(1)

به نام خدا　　　　خلاصه مبحث برآوردیابی

* جامعه ← جامعه یا جمعیت مجموعه‌ی تمام افراد یا اشیایی است که در مطالعه آماری در مورد یک یا چند صفت آن در یک مکان و زمان معین بررسی صورت می‌گیرد.

✓ هر یک از افراد یا اشیاء را یک عضو جمعیت می‌نامند.

✓ تعداد اعضای جمعیت را اندازه‌ی جمعیت می‌نامند.

* نمونه ← زیرمجموعه‌ای از جمعیت که طبق قاعده و ضوابطی خاص برای مطالعه‌ی صفتی از جمعیت انتخاب می‌شود را یک نمونه می‌نامند.

✓ تعداد اعضای نمونه به اندازه‌ی نمونه موسوم است.

* استنباط آماری ← روشی است که به وسیله‌ی آن براساس نتایج حاصل از نمونه انتخابی از جمعیت، درمورد کل جمعیت یا صفات ناشناخته آن نتیجه‌گیری‌هایی انجام می‌گیرد. دو روش مهم استنباط آماری ← برآورد، آزمون فرض

* پارامتر ← هر ویژگی یک جمعیت را یک پارامتر جمعیت یا به اختصار پارامتر می‌نامند.

* آماره ← ویژگی متناظر با ویژگی‌های جامعه را در نمونه یک آماره می‌نامیم.

* نمونه تصادفی ← متغیرهای تصادفی $X_n, \dots, X_1$ که مستقل و هم‌توزیع با توزیع مشترک $f(x)$ باشند یک نمونه تصادفی می‌نامیم و با نماد $X_1, \dots, X_n \overset{i.i.d.}{\sim} f(x)$ نمایش می‌دهیم.

independent identically distribution (مستقل و هم توزیع)

هدف از فصل برآوردیابی ← تخمین زدن پارامترهای مجهول جامعه به وسیله‌ی آماره‌های از نمونه.

دو نوع روش برآورد وجود دارد:
نقطه‌ای ← دقیقاً یک آماره را تعیین می‌کنیم به عنوان تخمین پارامتر مورد نظر.
فاصله‌ای ← یک بازه را به عنوان حدود تخمین پارامتر تعیین می‌کنیم.

* مهم‌ترین پارامترهای جامعه عبارتند از :

میانگین جامعه ← μ

واریانس جامعه ← σ^2

انحراف معیار ← $\sigma = \sqrt{\sigma^2}$

* مهم‌ترین آماره‌های یک نمونه تصادفی عبارتند از :

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

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

انحراف استاندارد ← $\sqrt{S^2} = S$

نکته مهم $S^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 - n \bar{X}^2 \right]$

روش محاسباتی دیگری برای واریانس نمونه

* برآورد نقطه‌ای : در بین آماره‌های مختلفی که با یک نمونه تصادفی می‌توان ساخت، با استفاده از معیارها و روش‌هایی می‌توان یک برآوردیاب (برآوردگر) برای پارامتر مجهول جامعه پیدا کرد. بعد از مشاهده نمونه و جایگذاری مقادیر مشاهده شده در آن آماره، مقداری به دست می‌آید که به آن برآورد (تخمین) پارامتر مجهول جامعه گویند.

* بهترین برآوردگر برای میانگین جامعه یعنی μ عبارت است از میانگین نمونه یعنی $\bar{X}$.

* 〃 〃 〃 واریانس 〃 〃 σ^2 〃 〃 〃 واریانس 〃 〃 S^2 .

3

* برآورد فاصله‌ای (فاصله اطمینان):

برای پارامتر θ بازه‌ی (L, U) که در آن L و U توابعی از نمونه تصادفی هستند را یک برآورد فاصله‌ای یا فاصله اطمینان $100(1-\alpha)\%$ می‌گویند هرگاه:

$$P(L<\theta<U)=1-\alpha$$

به طور کلی برای یافتن برآورد فاصله‌ای برای هر پارامتری ابتدا یک کمیت محوری (Q) پیدا می‌کنیم. کمیت محوری عبارت است از تابعی از نمونه تصادفی و پارامترهای جامعه که توزیع آن به پارامتر مجهول جامعه وابسته نیست.

سپس از یافتن Q بازه‌ی (a, b) را براساس اینکه $P(a<Q<b)=1-\alpha$ را یافته و به کمک آن به (L, U) برای θ می‌رسیم.

مثلاً فاصله اطمینان برای μ وقتی σ^2 معلوم است (در جامعه‌ی نرمال) به صورت زیر به دست می‌آید:

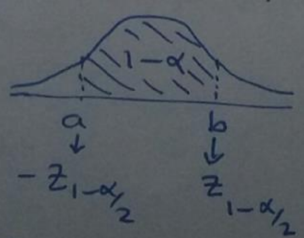
چون μ میانگین جامعه است برای برآورد فاصله‌ای آن از برآورد نقطه‌ای شروع می‌کنیم. می‌دانیم $\bar{X}$ برآوردگر نقطه‌ای μ است و داریم:

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

پس کمیت محوری عبارت است از:

$$Q=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$P\left(a<\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<b\right)=1-\alpha$$



$$\Rightarrow P\left(-z_{1-\frac{\alpha}{2}}<\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<z_{1-\frac{\alpha}{2}}\right)=P\left(\underbrace{\bar{X}-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}_{L}<\mu<\underbrace{\bar{X}+z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}_{U}\right)=1-\alpha$$

در زیر فاصله اطمینان ٪100(1−α) برای پارامترهای مختلف در جامعه نرمال بیان شده است:

* فاصله اطمینان ٪100(1−α) میانگین جامعه‌ی نرمال (μ) وقتی واریانس جامعه (σ^2) معلوم است:

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

* فاصله اطمینان ٪100(1−α) برای میانگین جامعه نرمال (μ) وقتی واریانس جامعه (σ^2) مجهول است:

$$\left(\bar{x} - t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}},\ \bar{x} + t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}}\right)$$

* فاصله اطمینان ٪100(1−α) برای واریانس جامعه‌ی نرمال:

$$\left(\frac{(n-1)s^2}{\chi^{2\,(n-1)}_{1-\frac{\alpha}{2}}},\ \frac{(n-1)s^2}{\chi^{2\,(n-1)}_{\frac{\alpha}{2}}}\right)$$

* فاصله اطمینان ٪100(1−α) برای تفاضل میانگین‌های دو جامعه نرمال ($\mu_1-\mu_2$) وقتی واریانس‌ها معلوم اند:

$$\left(\bar{x}_1-\bar{x}_2 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \bar{x}_1-\bar{x}_2 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

* فاصله اطمینان ٪100(1−α) برای تفاضل میانگین‌های دو جامعه نرمال ($\mu_1-\mu_2$) وقتی واریانس‌ها مجهول اند (ولی برابرند)

$$\left(\bar{x}_1-\bar{x}_2 - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \bar{x}_1-\bar{x}_2 + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

که در آن

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

* فاصله اطمینان برای نسبت واریانس‌های دو جامعه‌ی نرمال ($\frac{\sigma_2^2}{\sigma_1^2}$)

$$\left(\frac{S_2^2}{S_1^2}F_{\frac{\alpha}{2}}(n_1-1, n_2-1),\ \frac{S_2^2}{S_1^2}F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)\right)$$

که در آن

$$F_{\frac{\alpha}{2}}(n_1-1, n_2-1) = \frac{1}{F_{1-\frac{\alpha}{2}}(n_2-1, n_1-1)}$$

به نام خدا

(۱)

خلاصه مبحث آزمون فرض های آماری

* فرض آماری ⟵ ادعایی درمورد یک یا چند جامعه ی مورد بررسی است که ممکن است درست یا نادرست باشد.

* هدف از این مبحث مقایسه ی دو فرض آماری در مقابل یکدیگر است.

↓

یکی فرض صفر یا صحیح نامیده می‌شود و با H_0 نمایش داده می‌شود.

" " مقابل نامیده می‌شود و با H_1 " " " .

* آماره ی آزمون ⟵ تابعی از نمونه تصادفی مانند $T = T(X_1, \dots, X_n)$ که بر اساس آن یک فرض را رد یا قبول می‌کنیم.

* ناحیه ی بحرانی ⟵ مجموعه مقادیری از آماره آزمون که به ازای آن فرض H_0 را رد می‌کنیم، یا اصطلاحاً ناحیه ی رد H_0 را ناحیه بحرانی نامیده و با C نمایش می‌دهیم.

* در تصمیم گیری بین H_0 و H_1 ممکن است دو نوع خطا مرتکب شویم:

✓ خطای نوع اول: رد H_0 وقتی H_0 در واقع درست بوده.

✓ " " دوم: قبول H_0 وقتی در واقع H_1 درست بوده.

احتمال رخداد این خطاها عبارتند از:

$\alpha = P($ رد $H_0 \mid H_0$ درستی $)$

$\beta = P($ قبول $H_0 \mid H_1$ درستی $)$

در مقابل کمیتی به نام توان آزمون داریم که هرچه بیشتر باشد بهتر است:

$\beta^* = P($ رد $H_0 \mid H_1$ درستی $) = 1 - \beta$

* α را سطح معنی داری یا سطح تشخیص آزمون می‌نامند.

* مراحل انجام یک آزمون :

(1) تعیین فرض‌های H_0 و H_1 .

(2) تعیین یک سطح معنی‌داری α که معمولاً آن را 0.01 ، 0.05 و 0.1 می‌گیرند.

(3) تعیین آماره آزمون که معمولاً بر اساس برآوردگر نقطه‌ای پارامتر مجهول تشکیل می‌شود.

(4) تعیین ناحیه بحرانی (C) بر اساس آماره آزمون ، فرض مقابل و سطح معنی‌داری α .

(5) محاسبه‌ی مقدار آماره آزمون از روی مشاهدات نمونه و رسیدن به یک عدد.

(6) نتیجه‌گیری : اگر مقدار عددی حاصل از مرحله‌ی (5) در ناحیه‌ی بحرانی قرار بگیرد آن‌گاه H_0 را رد می‌کنیم و در غیر این صورت فرض H_0 را رد نمی‌کنیم.

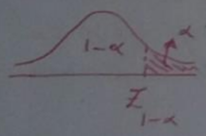
* در زیر آزمون های فرض مختلف روی پارامترهای جامعه نرمال را انجام می دهیم.

(۱) آزمون فرض برای میانگین جامعه نرمال (μ) وقتی واریانس (σ^2) معلوم است (در سطح معنی داری α)

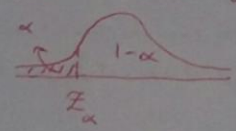
آماره آزمون برای μ در این حالت عبارت است از:

$$H_0 \text{ تحت فرض}: \quad Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

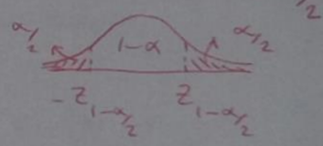
$$\begin{cases} H_0 : \mu = \mu_0 \text{ یا } \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad \downarrow \quad RH_0 \iff Z > Z_{1-\alpha}$$



$$\begin{cases} H_0 : \mu = \mu_0 \text{ یا } \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \downarrow \quad RH_0 \iff Z < Z_{\alpha}$$



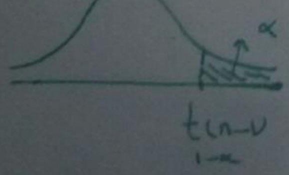
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad \downarrow \quad RH_0 \iff |Z| > Z_{1-\alpha/2}$$



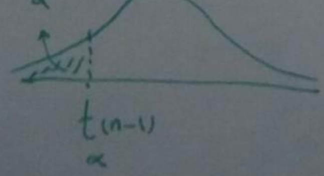
(۲) آزمون فرض برای میانگین جامعه نرمال (μ) وقتی واریانس (σ^2) مجهول است:

$$H_0 \text{ آماره آزمون تحت فرض}: \quad T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$$

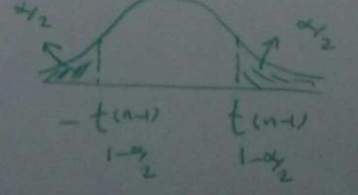
$$\begin{cases} H_0 : \mu = \mu_0 \text{ یا } \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad \downarrow \quad RH_0 \iff T > t_{1-\alpha}^{(n-1)}$$



$$\begin{cases} H_0 : \mu = \mu_0 \text{ یا } \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \downarrow \quad RH_0 \iff T < t_{\alpha}^{(n-1)}$$



$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad \downarrow \quad RH_0 \iff |T| > t_{1-\alpha/2}^{(n-1)}$$

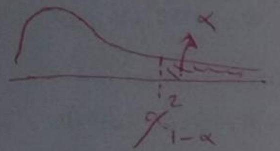


(4)

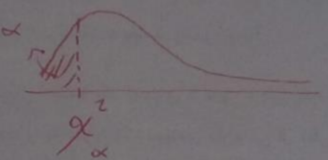
(3) آزمون فرض برای واریانس جامعه‌ی نرمال (σ^2)

آماره‌ی آزمون تحت فرض H_0 :
$$X^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{(n-1)}$$

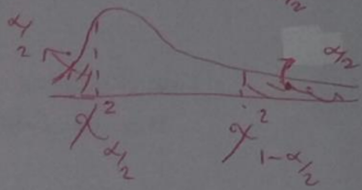
$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \text{ یا } \sigma^2 \le \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$$
$\downarrow$
$RH_0 \Longleftrightarrow X^2 > \chi^2_{1-\alpha\,(n-1)}$



$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \text{ یا } \sigma^2 \ge \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$$
$\downarrow$
$RH_0 \Longleftrightarrow X^2 < \chi^2_{\alpha\,(n-1)}$



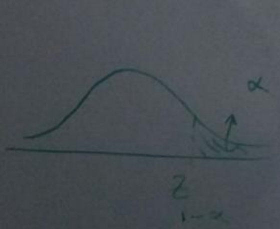
$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \ne \sigma_0^2 \end{cases}$$
$\downarrow$
$RH_0 \Longleftrightarrow X^2 > \chi^2_{1-\alpha/2\,(n-1)}$ or $X^2 < \chi^2_{\alpha/2\,(n-1)}$



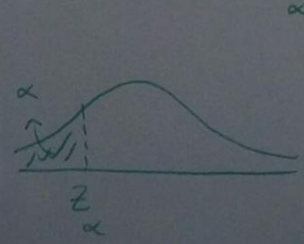
(3) آزمون فرض برای تفاضل میانگین‌های دو جامعه‌ی نرمال ($\mu_1-\mu_2$) وقتی واریانس‌ها معلوم اند:

آماره‌ی آزمون تحت H_0 :
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

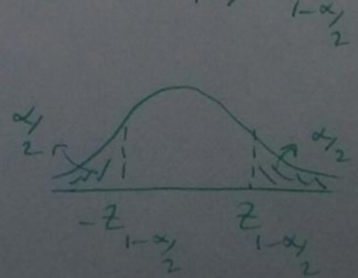
$$\begin{cases} H_0 : \mu_1-\mu_2 = d_0 \text{ یا } \le d_0 \\ H_1 : \mu_1-\mu_2 > d_0 \end{cases}$$
$\downarrow$
$RH_0 \Longleftrightarrow Z > Z_{1-\alpha}$



$$\begin{cases} H_0 : \mu_1-\mu_2 = d_0 \text{ یا } \ge d_0 \\ H_1 : \mu_1-\mu_2 < d_0 \end{cases}$$
$\downarrow$
$RH_0 \Longleftrightarrow Z < Z_{\alpha}$



$$\begin{cases} H_0 : \mu_1-\mu_2 = d_0 \\ H_1 : \mu_1-\mu_2 \ne d_0 \end{cases}$$
$\downarrow$
$RH_0 \Longleftrightarrow |Z| > Z_{1-\alpha/2}$



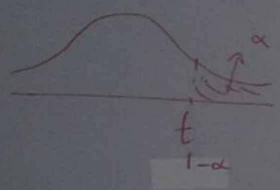
(5)

(4) آزمون فرض آماری برای تفاضل میانگین های دو جامعه نرمال $(\mu_1-\mu_2)$ وقتی واریانس ها مجهول و برابرند :

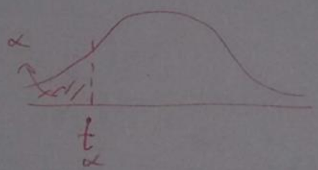
آماره آزمون تحت فرض H_0 :
$$T=\frac{(\bar{X}_1-\bar{X}_2)-d_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(n_1+n_2-2)}$$

که در آن :
$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

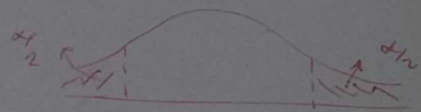
$$\begin{cases} H_0: \mu_1-\mu_2=d_0 \text{ یا } \le d_0 \\ H_1: \mu_1-\mu_2>d_0 \end{cases}$$
$\downarrow$
$$RH_0 \iff T>t_{1-\alpha}(n_1+n_2-2)$$



$$\begin{cases} H_0: \mu_1-\mu_2=d_0 \text{ یا } \ge d_0 \\ H_1: \mu_1-\mu_2<d_0 \end{cases}$$
$\downarrow$
$$RH_0 \iff T<t_{\alpha}(n_1+n_2-2)$$



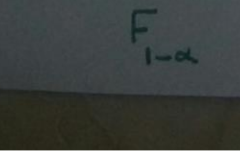
$$\begin{cases} H_0: \mu_1-\mu_2=d_0 \\ H_1: \mu_1-\mu_2\ne d_0 \end{cases}$$
$\downarrow$
$$RH_0 \iff |T|>t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$$



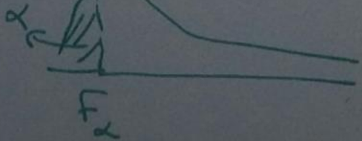
(5) آزمون فرض برای نسبت واریانس های دو جامعه نرمال $\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$:

آماره آزمون تحت H_0 :
$$F=\frac{S_1^2}{S_2^2}$$

$$\begin{cases} H_0: \sigma_1^2=\sigma_2^2 \text{ یا } \le \sigma_2^2 \\ H_1: \sigma_1^2>\sigma_2^2 \end{cases}$$
$\downarrow$
$$RH_0 \iff F>F_{1-\alpha}(n_1-1, n_2-1)$$



$$\begin{cases} H_0: \sigma_1^2=\sigma_2^2 \text{ یا } \ge \sigma_2^2 \\ H_1: \sigma_1^2<\sigma_2^2 \end{cases}$$
$\downarrow$
$$RH_0 \iff F<F_{\alpha}(n_1-1, n_2-1)$$



$$\begin{cases} H_0: \sigma_1^2=\sigma_2^2 \\ H_1: \sigma_1^2\ne\sigma_2^2 \end{cases}$$
$\downarrow$
$$RH_0 \iff F<F_{\frac{\alpha}{2}}(n_1-1, n_2-1) \text{ or } F>F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)$$