اثبات توزیع پارامتر و برآوردگر واریانس

$$f_{Z^2}(t) = \frac{t^{\frac{1}{2}-1}e^{-\frac{t}{t}}}{\Gamma(\frac{1}{2})2^{\frac{1}{2}}} \xRightarrow{\Gamma(\alpha,\beta):f_T(t)=\frac{t^{\alpha-1}e^{-\frac{t}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}} Z^2 \sim \Gamma(\frac{1}{2},2) \xRightarrow{\Gamma(\alpha,2)\sim\chi^2_{(2\alpha)}} Z^2 \sim \chi^{2^{(1)}} \tag{1}$$

$$\begin{aligned} &X_1, X_2, \ldots, X_n \sim \Gamma(\alpha_i,\beta) \Longrightarrow \sum_{i=1}^{n} X_i \sim \Gamma(\sum_{i=1}^{n}\alpha_i,\beta) \\ \xRightarrow{\Gamma(\alpha,2)\sim\chi^2_{(2\alpha)}} &X_1, X_2, \ldots, X_n \sim \chi^2_{(r_i)} \Longrightarrow \sum_{i=1}^{n} X_i \sim \chi^2_{(\sum_{i=1}^{n} r_i)} \end{aligned} \tag{2}$$

$$(1),(2) \implies \sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)} \tag{3}$$

$$\begin{aligned} X_i \sim N(0,1) \xRightarrow{(1)} (\frac{X_i-\mu}{\sigma})^2 \sim \chi^2_{(1)} \xRightarrow{(2)} \\ \sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2_{(n)} \end{aligned} \tag{4}$$

$$\begin{aligned} \sum_{i=1}^{n}(X_i-\mu)^2 &= \sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu)^2 \\ &= \sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{n}(\bar{X}-\mu)^2 + 2\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu) \\ &= \sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{n}(\bar{X}-\mu)^2 + 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X}) \\ &\xlongequal{\sum_{i=1}^{n}(X_i-\bar{X})=0} \sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2 \end{aligned} \tag{5}$$

$$(4),(5) \implies \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2_{(n)} \tag{6}$$

$$\begin{aligned} \bar{X} \sim N(\mu,\frac{\sigma^2}{n}) \implies \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \xRightarrow{(1)} \\ (\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}})^2 = n\frac{(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2_{(1)} \end{aligned} \tag{7}$$

$$\begin{aligned} (6),(7) \Longrightarrow \frac{\sum_{i=1}^{n}(X_i-\bar{X})}{\sigma^2} + \chi^2_{(1)} \sim \chi^2_{(n)} \\ \xRightarrow{(2)} \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)} \end{aligned} \tag{8}$$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} \implies \sum_{i=1}^{n}(X_i-\bar{X})^2 = (n-1)S^2 \tag{9}$$

$$(8),(9) \implies \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \tag{10}$$

Proof

$$X \sim \Gamma(\alpha,\beta) \implies M_X(t) = (1-\beta t)^{-\alpha} \tag{11}$$

$$Q = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}, \quad R = \frac{(n-1)S^2}{\sigma^2}, \quad N = \frac{n(\bar{X}-\mu)^2}{\sigma^2}$$

$$
\begin{aligned}
Q = R + N \implies\ & M_Q(t) = M_{R+N}(t) \\
\implies\ & M_Q(t) = M_R(t) \times M_N(t) \\
\xrightarrow[(7),(1)]{(4),(2)}\ & (1-2t)^{-\frac{n}{2}} = M_R(t) \times (1-2t)^{-\frac{1}{2}} \\
\implies\ & M_R(t) = (1-2t)^{-\frac{n-1}{2}} \\
\implies\ & R = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}
\end{aligned}
\tag{12}
$$

اثبات دو جامعهٔ نرمال مستقل

اگر داشته باشیم $Y = aX + b$ جایی‌که $X \sim N(\mu, \sigma^2)$ و a و b مقادیر ثابت باشند، آن‌گاه داریم:

$$E(Y) = E(aX + b) = aE(X) + b = a\mu + b \tag{1}$$

$$Var(Y) = Var(aX + b) = Var(aX) = a^2 Var(X) = a^2\sigma^2 \tag{2}$$

$$(1), (2) \implies Y = aX + b \sim N(a\mu + b, a^2\sigma^2) \tag{3}$$

اگر داشته باشیم $Z = \dfrac{X - \mu}{\sigma}$ جایی‌که $X \sim N(\mu, \sigma^2)$ آن‌گاه داریم:

$$Z = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \overset{(3)}{\Longrightarrow} Z \sim N(0, 1) \tag{4}$$

اگر داشته باشیم $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), \ldots, X_n \sim N(\mu_n, \sigma_n^2)$ به‌طوری‌که مستقل باشند، آن‌گاه برای هر $a_1, a_2, \ldots, a_n \in \mathbb{R}$ داریم:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} E\left(a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i \tag{5}$$

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} Var(a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) = \sum_{i=1}^{n} a_i^{2\sigma_i^2} \tag{6}$$

$$(1), (2) \implies \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right) \tag{7}$$

اگر داشته باشیم $X_1 \sim \chi^2_{r_1}, X_2 \sim \chi^2_{r_2}, \ldots, X_n \sim \chi^2_{r_n}$ آن‌گاه داریم:

$$\sum_{i=1}^{n} X_i \sim \chi^2_{\left(\sum_{i=1}^{n} r_i\right)} \tag{8}$$

اگر داشته باشیم $Z \sim N(0, 1)$ و $X \sim \chi^2_{(r)}$، آن‌گاه داریم:

$$\frac{Z}{\sqrt{\dfrac{X}{r}}} \sim t_{(r)} \tag{9}$$

اگر داشته باشیم $X_1, X_2, \ldots, X_n \sim N(\mu_1, \sigma_1^2)$ و $Y_1, Y_2, \ldots Y_n \sim N(\mu_2, \sigma_2^2)$ به‌طوری‌که این دو جامعه مستقل باشند، آن‌گاه داریم:

حالت ۱

واریانس دو جامعه معلوم

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \tag{10}$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right) \tag{11}$$

$$(10), (11) \overset{(7)}{\Longrightarrow} \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right) \tag{12}$$

$$(12) \overset{(4)}{\Longrightarrow} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0, 1) \tag{13}$$

حالت ۲

واریانس دو جامعه مجهول و برابر $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \tag{14}$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma^2}{m}\right) \tag{15}$$

$$(14), (15) \overset{(7)}{\Longrightarrow} \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right) \tag{16}$$

$$(16) \overset{(4)}{\Longrightarrow} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1) \tag{17}$$

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2_{(n-1)} \tag{18}$$

$$\frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2_{(m-1)} \tag{19}$$

$$(18), (19) \overset{(8)}{\Longrightarrow} \frac{(n-1)S_X^2 + (m-1)S_Y^2}{\sigma^2} \sim \chi^2_{(n+m-2)} \tag{20}$$

$$S_p := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n + m - 2} \tag{21}$$

$$(20), (21) \implies \frac{(n+m-2)S_p}{\sigma^2} \sim \chi^2_{(n+m-2)} \tag{22}$$

$$(16), (22) \overset{(9)}{\Longrightarrow} \frac{\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\dfrac{\frac{(n+m-2)S_p}{\sigma^2}}{n+m-2}}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t_{(n+m-2)} \tag{23}$$

فصل ۲ - قضیهٔ حد مرکزی

-

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

- $$\sum_{i=1}^{n} a_i X_i \sim N(\mu \sum_{i=1}^{n} a_i, \sigma^2 \sum_{i=1}^{n} a_i^2)$$

توزیع نرمال استاندارد (Standard Normal Distribution)

$$Z \sim N(0,1) \implies f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

- $$Z^2 \sim \chi^2_{(1)}$$
- $$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$$
- $$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \iff Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Central Limit Theorem

Let $X_1, X_2, \ldots X_n$ denote a random sample of n independent observations from a population with overall **expected value (average)** μ and finite **variance** σ^2 and let $\bar{X}$ denote the sample mean of that sample (which is itself a random variable). Then as $n \to \infty$ ($n > 30$), we have

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

توزیع تی-استودنت (Student's t-distribution)

$$\frac{Z}{\sqrt{\frac{X}{r}}} \sim t_{(r)} \qquad ; \qquad Z \sim N(0,1), \quad X \sim \chi^2_{(r)}$$

فصل ۳ - برآورد پارامترها

پارامتر: هر جزء قابل‌محاسبه و مجهول از جامعه
برای محاسبهٔ کامل پارامتر باید اطلاعات تمام اعضای جامعه را داشته باشیم.

- پارامتر جامعه یک عدد ثابت است.

برآوردگر: تابعی از مشاهدات که به‌منظور برآورد یا به دست آوردن تقریبی پارامتر جامعه استفاده می‌شود

- مقدار حاصل از برآوردگر یک جامعه در نمونه‌های مختلف، متفاوت است؛ بنابراین هر برآوردگر یک متغیر تصادفی است و توزیع آماری خود را دارد.
- تنها یکبار نمونه‌گیری انجام می‌دهیم و می‌بایست براساس همان مشاهدات، استنباط دربارهٔ پارامترهای جامعه را انجام داد.

برآوردگر نقطه‌ای (Point Estimation)

نماد	برآوردگر (Estimator)	نماد	پارامتر (Parameter)
$\bar{X}$	میانگین نمونه	μ	میانگین جامعه
S^2	واریانس نمونه	σ^2	واریانس جامعه
$\hat{p}$	نسبت نمونه	P or π	نسبت جامعه

برآوردگر فاصله‌ای (Interval Estimation)

بازهٔ (L, U) فاصلهٔ اطمینان یا برآوردگر فاصله‌ای برای پارامتر مدنظر در سطح اطمینان $100(1-\alpha)$درصدی است.

- امکان به دست آوردن فاصلهٔ اطمینان ۱۰۰درصدی (یعنی $\alpha = 0$) وجود ندارد.
- منظور از $P(\theta \in (L, U)) = 1 - \alpha$ این است که در هر ۱۰۰ عدد فاصلهٔ اطمینان مختلف که بر اساس ۱۰۰ نمونهٔ متفاوت برای پارامتر θ ساخته می‌شود، انتظار داریم به‌طور متوسط $100(1-\alpha)$ عدد از این فاصلهٔ اطمینان‌ها شامل پارامتر واقعی شوند.
- این‌طور نیست که $100(1-\alpha)$ درصد احتمال داشته باشد که پارامتر θ درون بازهٔ (L, U) قرار گیرد.

$$P(\theta \in (L, u)) = 1 - \alpha$$

$$P(\theta \geq L) = P(\theta \in (L, \infty)) = 1 - \alpha$$

$$P(\theta \leq U) = P(\theta \in (-\infty, U)) = 1 - \alpha$$

فاصلهٔ اطمینان (Confidence Interval)

میانگین جامعه (Population Mean) (μ) ›

حالت ۱

تعداد نمونه زیاد $(n > 30)$ و واریانس جامعه معلوم

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

- دوطرفه

$$(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$$

- یک‌طرفهٔ پایینی

$$(\bar{X} - Z_{\alpha} \frac{\sigma}{\sqrt{n}}, \infty)$$

- یک‌طرفهٔ بالایی

$$(-\infty, \bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}})$$

حالت ۲

تعداد نمونه زیاد ($n > 30$) و واریانس جامعه نامعلوم

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \to N(0,1)$$

- دوطرفه

$$(\bar{X} - Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}})$$

- یک‌طرفهٔ پایینی

$$(\bar{X} - Z_\alpha \frac{S}{\sqrt{n}}, \infty)$$

- یک‌طرفهٔ بالایی

$$(-\infty, \bar{X} + Z_\alpha \frac{S}{\sqrt{n}})$$

حالت ۳

مشاهدات نرمال و واریانس جامعه معلوم

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

- دوطرفه

$$(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$$

- یک‌طرفهٔ پایینی

$$(\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}}, \infty)$$

- یک‌طرفهٔ بالایی

$$(-\infty, \bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}})$$

حالت ۴

مشاهدات نرمال و واریانس جامعه نامعلوم

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

- دوطرفه

$$(\bar{X}-t_{\frac{\alpha}{2}}(n-1)\frac{\sigma}{\sqrt{n}}, \bar{X}+t_{\frac{\alpha}{2}}(n-1)\frac{\sigma}{\sqrt{n}})$$

- یک‌طرفهٔ پایینی

$$(\bar{X}-t_{\alpha}(n-1)\frac{\sigma}{\sqrt{n}}, \infty)$$

- یک‌طرفهٔ بالایی

$$(-\infty, \bar{X}+t_{\alpha}(n-1)\frac{\sigma}{\sqrt{n}})$$

نسبت در جامعه (Population Proportion) (π)

تعداد نمونه زیاد $(n > 30)$

- دوطرفه

$$(\hat{p}-Z_{\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p}+Z_{\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}})$$

- یک‌طرفهٔ پایینی

$$(\hat{p}-Z_{\alpha}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, 1)$$

- یک‌طرفهٔ بالایی

$$(0, \hat{p}+Z_{\alpha}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}})$$

واریانس جامعه (Population Variance) (σ^2)

مشاهدات نرمال و واریانس جامعه نامعلوم

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

- دوطرفه

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

- یک‌طرفهٔ پایینی

$$(\frac{n-1}{\chi^2_{\alpha}(n-1)}S^2, \infty)$$

- یک‌طرفهٔ بالایی

$$(-\infty, \frac{n-1}{\chi^2_{\alpha}(n-1)}S^2)$$

فصل ۴ - آزمون فرض

آزمون فرضیه: ابزاری آماری که به بررسی صحت دو ادعای مکمل دربارهٔ یک پارامتر در یک جامعه می‌پردازد

- هدف از آزمون فرضیه، پذیرش یکی از فرضیه‌ها و رد دیگری است.
- فرضیهٔ صفر به‌عنوان فرض اصلی و موردتحقیق بوده و قابل‌تحقیق است؛ یعنی حداقل یک مقدار مشخص در آن وجود دارد.

فرایند آزمون فرضیه

1. انتخاب فرضیه‌های صفر و یک
2. ایجاد ناحیهٔ رد (شرایطی که باعث رد فرض صفر می شود)
3. بررسی مشاهدات برای یافتن دلایلی بر رد فرض صفر (قرارگیری مقادیر حاصل از مشاهدات در ناحیهٔ رد)
4. تصمیم‌گیری (نتیجه‌گیری و تفسیر نتایج)

$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$	دوطرفه
$\begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$	یک‌طرفهٔ پایینی
$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$	یک‌طرفهٔ بالایی

خطاها (Errors)

$$\alpha = P(\text{Type I error occuring}) = P(R|H_0) = P_{H_0}(R)$$

$$\beta = P(\text{Type II error occuring}) = P(R'|H_1) = P_{H_1}(R')$$

- به α سطح آزمون می‌گویند.
- سطح آزمون را همیشه از قبل مشخص نموده و سپس تلاش می‌کنیم احتمال مربوط به β را کاهش دهیم.

ناحیهٔ رد (Rejection Region (Critical Region))

میانگین جامعه (Population Mean) (μ)

حالت ۱

تعداد نمونه زیاد $(n > 30)$ و واریانس جامعه معلوم

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

- دوطرفه: $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

$$R = \left\{ \left| \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > Z_{\frac{\alpha}{2}} \right\}$$

- یک‌طرفهٔ پایینی: $\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -Z_{\alpha} \right\}$$

- یک‌طرفهٔ بالایی: $\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha \right\}$$

حالت ۲

تعداد نمونه زیاد ($n > 30$) و واریانس جامعه نامعلوم

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \to N(0,1)$$

- دوطرفه: $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

$$R = \left\{ \left| \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \right| > Z_{\frac{\alpha}{2}} \right\}$$

- یک‌طرفهٔ پایینی: $\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} < -Z_\alpha \right\}$$

- یک‌طرفهٔ بالایی: $\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} > Z_\alpha \right\}$$

حالت ۳

مشاهدات نرمال و واریانس جامعه معلوم

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

- دوطرفه: $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

$$R = \left\{ \left| \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > Z_{\frac{\alpha}{2}} \right\}$$

- یک‌طرفهٔ پایینی: $\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} < -Z_\alpha \right\}$$

- یک‌طرفهٔ بالایی: $\begin{cases} H_0 : \mu \le \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha \right\}$$

حالت ۴

مشاهدات نرمال و واریانس جامعه نامعلوم

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

- دوطرفه: $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \ne \mu_0 \end{cases}$

$$R = \left\{ \left| \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \right| > t_{\frac{\alpha}{2}}(n-1) \right\}$$

- یک‌طرفهٔ پایینی: $\begin{cases} H_0 : \mu \ge \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} < -t_\alpha(n-1) \right\}$$

- یک‌طرفهٔ بالایی: $\begin{cases} H_0 : \mu \le \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

$$R = \left\{ \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} > t_\alpha(n-1) \right\}$$

نسبت در جامعه (Population Proportion) (π) ›

تعداد نمونه زیاد ($n > 30$)

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \to N(0,1)$$

- دوطرفه: $\begin{cases} H_0 : P = P_0 \\ H_1 : P \ne P_0 \end{cases}$

$$R = \left\{ \left| \frac{\hat{p} - P_0}{\frac{\sqrt{P_0(1-P_0)}}{\sqrt{n}}} \right| > Z_{\frac{\alpha}{2}} \right\}$$

- یک‌طرفهٔ پایینی: $\begin{cases} H_0 : P \ge P_0 \\ H_1 : P < P_0 \end{cases}$

$$R = \left\{ \frac{\hat{p} - P_0}{\frac{\sqrt{P_0(1-P_0)}}{\sqrt{n}}} < -Z_\alpha \right\}$$

- یک‌طرفهٔ بالایی: $\begin{cases} H_0 : P \leq P_0 \\ H_1 : P > P_0 \end{cases}$

$$R = \left\{ \frac{\hat{p} - P_0}{\frac{\sqrt{P_0(1-P_0)}}{\sqrt{n}}} > Z_\alpha \right\}$$

واریانس جامعه (Population Variance) (σ^2)

مشاهدات نرمال و واریانس جامعه نامعلوم

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

- دوطرفه: $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$

$$R = \left\{ \left| \frac{n-1}{\sigma_0^2} S^2 \right| > \chi_{\frac{\alpha}{2}}(n-1) \right\}$$

- یک‌طرفهٔ پایینی: $\begin{cases} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$

$$R = \left\{ \frac{n-1}{\sigma_0^2} S^2 < -\chi_\alpha(n-1) \right\}$$

- یک‌طرفهٔ بالایی: $\begin{cases} H_0 : \sigma^2 \leq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$

$$R = \left\{ \frac{n-1}{\sigma_0^2} S^2 > \chi_\alpha(n-1) \right\}$$

فصل ۵ - استنباط دو میانگین

فاصلهٔ اطمینان (Confidence Interval)

میانگین جامعه (Population Mean) (μ) <

حالت ۱

تعداد نمونه زیاد ($n > 30$) و واریانس جامعه معلوم (برای هردو جامعه)

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \to N(0,1)$$

- دوطرفه

$$\left(\bar{X} - \bar{Y} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \bar{X} - \bar{Y} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

- یک‌طرفهٔ پایینی

$$\left(\bar{X} - \bar{Y} - Z_{\alpha}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \infty\right)$$

- دوطرفهٔ پایینی

$$\left(-\infty, \bar{X} - \bar{Y} + Z_{\alpha}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

حالت ۲

تعداد نمونه زیاد ($n > 30$) و واریانس جامعه نامعلوم (برای هردو جامعه)

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \to N(0,1)$$

- دوطرفه

$$\left(\bar{X} - \bar{Y} - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}, \bar{X} - \bar{Y} + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}\right)$$

- یک‌طرفهٔ پایینی

$$\left(\bar{X} - \bar{Y} - Z_{\alpha}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}, \infty\right)$$

- دوطرفهٔ پایینی

$$\left(-\infty, \bar{X}-\bar{Y}+Z_{\alpha}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)$$

حالت ۳

مشاهدات نرمال و واریانس جامعه معلوم (برای هردو جامعه)

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1)$$

- دوطرفه

$$\left(\bar{X}-\bar{Y}-Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}, \bar{X}-\bar{Y}+Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)$$

- یک‌طرفهٔ پایینی

$$\left(\bar{X}-\bar{Y}-Z_{\alpha}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}, \infty\right)$$

- دوطرفهٔ پایینی

$$\left(-\infty, \bar{X}-\bar{Y}+Z_{\alpha}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)$$

حالت ۴

مشاهدات نرمال و واریانس جوامع برابر و نامعلوم اثبات

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}} \sim t(n+m-2)$$

$$S_p^2 = \frac{(n-1)S_{X^2}+(m-1)S_{X^2}}{m+n-2}$$

- دوطرفه

$$\left(\bar{X}-\bar{Y}-t_{\frac{\alpha}{2}}(n+m-2)\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}, \bar{X}-\bar{Y}+t_{\frac{\alpha}{2}}(n+m-2)\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}\right)$$

- یک‌طرفهٔ پایینی

$$\left(\bar{X}-\bar{Y}-t_{\frac{\alpha}{2}}(n+m-2)\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})},\infty\right)$$

- دوطرفهٔ پایینی

$$\left(-\infty,\bar{X}-\bar{Y}-t_{\frac{\alpha}{2}}(n+m-2)\sqrt{S_p^2(\frac{1}{n}+\frac{1}{m})}\right)$$

داده‌های زوجی

هرگاه n نمونه داشته باشیم و اطلاعات حاصل از آن‌ها مربوط به دو زمان مختلف، به منظور بررسی یک روش طراحی‌شده، باشد، عملاً از هر نمونه دو داده (دو مشاهده) داریم؛ یعنی این داده‌ها مستقل نیستند. این داده‌ها غالباً حاصل از تکرار یک شاخص اندازه‌گیری‌ست که به داده‌های قبل و بعد نیز معروف است.
هدف از جمع‌آوری این نوع داده‌ها، بررسی اثربخشی انجام یک روش، استراتژی، روند درمان و ... است. عملاً به دنبال بررسی $\mu_D=\mu_A-\mu_B$ هستیم؛ جایی‌که μ_A میانگین بعد از اعمال موضوع و μ_B میانگین قبل از اعمال است.

حالت ۱ (۲)

تعداد نمونه زیاد ($n>30$) و واریانس جامعه نامعلوم

$$\bar{D}\pm Z_{\frac{\alpha}{2}}\frac{S_D}{\sqrt{n}}$$

حالت ۲ (۴)

بررسی فرایند طبیعی بدون مداخله

$$\bar{D}\pm t_{\frac{\alpha}{2}}\frac{S_D}{\sqrt{n}}$$

فصل ۷ - رگرسیون

ضریب همبستگی

$$\hat{\rho}_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)\sqrt{S_X^2 \times S_Y^2}}$$

هرگاه ضریب همبستگی بین مشاهدات (X_i, Y_i) برای $i = 1, 2, \dots, n$ مقادیر نزدیک 1 یا -1 باشد، گوییم بین این دو متغیر یک رابطهٔ خطی معنادار وجود دارد. این رابطهٔ خطی را به فرم زیر تعریف می‌کنیم:

$$y_i = a + bx_i + \epsilon_i; \qquad i = 1, 2, \dots, n$$

برآورد پارامترها

می‌خواهیم a و b را چنان بیابیم که مجموع مربعات خطا کمینه شود. تابع L را به‌صورت زیر تعریف می‌کنیم:

$$L(a,b) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n}(y_i - (a + bx_i))^2$$

برای کمینه کردن این تابع، از آن نسبت به a و b مشتق می‌گیریم تا برآوردگرهای a و b را بیابیم.

$$\begin{aligned}
\frac{dL(a,b)}{da} = 0 &\implies \frac{d}{da}\sum_{i=1}^{n}(y_i - (a + bx_i))^2 = 0 \\
&\implies \sum_{i=1}^{n}\frac{d}{da}(y_i - (a + bx_i))^2 = 0 \\
&\implies -2\sum_{i=1}^{n}(y_i - (a + bx_i)) = 0 \\
&\implies n\bar{y} - na - nb\bar{x} = 0 \\
&\implies \bar{y} - b\bar{x} - a = 0 \\
&\implies \hat{a} = \bar{y} - \hat{b}\bar{x}
\end{aligned} \tag{1}$$

$$\begin{aligned}
\frac{dL(a,b)}{db} = 0 &\implies \frac{d}{db}\sum_{i=1}^{n}(y_i - (a + bx_i))^2 = 0 \\
&\implies \sum_{i=1}^{n}\frac{d}{db}(y_i - (a + bx_i))^2 = 0 \\
&\implies -2\sum_{i=1}^{n}x_i(y_i - (a + bx_i)) = 0 \\
&\implies \sum_{i=1}^{n}x_iy_i - a\sum_{i=1}^{n}x_i - b\sum_{i=1}^{n}x_i^2 = 0 \\
&\xRightarrow{(*)} \sum_{i=1}^{n}x_iy_i - (\bar{y} - b\bar{x})n\bar{x} - b\sum_{i=1}^{n}x_i^2 = 0 \\
&\implies \left(\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}\right) - b\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right) = 0 \\
&\implies \hat{b} = \frac{\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}
\end{aligned} \tag{2}$$

$$\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y} = \sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}y_i\bar{x} = \sum_{i=1}^{n}y_i(x_i - \bar{x}) \tag{3}$$

$$SSX := \sum_{i=1}^{n}x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}x_i\bar{x} = \sum_{i=1}^{n}x_i(x_i - \bar{x}) \tag{4}$$

$$(2) \xRightarrow{(3),(4)} \hat{b} = \frac{\sum_{i=1}^{n}y_i(x_i - \bar{x})}{SSX} = \sum_{i=1}^{n}y_i\frac{x_i - \bar{x}}{SSX} \tag{5}$$

- مقادیر $\hat{a}$ و $\hat{b}$ به‌ازای نمونه‌های مختلف، متفاوت هستند؛ بنابراین این دو برآوردگر، متغیر تصادفی هستند.

توزیع پارامترها

توزیع y_i

فرض می‌کنیم $\epsilon_i \sim N(0, \sigma^2)$ زیرا عوامل تأثیرگذار روی خطا، رفتاری نرمال دارند. آن‌گاه داریم:

$$E(y_i) = E(a + bx_i + \epsilon_i) = a + bx_i + E(\epsilon_i) = a + bx_i \tag{6}$$

$$Var(y_i) = Var(a + bx_i + \epsilon_i) = Var(\epsilon_i) = \sigma^2 \tag{7}$$

$$(6), (7) \implies y_i = a + bx_i + \epsilon_i \sim N(a + bx_i, \sigma^2) \tag{8}$$

توزیع $\hat{b}$

با توجه به آن‌که $\hat{b}$ یک ترکیب خطی از y_i است، این برآوردگر دارای توزیع نرمال است و داریم:

$$\begin{aligned} E(\hat{b}) &= E\left(\sum_{i=1}^{n} y_i \frac{x_i - \bar{x}}{SSX}\right) \\ &= \sum_{i=1}^{n} E(y_i) \frac{x_i - \bar{x}}{SSX} \\ &= \sum_{i=1}^{n} (a + bx_i) \frac{x_i - \bar{x}}{SSX} \\ &= a \sum_{i=1}^{n} \frac{x_i - \bar{x}}{SSX} + b \sum_{i=1}^{n} x_i \frac{x_i - \bar{x}}{SSX} \\ &= 0 + b = b \end{aligned} \tag{9}$$

$$\begin{aligned} Var(\hat{b}) &= Var\left(\sum_{i=1}^{n} y_i \frac{x_i - \bar{x}}{SSX}\right) \\ &= \sum_{i=1}^{n} Var(y_i) \frac{(x_i - \bar{x})^2}{SSX^2} \\ &= \sum_{i=1}^{n} \sigma^2 \frac{(x_i - \bar{x})^2}{SSX^2} \\ &= \sigma^2 \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{SSX^2} \\ &= \sigma^2 \frac{SSX}{SSX^2} \\ &= \frac{\sigma^2}{SSX} \end{aligned} \tag{10}$$

$$(9), (10) \implies \hat{b} = \sum_{i=1}^{n} y_i \frac{x_i - \bar{x}}{SSX} \sim N\left(b, \frac{\sigma^2}{SSX}\right) \tag{11}$$

$$(11) \implies \frac{\hat{b} - b}{\frac{\sigma}{\sqrt{SSX}}} \sim N(0, 1) \tag{12}$$

توزیع $\hat{a}$

با توجه به آن‌که $\hat{a}$ یک ترکیب خطی از $\bar{y}$ و $\hat{b}$ است، این برآوردگر دارای توزیع نرمال است و داریم:

$$\begin{aligned} E(\hat{a}) &= E(\bar{y} - \hat{b}\bar{x}) \\ &= E(\bar{y}) - E(\hat{b}\bar{x}) \\ &= a + b\bar{x} - b\bar{x} \\ &= a \end{aligned} \tag{13}$$

$$
\begin{aligned}
Var(\hat{a}) &= Var(\bar{y} - \hat{b}\bar{x}) \\
&= Var\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i \frac{\bar{x}(x_i - \bar{x})}{SSX}\right) \\
&= Var\left(\sum_{i=1}^{n} y_i \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{SSX}\right)\right) \\
&= \sum_{i=1}^{n} Var(y_i) \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{SSX}\right)^2 \\
&= \sum_{i=1}^{n} \sigma^2 \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{SSX}\right)^2 \\
&= \sigma^2 \sum_{i=1}^{n} \left(\frac{1}{n^2} + \frac{\bar{x}^2(x_i - \bar{x}^2)}{SSX^2} - 2\frac{\bar{x}(x_i - \bar{x})}{nSSX}\right) \\
&= \sigma^2 \left(\frac{n}{n^2} + \frac{\bar{x}^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{SSX^2} - 2\bar{x}\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{SSX}\right) \\
&= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)
\end{aligned} \tag{14}
$$

$$
(13), (14) \implies \hat{a} = \bar{y} - \hat{b}\bar{x} \sim N\left(a, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)\right) \tag{15}
$$

$$
(14) \implies \frac{\hat{a} - a}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SSX}}} \sim N(0,1) \tag{16}
$$

فاصلهٔ اطمینان پارامترها

فاصلهٔ اطمینان b

بازهٔ (L, U) را برای b در نظر می‌گیریم به‌طوری که $L = \hat{b} - c$ و $U = \hat{b} + d$.

$$
\begin{aligned}
1 - \alpha &= P\left(b \in (L, U)\right) \\
&= P\left(P \in (\hat{b} - c, \hat{b} + d)\right) \\
&= P\left(\hat{b} - c < b < \hat{b} + d\right) \\
&= P\left(-c < b - \hat{b} < d\right) \\
&= P\left(-d < \hat{b} - b < c\right) \\
&= P\left(\frac{-d}{\sqrt{\frac{\sigma^2}{SSX}}} < \frac{\hat{b} - b}{\sqrt{\frac{\sigma^2}{SSX}}} < \frac{c}{\sqrt{\frac{\sigma^2}{SSX}}}\right) \\
&= P\left(-d^* < Z < c^*\right) \\
&\xRightarrow{c^* = d^* = Z_{\frac{\alpha}{2}}} c = d = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{SSX}}
\end{aligned}
$$

بنابراین فاصلهٔ اطمینان $100(1-\alpha)$درصدی برای $\hat{b}$ برابر است با:

$$
\left(\hat{b} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{SSX}}, \hat{b} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{SSX}}\right) = \hat{b} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{SSX}}
$$

فاصلهٔ اطمینان $\hat{a}$

بازهٔ (L, U) را برای a در نظر می‌گیریم به‌طوری‌که $L = \hat{a} - c$ و $U = \hat{a} + d$.

$$
\begin{aligned}
1-\alpha &= P(a \in (L,U)) \\
&= P(a \in (\hat{a}-c, \hat{a}+d)) \\
&= P(\hat{a}-c < a < \hat{a}+d) \\
&= P(-c < a-\hat{a} < d) \\
&= P(-d < \hat{a}-a < c) \\
&= P\left(-\frac{d}{\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SSX}}} < \frac{\hat{a}-a}{\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SSX}}} < \frac{c}{\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SSX}}}\right) \\
&= P(-d^* < Z < c^*) \\
&\xRightarrow{c^*=d^*=Z_{\frac{\alpha}{2}}} c = d = Z_{\frac{\alpha}{2}}\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SSX}}
\end{aligned}
$$

بنابراین فاصلهٔ اطمینان $100(1-\alpha)$درصدی برای a برابر است با:

$$
\left(\hat{a}-Z_{\frac{\alpha}{2}}\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SSX}}, \hat{a}+Z_{\frac{\alpha}{2}}\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SSX}}\right) = \hat{a} \pm Z_{\frac{\alpha}{2}}\sigma\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SSX}}
$$

آزمون فرض

پارامتر b

$$
\begin{cases} H_0: b = b_0 \\ H_1: b \neq b_0 \end{cases}
$$

فرض صفر رد می شود هرگاه میزان تغییر در مقدار y به‌ازای یک واحد تغییر در متغیر x از مقدار ادعاشده در فرض صفر خیلی بزرگ‌تر یا کوچک‌تر باشد؛ یعنی برای $c < b_0 < d$ داشته باشیم $R = \{\hat{b} < c \vee \hat{b} > d\}$.

$$
\begin{aligned}
\alpha = P(R|H_0) &= P\left(\hat{b} < c \vee \hat{b} > d | b = b_0\right) \\
&= P_{b=b_0}\left(\hat{b}-b < c-b \vee \hat{b}-b > d-b\right) \\
&= P_{b=b_0}\left(\frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} < \frac{c-b}{\sqrt{\frac{\sigma^2}{SSX}}} \vee \frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} > \frac{d-b}{\sqrt{\frac{\sigma^2}{SSX}}}\right) \\
&= P(Z < c^* \vee Z > d^*) \\
&\implies -c^* = d^* = Z_{\frac{\alpha}{2}}
\end{aligned}
$$

بنابراین ناحیهٔ رد این‌طور به دست می‌آید:

$$
R = \left\{\frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} < -Z_\alpha \vee \frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} > Z_\alpha\right\} = \left\{|\frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}}| > Z_{\frac{\alpha}{2}}\right\}
$$

یک‌طرفهٔ پایینی:

$$
\begin{cases} H_0: b \geq b_0 \\ H_1: b < b_0 \end{cases}
$$

فرض صفر رد می‌شود هرگاه میزان تغییر در مقدار y به‌ازای یک واحد تغییر در متغیر x از مقدار ادعاشده در فرض صفر خیلی کوچک‌تر باشد؛ یعنی برای $c < b_0$ داشته باشیم $R = \{\hat{b} < c\}$.

$$
\begin{aligned}
\alpha = P(R|H_0) &= P\left(\hat{b} < c | b = b_0\right) \\
&= P_{b=b_0}\left(\hat{b}-b < c-b\right) \\
&= P_{b=b_0}\left(\frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} < \frac{c-b}{\sqrt{\frac{\sigma^2}{SSX}}}\right) \\
&= P(Z < c^*) \implies -c^* = Z_\alpha
\end{aligned}
$$

$$
R = \{\frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} < -Z_\alpha\}
$$

پارامتر a

$$\begin{cases} H_0: a = a_0 \\ H_1: a \neq a_0 \end{cases}$$

$$\begin{aligned} \alpha = P(R|H_0) &= P\left(\hat{a} < c \vee \hat{a} > d | a = a_0\right) \\ &= P_{a=a_0}\left(\hat{a} - a < c - a \vee \hat{a} - a > d - a\right) \\ &= P_{a=a_0}\left(\frac{\hat{a} - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{SSX}\right)}} < \frac{c - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{SSX}\right)}} \vee \frac{\hat{a} - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{SSX}\right)}} > \frac{d - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{SSX}\right)}}\right) \\ &= P\left(Z < c^* \vee Z > d^*\right) \\ &\implies -c^* = d^* = Z_{\frac{\alpha}{2}} \end{aligned}$$

$$R = \left\{\frac{\hat{a} - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{SSX}\right)}} < -Z_\alpha \vee \frac{\hat{a} - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{SSX}\right)}} > Z_\alpha\right\} = \left\{\left|\frac{\hat{a} - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{SSX}\right)}}\right| > Z_{\frac{\alpha}{2}}\right\}$$

بررسی وجود مدل رگرسیون

می‌خواهیم ببینیم آیا برای مشاهدات به‌دست‌آمده، می‌توان مدل رگرسیون خطی ارائه کرد یا نه.

1. رسم نمودار پراکندگی
2. برآورد a و b (به دست آوردن $\hat{a}$ و $\hat{b}$)
3. بررسی توزیع جملهٔ خطا (نرمال بودن)
4. برآورد فاصلهٔ اطمینان a و b
5. آزمون فرضیهٔ a و b
