

$$110.27 \pm 1.9842 \frac{28.95}{\sqrt{100}} \Rightarrow 104.53 \le \mu \le 116.01$$

نکته: در مساله اگر در ابتدا انحراف معیار بیان گردد، مربوط به جامعه بوده ولی اگر پس از نمونه گیری در خصوص انحراف معیار صحبت شود، مربوط به نمونه می باشد.

درجه آزادی (df) بدین معنی است که تنها یکی از نمونه ها تحت کنترل نبوده ولی مابقی تصادفی می‌باشند.

نکته: $\bar{x}$ همواره تخمین نقطه ای می‌باشد.

تخمین صفات کیفی:

P: نسبت جامعه (نامشخص می باشد)

$P_s = \frac{x}{n}$ نسبت نمونه (تخمین نقطه ای)

$$P_s \pm Z_{\alpha/2} \sqrt{\frac{P_s(1-P_s)}{n}}$$

نکته اگر نمونه بزرگ نباشد و شرایط ذیل نیز بر قرار نباشد باید از توزیع دو جمله ای، احتمال را محاسبه نمود.

$$\begin{cases} np \ge 5 \\ n(1-p) \ge 5 \end{cases}$$

مثال: مدیر یک روزنامه در نظر دارد که تخمین بزند درصد روزنامه‌های چاپ شده در چند درصد موارد دارای مشکلاتی می‌باشد. یک نمونه ۲۰۰ تایی از روزنامه‌ها بطور تصادفی انتخاب و ۳۵ روزنامه دارای مشکلاتی می‌باشند. با ۹۰٪ فاصله اطمینان نسبت روزنامه های چاپ شده که دارای مشکل می باشد را تخمین بزنید.

$n = 200$
$x = 35$
$P_s = \frac{35}{200} = 0.175$
$\alpha = 90\%$

$$0.175 \pm 1.645\sqrt{\frac{0.175(1-0.175)}{200}} \Rightarrow\ 0.1308\ \leq\ P\ \leq 0.2192$$