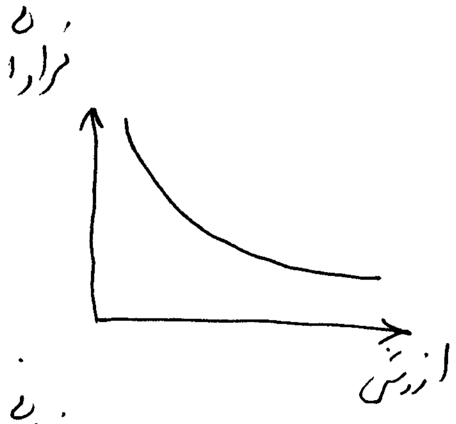
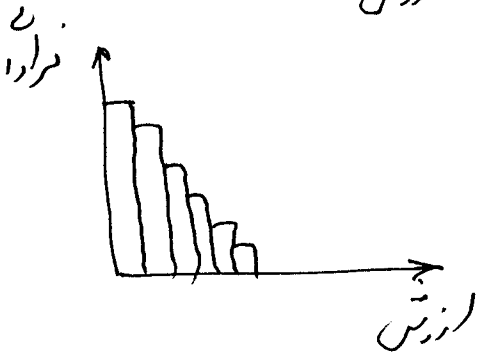
انواع توزیع فراوانی:

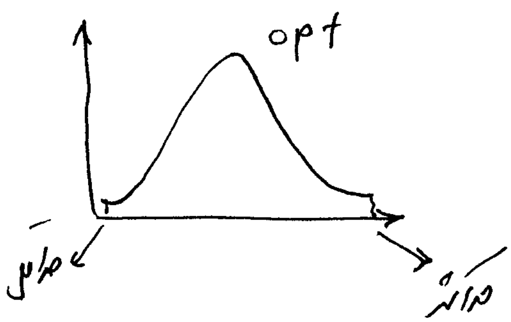


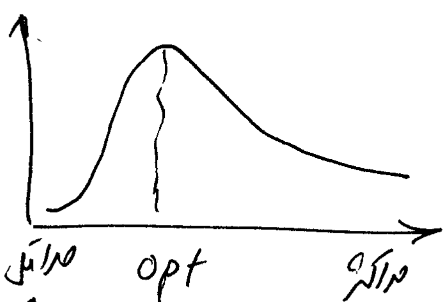


① توزیع گاما (γ): γ-distribution

شبیه یک تابع نمایی است. افرادی که ارزش آنها زیاد است، تعدادکم و افرادی که ارزش آنها کم است تعدادشان زیاد است

کاربرد: حشره، علوم اجتماعی، هواشناسی

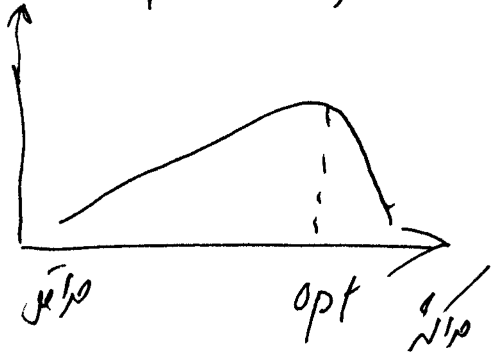




② توزیع بتا (β): β-distribution

برای توابعی که دارای یک نقطه حداقل، یک نقطه حداکثر و یک نقطه بهینه هستند

مثل: رشد باکتری‌ها در دماهای مختلف



③ توزیع دوجمله‌ای: Binomial Distribution

حوادثی که احتمال بروز پدیده دو حالت دارد:

p: موفقیت ، q: شکست

$p+q=1$

مثل باکتری: زنده یا مرده   محصول غذایی: سالم یا معیوب

پرتاب سکه: شیر یا خط   تولد: پسر یا دختر

ویژگی این توزیع: اگر یک پدیده n بار تکرار شود، در حالت بعد احتمال موفقیت و شکست مستقل از دفعات قبلی است.

$X \sim b(\bar{X}, \sigma)$

$\bar{X} = np$

$\sigma = \sqrt{npq}$

n: تعداد دفعات تکرار

p: احتمال موفقیت

q: احتمال شکست

مثال: فرض کنید سکه سالمی را ۱۰۰ بار پرتاب کنید

$n = 100 \;,\; p = \frac{1}{2} \;,\; q = \frac{1}{2}$

$\bar{X} = 100 \times \frac{1}{2} = 50$

$\sigma = \sqrt{100 \times \frac{1}{2} \times \frac{1}{2}} = 5$

$X \sim b(50, 5)$

X دارای توزیع دوجمله‌ای با

$\bar{X} = 50 \;,\; \sigma = 5$

④ توزیع پواسن Poisson Distribution

حالت خاصی از توزیع دوجمله‌ای است

$\bar{X} = np \;,\; \sigma = \sqrt{np}$

$X \sim p(np, \sqrt{np})$

زمانی اهمیت است که $np \leq 5$

مثال: احتمال وقوع یک بیماری در یک جمعیت (جامعه) ۵ در هزار است. از این جامعه نمونه‌ای با $n = ۳۶۰$ انتخاب شده است. احتمال وجود ۴ بیمار در این جامعه چقدر است؟

$$(p+q)^n \qquad p = 0.005$$

$$q = 0.995$$

$$n = 360$$

$$\bar{X} = np = 360 \times 0.005 = 1.8 \leq 5$$

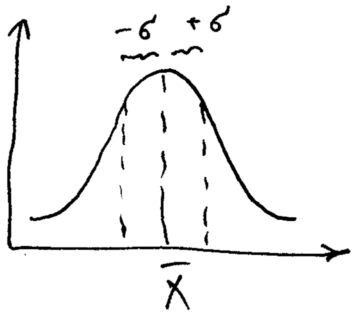
$$p(x) = \frac{e^{-\bar{x}} \cdot \bar{x}^{x}}{x!}$$

$$p(4) = \frac{2.72^{-1.8} \times 1.8^{4}}{4!} = ?$$

استفاده از جدول تجمعی پواسن

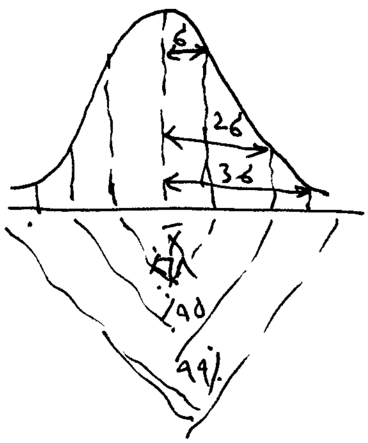
۵) توزیع نرمال Normal Distribution

مهمترین توزیع آماری، ۸۵٪ توابع، ۸۰٪ علم آمار



ویژگی‌های توزیع نرمال

- زنگوله‌ای شکل و سطح زیر منحنی = ۱
- متقارن در یک محور تقارن
- میانگین، میانه و نما بر هم منطبقند
- دارای دو نقطه عطف در فاصله $\bar{X} \pm \sigma$



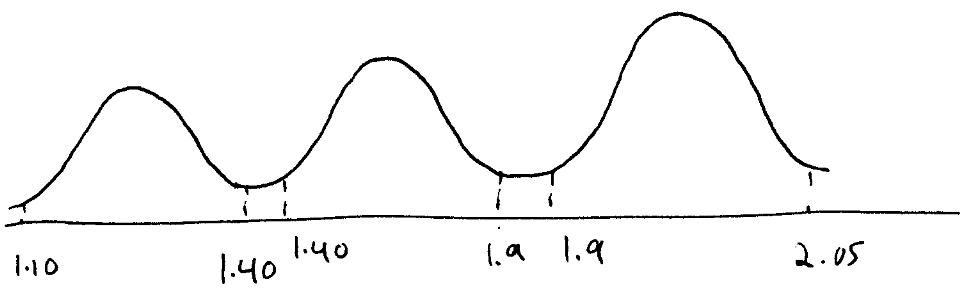
- ۶۸٪ افراد در محدوده $\bar{X} \pm \sigma$
- ۹۵٪ افراد در محدوده $\bar{X} \pm 2\sigma$
- ۹۹٪ افراد در محدوده $\bar{X} \pm 3\sigma$

$$\left.\begin{array}{l} \bar{X} \pm \sigma = 68\% \\ \bar{X} \pm 2\sigma = 95\% \end{array}\right) 27\%$$

$$\left.\begin{array}{l} \bar{X} \pm 2\sigma = 95\% \\ \bar{X} \pm 3\sigma = 99\% \end{array}\right) 4\%$$

غیرخطی

- همواره ~~[illegible]~~ بصورت مجانب با محور xها و تماسی با آن ندارد، نمودار توزیع نرمال پیوسته



$X \sim N(\bar{X}, \sigma) \longrightarrow X \sim N(5, 3)$

سه جامعه متفاوت

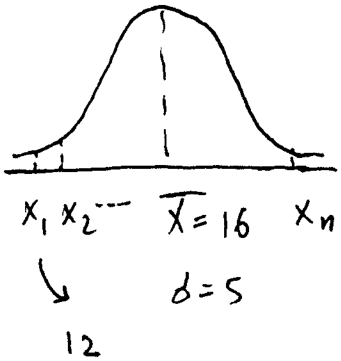
$$\begin{cases} X \sim N(15, 3) \\ X \sim N(16, 5) \\ X \sim N(14, 2) \end{cases}$$

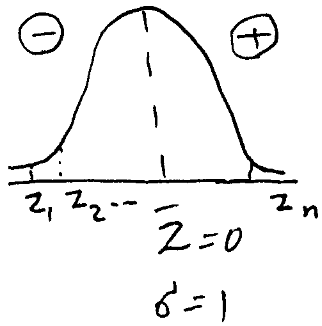
معرفی توزیع Z : توزیع نرمال استاندارد

$$Z \sim N(0, 1)$$

$$Z = \frac{X - \bar{X}}{\sigma}$$

تک‌تک افراد با این رابطه استاندارد می‌شوند





$$Z_1 = \frac{12-16}{5} = -0.8$$

کاربرد توزیع Z

۱- تعیین حدود افراد در نمونه (جامعه)

۲- برآورد میانگین جمعیت از روی میانگین نمونه

۳- تعیین تعداد نمونه لازم برای رسیدن به خطای مشخص

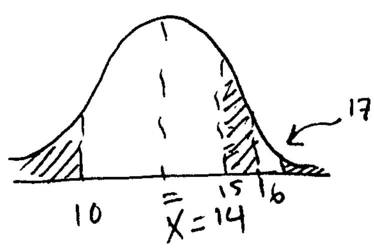
۴- مقایسه میانگین‌ها

① تعیین حدود افراد در جامعه (نمونه)

مثال: در نمونه‌گیری از یک انبار رب گوجه‌فرنگی اطلاعات زیر درباره بریکس به دست آمده است:

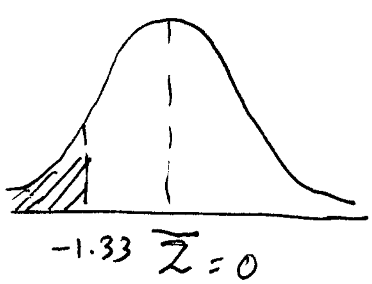
$\bar{X} = 14$

$\sigma = 3$



- چند درصد نمونه‌ها Bx کمتر از ۱۰ دارند
- چند درصد نمونه‌ها Bx بیشتر از ۱۷ دارند
- چند درصد بین ۱۴-۱۵ هستند

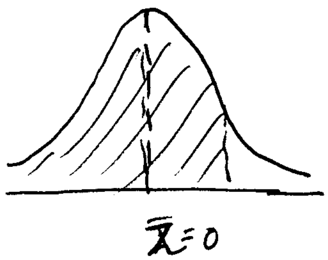
$$x_i \longrightarrow Z = \frac{x_i - \bar{x}}{\sigma} = \frac{10-14}{3} = -1.33$$



محاسبه سطح زیر منحنی از ∞- تا 10

برای محاسبه از جدول Z استفاده می‌شود

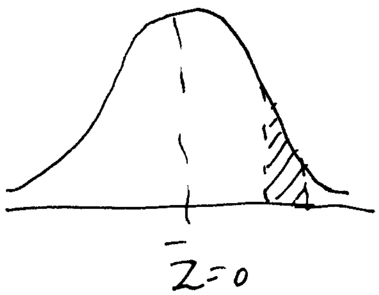
$$Z_{-1.33} = 0.0918 \longrightarrow 9.18\%$$



$$x_i \longrightarrow Z = \frac{x_i - \bar{x}}{\sigma} = \frac{17-14}{3} = 1$$

جدول $Z_1 = 84.13\%$

$100 - 84.13 = 15.87\%$ $\qquad BX > 17$



$$Z = \frac{x_i - \bar{x}}{\sigma} = \frac{15-14}{3} = 0.33 \rightarrow 62.93\%$$

$$Z = \frac{x_i - \bar{x}}{\sigma} = \frac{16-14}{3} = 0.66 \rightarrow 74.54\%$$

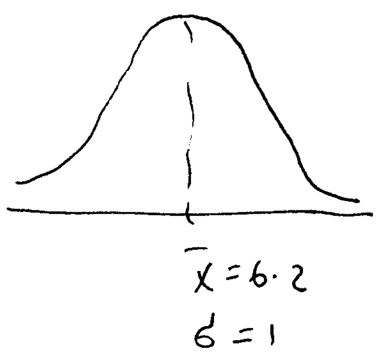
$15 < BX < 16$ $\qquad$ اختلاف 11.61%

مثال: از یک انبار حاوی تعدادی برنج قرار است بخشی به طول کمتر از 6^{mm} دارند جدا شوند، چند درصد از محصول باید جدا شود

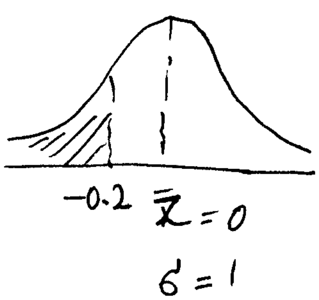
$n = 50$

$\bar{x} = 6.2$

$\sigma = 1$



$$Z = \frac{6-6.2}{1} = -0.2$$



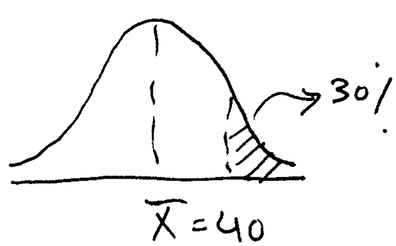
$$Z_{-0.2} = 42.7\%$$

مثال: از یک گله گوسفند میانگین وزن و انحراف معیار بصورت زیر است:

چنانچه دامدار بخواهد ۳۰٪ چاق ترین گوسفندها را بفروشد، از چند کیلو به بالا بفروشد

$\bar{x} = 40\ Kg$

$\sigma = 5\ Kg$



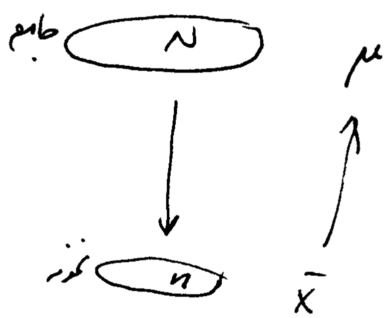
$Z_{70\%} = 0.53$

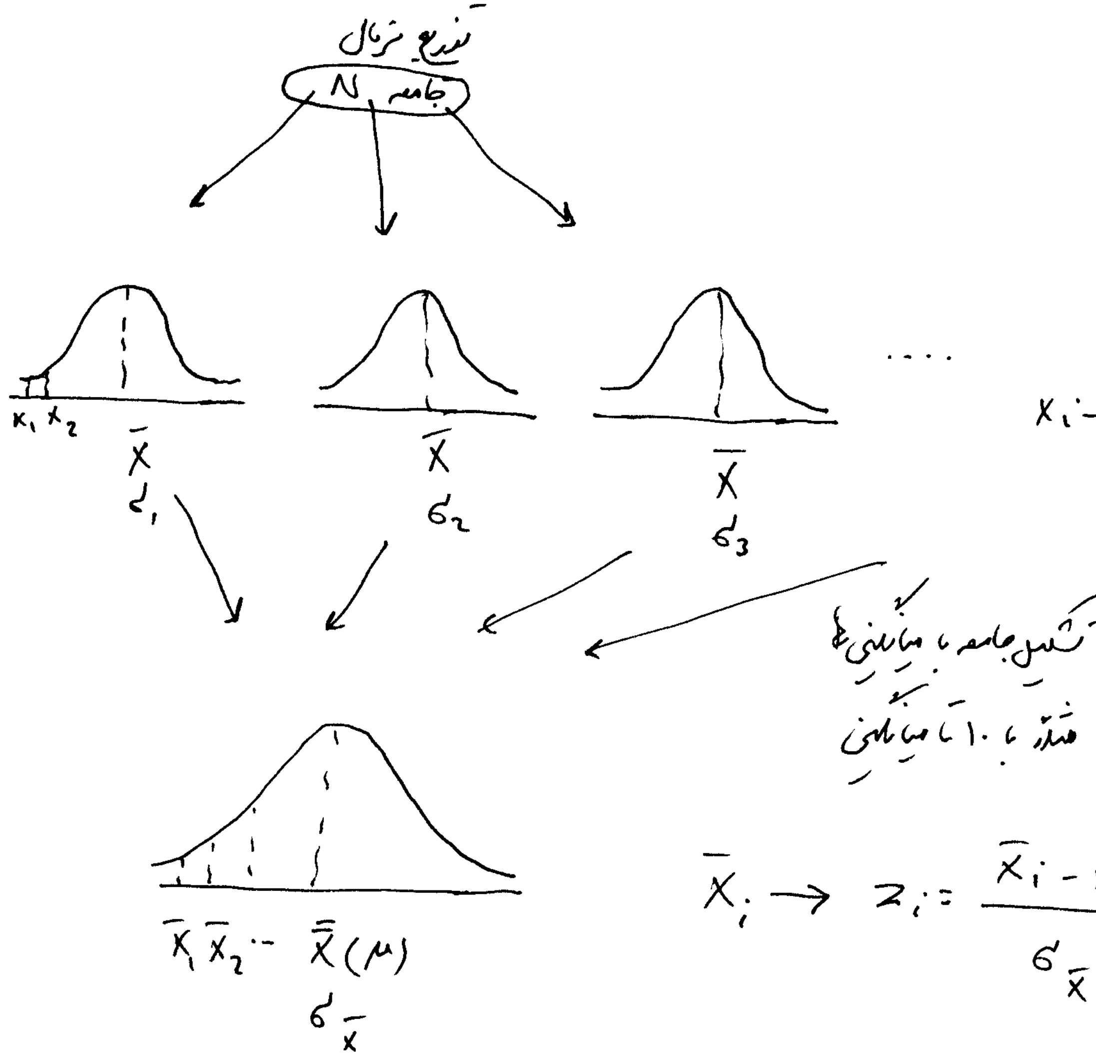
$$Z = \frac{x_i - \bar{x}}{\sigma}$$

$$0.53 = \frac{x_i - 40}{5}$$

$x_i =$ [illegible]

۲) برآورد میانگین جامعه از روی میانگین نمونه





$$x_i \rightarrow z_i = \frac{x_i - \bar{x}}{\sigma}$$

استاندارد کردن x_i

$$\bar{X}_i \rightarrow z_i = \frac{\bar{X}_i - \bar{\bar{x}}}{\sigma_{\bar{x}}}$$

استاندارد کردن $\bar{x}_i$

$$z_i = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

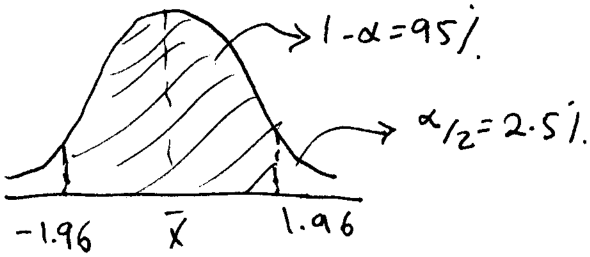
$\bar{\bar{x}}$ تقریب بهتری برای μ است تا $\bar{x}$

$$\sigma_x = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma (\bar{x}_i - \bar{\bar{x}})^2}{n-1}} \rightarrow \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \longrightarrow \bar{x} - \mu = Z \cdot \sigma/\sqrt{n} \longrightarrow \mu = \bar{x} \pm Z \cdot \sigma/\sqrt{n}$$

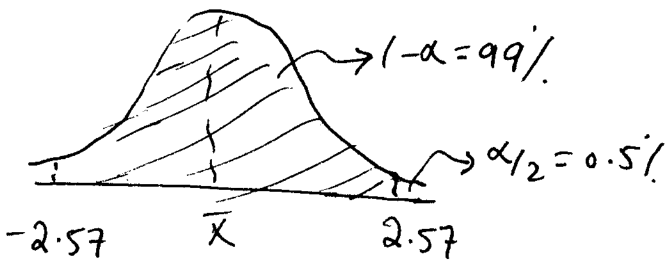
- مفهوم سطح اطمینان ، فاصله اطمینان یا حدود اطمینان



$\alpha = 5\% \longleftarrow 95\%$

$\alpha = 1\% \longleftarrow 99\%$

$$-1.96 < Z < 1.96$$



$$-2.57 < Z < 2.57$$

$$\mu = \bar{x} \pm Z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

مثال : کارخانه ای ادعا دارد پس از انتخاب یک نمونه ۳۶ تایی از یک محصول اطلاعات زیر را بدست آورده است. میانگین جامعه با سطح اطمینان ۹۵, ۹۹٪ چقدر می باشد؟

$\bar{x} = 2.6$

$\sigma = 0.3$

$n = 36$

$$1-\alpha = 95\% \longrightarrow \alpha = 5\% \;,\; \alpha/2 = 2.5\%$$

$$\mu = \bar{x} \pm Z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

$$\mu = 2.6 \pm 1.96 \times \frac{0.3}{6}$$

$$\mu = 2.6 \pm 0.1$$

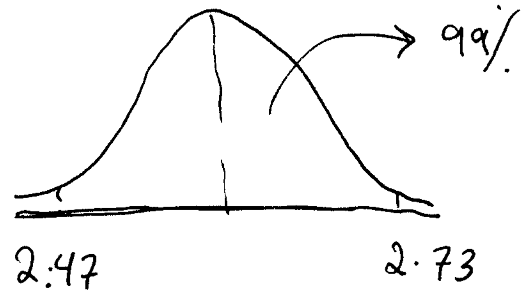
$$2.5 < \mu < 2.7$$

$$\mu = \bar{x} \pm 2.57 \times \frac{\sigma}{\sqrt{n}}$$

$$\mu = 2.6 \pm 2.57 \times \frac{0.3}{\sqrt{36}}$$

$$\mu = 2.6 \pm 0.128$$

$$2.47 < \mu < 2.73$$



با اطمینان ۹۹٪ میانگین جامعه در محدوده فوق قرار دارد.

③ برآورد تعداد نمونه لازم برای رسیدن به خطای مشخص

$$n \longrightarrow N$$

$$\bar{x} \longrightarrow \mu$$

$$\bar{x} - \mu \longrightarrow 0$$

$$\bar{x} - \mu = \text{error}$$

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

$$Z = \frac{e}{\sigma / \sqrt{n}} \longrightarrow e = Z \cdot \frac{\sigma}{\sqrt{n}}$$

$$e^2 = \frac{Z^2 \cdot \sigma^2}{n} \longrightarrow n = \frac{Z^2 \cdot \sigma^2}{e^2}$$

خطای برآورد

مثال: در یک کارخانه تولید غذای کودک قرار است جهت اندازه‌گیری ترکیبات نمونه‌برداری شود. با فرض براینکه مقدار $e = 0.5$ باشد، با حدود اطمینان ۹۵٪ چند نمونه باید انتخاب کرد $(\sigma = 5)$؟

جامعه $\mu = ?$ $\sigma = 5$

نمونه $n = ?$

$$\mu = \bar{x} \pm 0.5$$

$$Z_{95\%} = 1.96 \qquad Z_{99\%} = 2.57$$

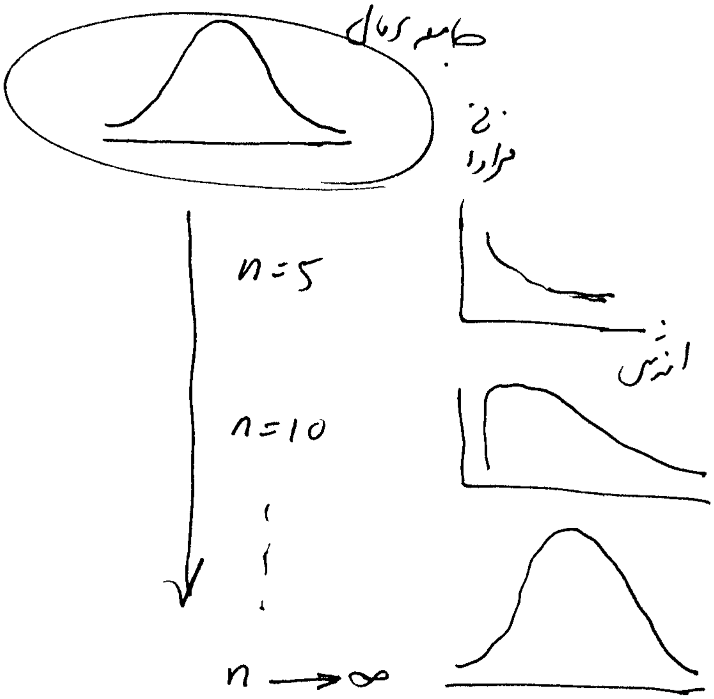
$$n = \frac{Z^2 \cdot \sigma^2}{e^2} = \frac{(1.96)^2 \cdot (5)^2}{(0.5)^2} = 384.16$$

حال با فرض اینکه $e=1$ باشد چند نمونه باید برداشت

$$n = \frac{Z^2 \cdot \sigma^2}{e^2} = \frac{(1.96)^2 (5)^2}{(1)^2} = 96.04$$

عدم ارتباط خطی e با تعداد نمونه

<u>قضیه حد مرکزی</u>: در نمونه‌های حاصل از جوامع نرمال، با هر نوع توزیع با افزایش تعداد نمونه به سمت بی‌نهایت، توزیع نمونه به سمت نرمال میل خواهد کرد



نمونه بزرگ : $n \geq 30$

نمونه کوچک : $n < 30$

<u>توزیع t یا t-student</u>

هزینه بر بودن، وقت گیر بودن $\}$ ← بسیاری از نمونه‌ها کوچک هستند

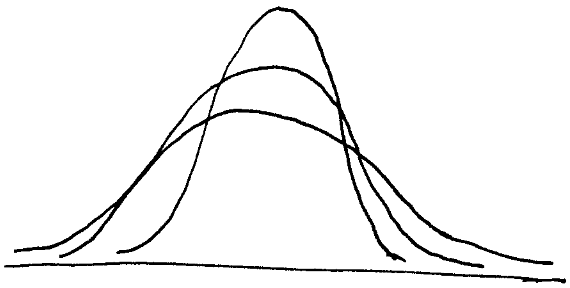
$Z \sim N(0,1)$

$t \sim N(0, \sigma > 1)$

$n \to \infty \Rightarrow \sigma_t \to 1$

$\Rightarrow$ توزیع t $\to$ توزیع Z

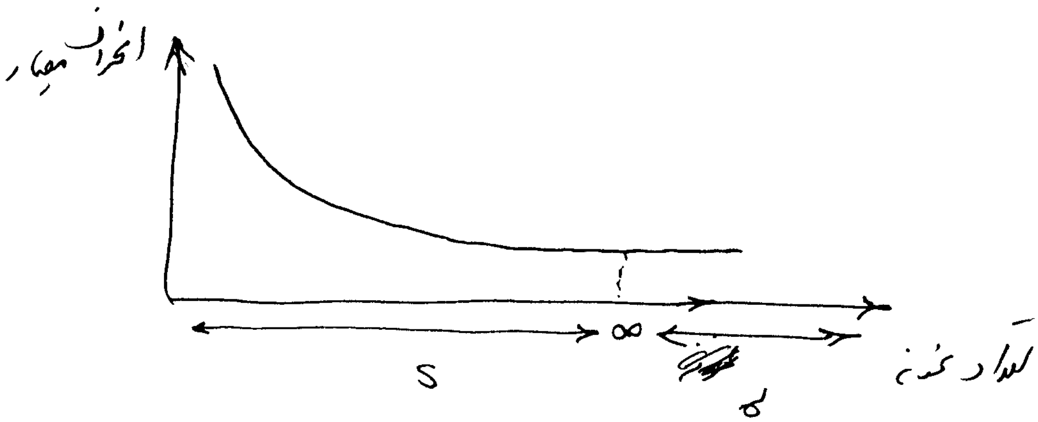
با افزایش تعداد نمونه σ کاهش می‌یابد.



σ انحراف معیار ← $n > 30$

S انحراف معیار ← $n < 30$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \qquad S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$



$n-1$ ← درجه آزادی Degree of Freedom (df)

کاربردهای توزیع t

① تعیین صدور افراد

$$x_i \longrightarrow z = \frac{x_i - \bar{x}}{\sigma}$$

انحراف معیار جامعه

$$x_i \longrightarrow t = \frac{x_i - \bar{x}}{s}$$

انحراف معیار نمونه

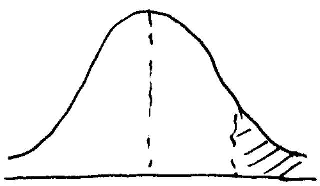
④ برآورد μ از روی $\bar{X}$:

$$\mu = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

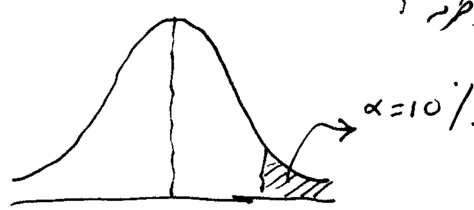
$$\mu = \bar{X} \pm t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

نحوه استفاده از جدول t :

استفاده یک طرفه و دوطرفه :



t_α

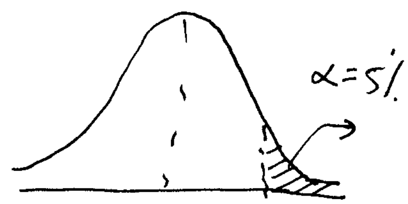


$t_\alpha = 1.383$

$n=10$

$df=9$

اطمینان = 90%



$t = 1.796$

$n=12$

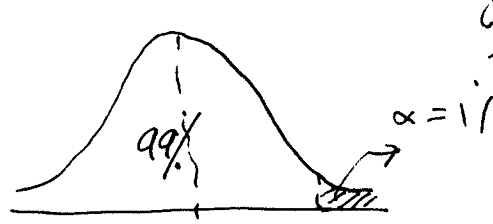
$df=9$

اطمینان = 95%

-1.833 1.833

$n=10$

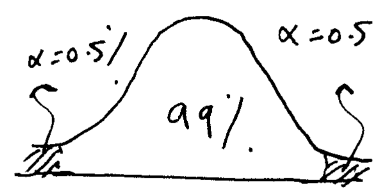
$df=9$



2.624

$n=15$

$df=14$



2.977 2.977

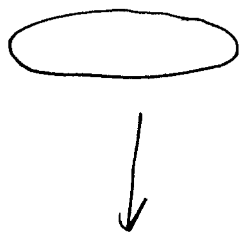
$n=15$

$df=14$

$n = 31 \longrightarrow \begin{cases} \text{سطح اطمینان 95\% دوطرفه} \\ \text{سطح اطمینان 99\% دوطرفه} \end{cases}$

$\longrightarrow \begin{cases} n \longrightarrow \infty \\ t \longrightarrow Z \end{cases}$

مثال ٤: جهت اندازه‌گیری pH یک ماده غذایی ۷ عدد نمونه تهیه شده است به طوری که مقدار μ و σ مشخص نیست. μ را برای این محصول با حدود اطمینان ۹۵٪ تعیین کنید؟



$\mu = ?$

$\sigma = ?$

$n = 7$

$x_1 = 9.8$

$x_2 = 10.2$

$x_3 = 10.4$

$x_4 = 9.8$

$x_5 = 10.0$

$x_6 = 10.2$

$x_7 = 9.6$

$n = 7$

$\bar{x} = 10$

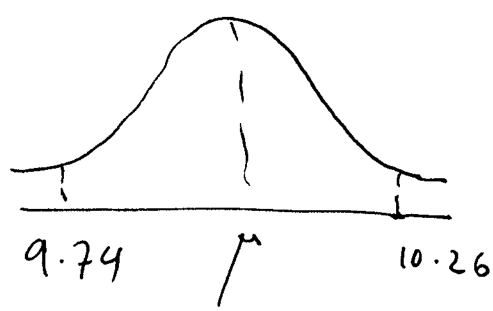
$$\mu = \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = 0.283$$

$$t \begin{cases} n = 7,\ df = 6 \\ \alpha/2 = 2.5\% \end{cases} \longrightarrow 2.447$$

$$\mu = 10 \pm 2.447 \times \frac{0.283}{\sqrt{7}}$$

$$9.74 < \mu < 10.26$$

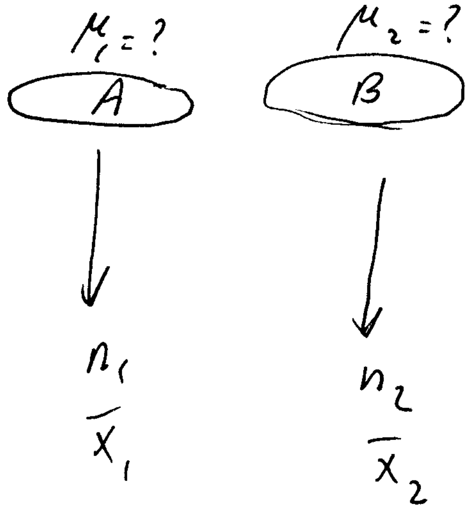


③ تعیین تعداد نمونه برای رسیدن به خطای مشخص

$$n = \frac{Z^2 \sigma^2}{e^2} \longrightarrow n = \frac{t^2 s^2}{e^2}$$

④ مقایسه دو میانگین با آزمون فرض‌ها

می‌خواهیم بدانیم اختلاف بین میانگین‌های دو جامعه حائز اهمیت است یا خیر



فرضیات:

فرض صفر $H_0: \mu_1=\mu_2 \rightarrow \mu_1-\mu_2=0$

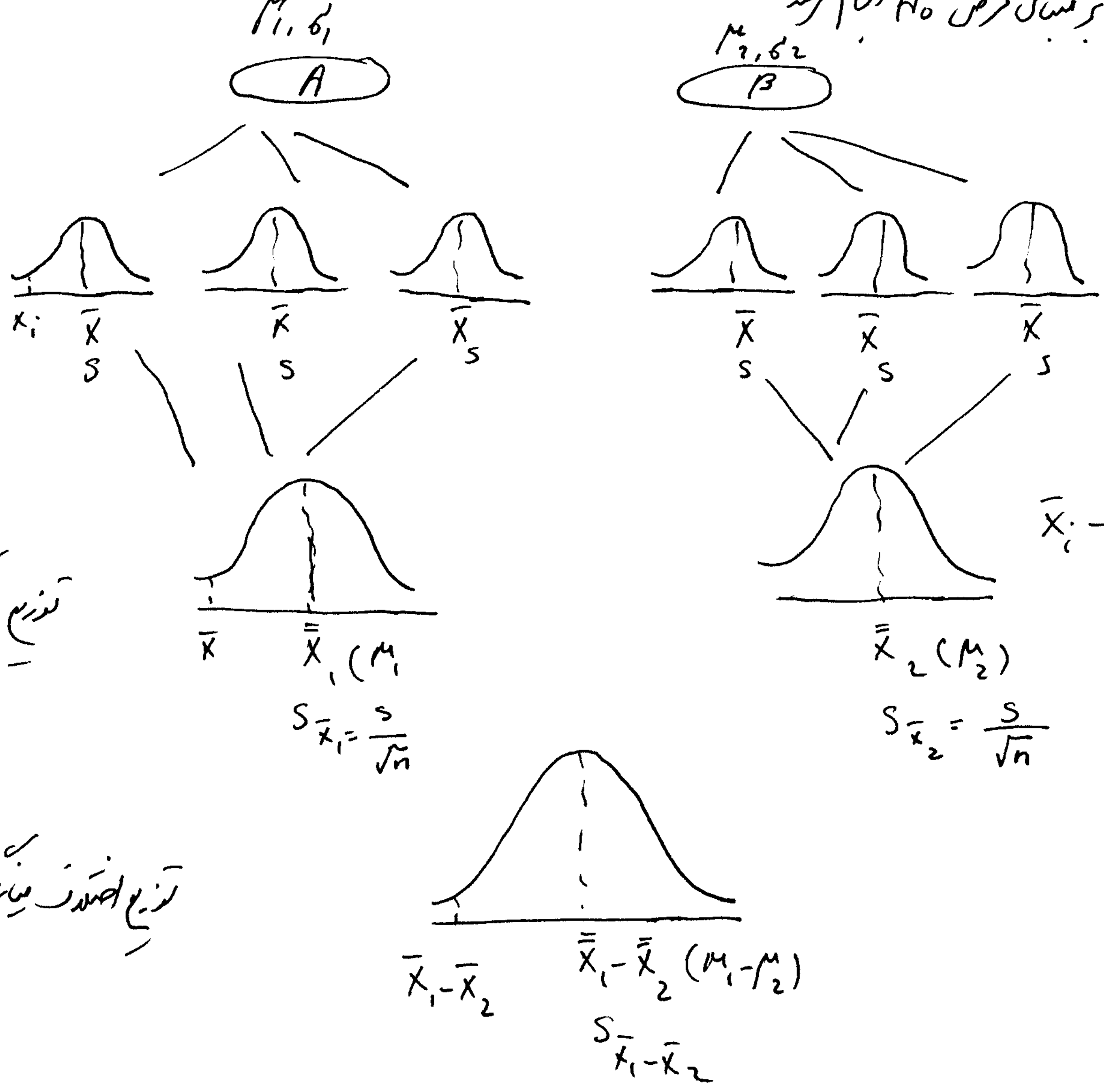
فرض مقابل $H_1: \mu_1\neq\mu_2 \rightarrow \mu_1-\mu_2\neq 0$

محاسبات بر مبنای فرض H_0 انجام می‌شود

توزیع نمرات

توزیع میانگین‌ها

توزیع اختلاف میانگین‌ها



$$x_i \rightarrow t=\frac{x_i-\bar{x}}{S}$$

$$\bar{X}_i \rightarrow t=\frac{\bar{X}_1-\mu}{S/\sqrt{n}}$$

$$\bar{X}_1-\bar{X}_2 \rightarrow t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_{\bar{X}_1-\bar{X}_2}}$$

با در نظر گرفتن فرض $H_0: \mu_1 = \mu_2$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \overset{0}{\cancel{(\mu_1 - \mu_2)}}}{S_{\bar{X}_1 - \bar{X}_2}}$$

$$S_{\bar{X}_1} = \frac{S_1}{\sqrt{n_1}} \qquad S_{\bar{X}_2} = \frac{S_2}{\sqrt{n_2}}$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\Sigma\left[(\bar{X}_1 - \bar{X}_2) - (\bar{\bar{X}}_1 - \bar{\bar{X}}_2)\right]^2}{n-1}}$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

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

مثال: مقایسه تاثیر یک غذای کودک با فرمولاسیون جدید بر رشد کودکان نسبت به یک غذای کودک قدیمی:

۲۰ کودک انتخاب می‌شوند، ۱۰ تا غذای کودک جدید و ۱۰ تا غذای کودک قدیم =

افزایش وزن (A)	افزایش وزن (B)
x_1	x_1
x_2	x_2
x_3	x_3
⋮	⋮
x_{10}	x_{10}
$\bar{X}_1 = 0.5$	$\bar{X}_2 = 0.3$

$\bar{X}_1 - \bar{X}_2 = 0.2$

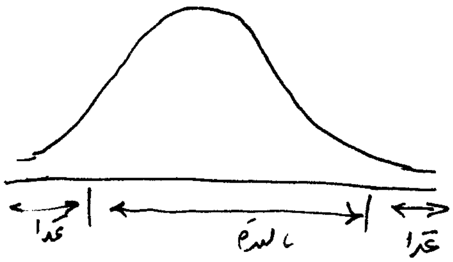
$H_0: \mu_1 = \mu_2$

اختلاف ناشی از چیست؟

۱- نوع غذا

۲- با مجموعه ← ژنتیک و ...

راه حل ← ؟



اهمیت تجزیه واریانس ← ؟

مقایسه میانگین ← نمونه‌های جفت شده ← یک واریانس

مقایسه میانگین ← نمونه‌های جفت نشده ← تنوع واریانس

نمونه‌های جفت شده :

مثال : در کارخانه‌ای برای ایجاد رنگ از دو روش A و B استفاده می‌شود. می‌خواهیم بدانیم تاثیر کدام روش بر ماندگاری رنگ بیشتر است؟

جفت	روش A	روش B	$X_1 - X_2 = D$
1	2.0	2.2	-0.2
2	2.0	1.9	+0.1
3	2.3	2.5	-0.1
4	2.1	2.3	-0.1
5	2.4	2.4	0.0
	$\bar{X}_A = 2.16$	$\bar{X}_B = 2.26$	$\bar{D} = \frac{\Sigma d}{n} = -0.1$

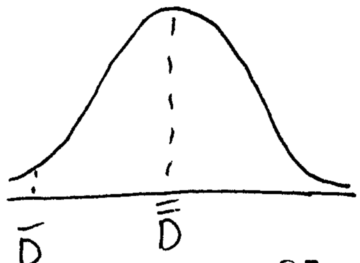
$X_1 - X_2 = D$ → Difference

چگونه نمونه جفت شده تهیه می‌شود؟

$\bar{X}_A \rightarrow \mu_A \qquad \bar{X}_B \rightarrow \mu_B \qquad \bar{D} \rightarrow \mu_1 - \mu_2 = D$

$$\begin{cases} H_0: \mu_1 = \mu_2 \longrightarrow \mu_1 - \mu_2 = 0 \longrightarrow \mu_D = 0 \\ H_1: \mu_1 \neq \mu_2 \longrightarrow \mu_1 - \mu_2 \neq 0 \longrightarrow \mu_D \neq 0 \end{cases}$$

آزمون برای فرض H_0 نوشته می‌شود



$s_{\bar{D}} = \frac{S_D}{\sqrt{n}}$

$\bar{D} \rightarrow \bar{x}_1 - \bar{x}_2$

$\bar{\bar{D}} \rightarrow \bar{\bar{x}}_1 - \bar{\bar{x}}_2 \; (\mu_1 - \mu_2)$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \cancel{(\mu_1 - \mu_2)}}{S_{\bar{x}_1 - \bar{x}_2}}$$

$$\bar{D} \longrightarrow t = \frac{\bar{D} - \mu_D}{S_{\bar{D}}}$$

$$S_D = \sqrt{\frac{\sum D_i^2 - \frac{(\sum D_i)^2}{n}}{n-1}}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}}$$

$$S_D = \sqrt{\frac{(0.2)^2 + (0.1)^2 + \cdots - \frac{(-0.5)^2}{5}}{5-1}} = 0.14142$$

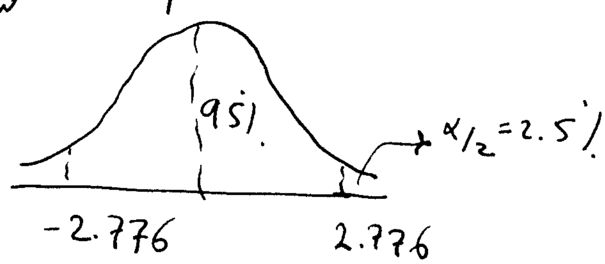
$$\bar{D} \longrightarrow \hat{t} = \frac{\bar{D} - \mu_D}{S_{\bar{D}}} = \frac{-0.1 - 0}{0.14142 / \sqrt{5}} = -1.6$$

$\hat{t} < t_{\text{جدول}} \rightarrow H_0$: قبول

$\hat{t} > t_{\text{جدول}} \rightarrow H_0$: رد

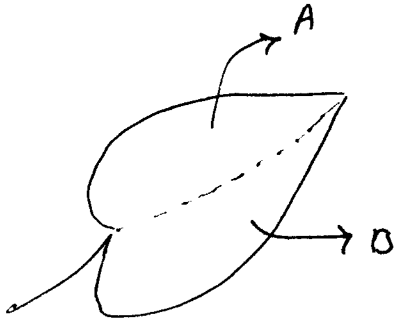
$\longleftarrow$ مقایسه $\hat{t}$ با t جدول

t جدول $\begin{cases} \alpha = 5\% \\ df = 4 \end{cases} \longrightarrow 2.776$



فرض H_0 پذیرفته می‌شود، پس دو روش ایجاد تاثیر معنی‌داری بر ماندگاری مرغ ندارند.

مثال: تاثیر دو ویروس A و B بر آلودگی برگ گیاه:



نحوه تهیه نمونه جفت شده

شماره جفت برگ	ویروس A	ویروس B	$X_1 - X_2 = D$
1	31	18	13
2	20	17	3
3	18	14	4
4	17	11	6
5	9	10	-1
6	8	7	1
7	10	5	5
8	7	6	1
جمع	120	88	$32 = \Sigma D$
میانگین	15	11	$4 = \bar{D} = \bar{x}_1 - \bar{x}_2$

$$15 \rightarrow \mu_1 \qquad 11 \rightarrow \mu_2 \qquad \bar{D} \rightarrow \mu_D = \mu_1 - \mu_2$$

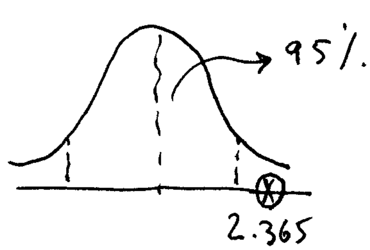
$$H_0: \mu_1 - \mu_2 = 0 \longrightarrow \mu_D = 0$$

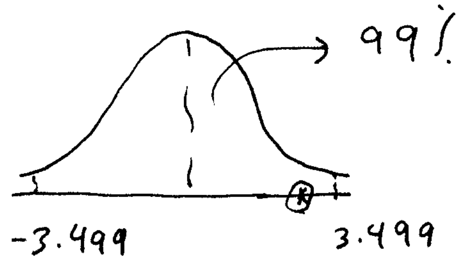
$$\bar{D} \longrightarrow \hat{t} = \frac{\bar{D} - \mu_D}{S_{\bar{D}}} \qquad S_{\bar{D}} \longrightarrow \frac{S_D}{\sqrt{n}}$$

$$\hat{t} = \frac{4 - 0}{4.34/\sqrt{8}} = 2.63$$

جدول t: $\alpha = 5\%$, $df = 7 \longrightarrow 2.365$

جدول t: $\alpha = 1\%$, $df = 7 \longrightarrow 3.499$





$\hat{t} > t_{\text{جدول}} \rightarrow H_0$: رد می‌شود

تاثیر دو ویروس بر آلودگی برگ معنی دار است

$\hat{t} < t_{\text{جدول}} \rightarrow H_0$: قبول

تاثیر دو ویروس بر آلودگی برگ معنی دار نیست