推定 (まとめ)

点推定

母数		推定量		推定値
母平均	μ	標本平均	$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$	$\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$
母分散	σ^2	標本不偏分散	$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$	$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$
母比率	p	標本比率	$\hat{p} = \overline{X} = \dfrac{R}{n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$	$\hat{p} = \overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

推定量の性質：

推定量	不偏性	一致性	有効性
$\overline{X}$	○	○	○
S^2	○	○	×
$\hat{p} = \overline{X}$	○	○	○

$\{X_1, X_2, \cdots, X_n\}$ **は大きさ** n **の標本，**$\{x_1, x_2, \cdots, x_n\}$ **はその実現値**

確率変数 R **の実現値** x

区間推定

母平均 μ の区間推定

◎ 正規母集団

大きさ n の無作為標本 $\{X_1, X_2, \cdots, X_n\}$。すべての $i = 1, 2, \cdots, n$ について，$X_i \sim N(\mu, \sigma^2)$ とする。

母分散 σ^2 が既知のとき：

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \qquad \Longrightarrow \qquad P\left(\left|\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}\right| < z_{\alpha/2}\right) = 1 - \alpha$$

$z_{\alpha/2}$ **は** $100 \times \dfrac{\alpha}{2}$ **% 点で，確率** α **が与えられると，正規分布表 (P.251) から得られる。**

したがって， $P\left(\overline{X} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

$\overline{X}$ **を** $\overline{x}$ **で置き換えて，信頼係数** $1 - \alpha$ **の** μ **の信頼区間：** $\Longrightarrow \boxed{\left(\overline{x} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)}$

母分散 σ^2 **が未知のとき：**

$$\frac{\overline{X} - \mu}{S / \sqrt{n}} \sim t(n-1) \qquad \Longrightarrow \qquad P\left(\left|\frac{\overline{X} - \mu}{S / \sqrt{n}}\right| < t_{\alpha/2}(n-1)\right) = 1 - \alpha$$

$t_{\alpha/2}(n-1)$ **は** $100 \times \dfrac{\alpha}{2}$ **% 点で，確率** α **と自由度** $n - 1$ **が与えられると，** t **分布表 (P.252) から得られる。**

したがって， $P\left(\overline{X} - t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}}\right) = 1 - \alpha$

$\overline{X}, S^2$ を $\overline{x}, s^2$ で置き換えて，信頼係数 $1-\alpha$ の μ の信頼区間：

$$\Longrightarrow \boxed{\left(\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right)}$$

◎ 非正規母集団 (大標本, すなわち，n が大きいとき)

大きさ n の無作為標本 $\{X_1, X_2, \cdots, X_n\}$

すべての $i = 1, 2, \cdots, n$ について，$X_i \sim (\mu, \sigma^2)$ とする (正規分布を仮定する必要ない)。

母分散 σ^2 が既知のとき：　n **が大きいとき，中心極限定理 (定理 6.1, P.90) により，以下が成り立つ。**

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \Longrightarrow \qquad P\left(\left|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right) \approx 1-\alpha$$

したがって， $P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha$

$\overline{X}$ **を** $\overline{x}$ **で置き換えて，信頼係数** $1-\alpha$ **の** μ **の信頼区間：**$\Longrightarrow$ $\boxed{\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)}$

母分散 σ^2 が未知のとき：　**さらに，分母の** σ^2 **を標本不偏分散** S^2 **で置き換えて，近似する。**

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \Longrightarrow \qquad \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim N(0,1) \qquad \Longrightarrow \qquad P\left(\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right| < z_{\alpha/2}\right) \approx 1-\alpha$$

したがって，$P\left(\overline{X} - z_{\alpha/2}\dfrac{S}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\dfrac{S}{\sqrt{n}}\right) \approx 1 - \alpha$

$\overline{X}, S^2$ を $\overline{x}, s^2$ で置き換えて，信頼係数 $1-\alpha$ の μ の信頼区間：$\Longrightarrow \boxed{\left(\overline{x} - z_{\alpha/2}\dfrac{s}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\dfrac{s}{\sqrt{n}}\right)}$

母分散 σ^2 の区間推定

正規母集団を仮定すると，

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

となる。

$$P\left(\chi_{1-\alpha/2}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi_{\alpha/2}(n-1)\right) = 1 - \alpha$$

$$P\Bigl(\frac{(n-1)S^2}{\chi_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\alpha/2}(n-1)}\Bigr) = 1-\alpha$$

S^2 を s^2 で置き換えて，信頼係数 $1-\alpha$ の σ^2 の信頼区間：$\Longrightarrow$ $\boxed{\Bigl(\frac{(n-1)S^2}{\chi_{\alpha/2}(n-1)},\ \frac{(n-1)S^2}{\chi_{1-\alpha/2}(n-1)}\Bigr)}$

母比率 p の区間推定

非正規母集団 (大標本, 母分散 σ^2 が未知のとき) に相当

中心極限定理により，近似的に，

$$\frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

を得る。

$\hat{p} = \overline{X}$ と考える。

さらに，分母の p をその推定量 $\overline{X}$ で置き換えて，近似する。

$$\frac{\overline{X} - p}{\sqrt{\overline{X}(1-\overline{X})/n}} \sim N(0,1) \qquad \Longrightarrow \qquad P\left(\left|\frac{\overline{X} - p}{\sqrt{\overline{X}(1-\overline{X})/n}}\right| < z_{\alpha/2}\right) \approx 1 - \alpha$$

したがって，$P\left(\overline{X} - z_{\alpha/2}\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}} < p < \overline{X} + z_{\alpha/2}\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}}\right) \approx 1 - \alpha$

$\overline{X} = \hat{p}$ を $\overline{x} = \hat{p}$ で置き換えて，信頼係数 $1-\alpha$ の p の信頼区間：

$$\Longrightarrow \boxed{\left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}, \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right)}$$

$\hat{p} = \overline{x}$ と考える。

仮説検定 (まとめ)

検定の種類

母平均 μ の検定

◎ 正規母集団

大きさ n の無作為標本 $\{X_1, X_2, \cdots, X_n\}$。$i = 1, 2, \cdots, n$ について，$X_i \sim N(\mu, \sigma^2)$ とする。

母分散 σ^2 は既知のとき： $\overline{X}$ **の分布は，** $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ **なので，帰無仮説** $H_0:\ \mu=\mu_0$ **が正しいもとで，**

$\boxed{\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)}$ **となる (**μ **を** μ_0 **で置き換える)。このとき，検定統計量** $\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$**。検定統計量の値** $\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$**。**

1. **対立仮説 $H_1:\ \mu<\mu_0$ (片側検定)**

   $P\left(\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha\right)=\alpha$ **なので，** $\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$ **のとき，有意水準** α **で帰無仮説** $H_0:\ \mu=\mu_0$ **を棄却する。**

2. **対立仮説 $H_1:\ \mu>\mu_0$ (片側検定)**

   $P\left(\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha\right)=\alpha$ **なので，** $\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$ **のとき，有意水準** α **で帰無仮説** $H_0:\ \mu=\mu_0$ **を**

棄却する。

3. **対立仮説** H_1 **:** $\mu \neq \mu_0$ **(両側検定)**

$P\left(\left|\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}\right) = \alpha$ なので，$\left|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$ のとき，有意水準 α で帰無仮説 H_0 : $\mu = \mu_0$ を棄却する。

母分散 σ^2 は未知のとき： $\overline{X}$ の分布は，$\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ なので，帰無仮説 H_0 : $\mu = \mu_0$ が正しいもとで，$\boxed{\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)}$ となる (μ を μ_0 で置き換える)。このとき，検定統計量 $\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$。検定統計量の値 $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}$。

1. **対立仮説** H_1 **:** $\mu < \mu_0$ **(片側検定)**

$P\left(\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} < -t_\alpha(n-1)\right) = \alpha$ **なので，** $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}} < -t_\alpha(n-1)$ **のとき，有意水準** α **で** $H_0 : \mu = \mu_0$ **を棄却する。**

2. 対立仮説 $H_1 : \mu > \mu_0$ (片側検定)

$P\left(\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} > t_\alpha(n-1)\right) = \alpha$ **なので，** $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}} > t_\alpha(n-1)$ **のとき，有意水準** α **で** $H_0 : \mu = \mu_0$ **を棄却する。**

3. 対立仮説 $H_1 : \mu \neq \mu_0$ (両側検定)

$P\left(\left|\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}\right| > t_{\alpha/2}(n-1)\right) = \alpha$ **なので，** $\left|\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}\right| > t_{\alpha/2}(n-1)$ **のとき，有意水準** α **で** $H_0 : \mu = \mu_0$ **を棄却する。**

◎ **非正規母集団 (大標本，すなわち，n が大きいとき)**

大きさ n **の無作為標本** $\{X_1, X_2, \cdots, X_n\}$。$i = 1, 2, \cdots, n$ **について，** $X_i \sim (\mu, \sigma^2)$ **とする (正規分布の仮定は不必要)。**

母分散 σ^2 **は既知のとき：　近似的に** $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ **が成り立つので，帰無仮説** $H_0:\ \mu = \mu_0$ **が正しいもとで，** $\boxed{\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)}$ **となる** (μ **を** μ_0 **で置き換える)。このとき，検定統計量** $\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$。**検定統計量の値** $\dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$。

1. **対立仮説** $H_1:\ \mu < \mu_0$ **(片側検定)**

   $P\Big(\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha\Big) \approx \alpha$ **なので，** $\dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha$ **のとき，有意水準** α **で帰無仮説** $H_0:\ \mu = \mu_0$

を棄却する。

2. 対立仮説 H_1 ： $\mu > \mu_0$ (片側検定)

$P\Big(\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha\Big) \approx \alpha$ **なので，** $\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$ **のとき，有意水準** α **で帰無仮説** $H_0:\ \mu = \mu_0$ **を棄却する。**

3. 対立仮説 H_1 ： $\mu \neq \mu_0$ (両側検定)

$P\Big(\Big|\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\Big| > z_{\alpha/2}\Big) \approx \alpha$ **なので，** $\Big|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\Big| > z_{\alpha/2}$ **のとき，有意水準** α **で帰無仮説** $H_0:\ \mu = \mu_0$ **を棄却する。**

母分散 σ^2 **は未知のとき：　近似的に，** $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$ **が成り立つので，帰無仮説** $H_0:\ \mu = \mu_0$ **が正しいもとで，** $\boxed{\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim N(0,1)}$ **となる (**μ **を** μ_0 **で置き換える)。このとき，検定統計量**

$\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$。**検定統計量の値** $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}$。

1. **対立仮説** H_1 **:** $\mu < \mu_0$ **(片側検定)**

   $P\left(\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} < -z_\alpha\right) \approx \alpha$ **なので，** $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}} < -z_\alpha$ **のとき，有意水準** α **で帰無仮説** H_0 : $\mu = \mu_0$ **を棄却する。**

2. **対立仮説** H_1 **:** $\mu > \mu_0$ **(片側検定)**

   $P\left(\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} > z_\alpha\right) \approx \alpha$ **なので，** $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}} > z_\alpha$ **のとき，有意水準** α **で帰無仮説** H_0 : $\mu = \mu_0$ **を棄却する。**

3. **対立仮説** H_1 **:** $\mu \neq \mu_0$ **(両側検定)**

   $P\left(\left|\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}\right| > z_{\alpha/2}\right) \approx \alpha$ **なので，** $\left|\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}\right| > z_{\alpha/2}$ **のとき，有意水準** α **で帰無仮説** H_0 : $\mu = \mu_0$ **を棄却する。**

2 つの標本の母平均の差 (H_0： $\mu_1 = \mu_2$) の検定

◎ 正規母集団

- 第 1 グループ：大きさ n_1 の無作為標本。$i = 1, 2, \cdots, n$ について，$X_{1i} \sim N(\mu_1, \sigma_1^2)$ とする。
- 第 2 グループ：大きさ n_2 の無作為標本。$i = 1, 2, \cdots, n$ について，$X_{2i} \sim N(\mu_2, \sigma_2^2)$ とする。

σ_1^2 と σ_2^2 は既知のとき： 母平均の差を検定したいので，統計量 $\overline{X}_1 - \overline{X}_2$ の分布を考える。

定理 4.1 (P.54)，定理 4.5 (P.60) より，$\mathrm{E}(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2$，また，定理 4.3 (P.55)，定理 4.8 (P.62) より，$\mathrm{V}(\overline{X}_1 - \overline{X}_2) = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$ を得る。したがって，$\overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2})$ が成

り立つ (証明略)。さらに，標準化によって，$\dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$ を得るので，帰無仮説 H_0： $\mu_1 = \mu_2$ が正しいもとで，$\boxed{\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)}$ となる ($\mu_1 - \mu_2 = 0$ を代入する)。このとき，検定統計量 $\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$。検定統計量の値 $\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$。

1. **対立仮説 H_1： $\mu_1 < \mu_2$ (片側検定)**

   $P\left(\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < -z_\alpha\right) \approx \alpha$ なので，$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < -z_\alpha$ のとき，有意水準 α で H_0 を棄却する。

2. **対立仮説 H_1： $\mu_1 > \mu_2$ (片側検定)**

$P\left(\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} > z_\alpha\right) \approx \alpha$ なので，$\frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} > z_\alpha$ のとき，有意水準 α で H_0 を棄却する。

3. **対立仮説 H_1： $\mu_1 \neq \mu_2$ (両側検定)**

$P\left(\left|\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right| > z_{\alpha/2}\right) \approx \alpha$ なので，$\left|\frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right| > z_{\alpha/2}$ のとき，有意水準 α で H_0 を棄却する。

σ_1^2 と σ_2^2 は未知のとき：　小標本の場合，検定統計量の分布を導出できないため，検定不可能 (正規分布や t 分布にはならない)。

大標本の場合，正規分布で近似。

◎ **非正規母集団 (大標本，すなわち，n_1, n_2 が共に大きいとき)**

・**第 1 グループ：大きさ** n_1 **の無作為標本。**$i = 1, 2, \cdots, n$ **について，**$X_{1i} \sim (\mu_1, \sigma_1^2)$ **とする (正規分布の仮定は不必要)。**

・**第 2 グループ：大きさ** n_2 **の無作為標本。**$i = 1, 2, \cdots, n$ **について，**$X_{2i} \sim (\mu_2, \sigma_2^2)$ **とする (正規分布の仮定は不必要)。**

$\boldsymbol{\sigma_1^2}$ **と** $\boldsymbol{\sigma_2^2}$ **は既知のとき：**　**近似的に，**$\dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$ **を得るので，帰無仮説** $H_0:\ \mu_1 = \mu_2$ **が正しいもとで，**$\boxed{\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)}$ **となる** ($\mu_1 - \mu_2 = 0$ **を代入する)。**

このとき，検定統計量 $\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$。検定統計量の値 $\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$。

1. **対立仮説 H_1 ： $\mu_1 < \mu_2$ (片側検定)**

   $P\left(\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < -z_\alpha\right) \approx \alpha$ なので，$\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < -z_\alpha$ のとき，有意水準 α で H_0 を棄却する。

2. **対立仮説 H_1 ： $\mu_1 > \mu_2$ (片側検定)**

   $P\left(\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} > z_\alpha\right) \approx \alpha$ なので，$\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} > z_\alpha$ のとき，有意水準 α で H_0 を棄却する。

3. **対立仮説 H_1 ： $\mu_1 \neq \mu_2$ (両側検定)**

$P\left(\left|\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right| > z_{\alpha/2}\right) \approx \alpha$ なので，$\left|\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right| > z_{\alpha/2}$ のとき，有意水準 α で H_0 を棄却する。

σ_1^2 と σ_2^2 は未知のとき： 近似的に，$\dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim N(0,1)$ を得るので，帰無仮説 H_0 : $\mu_1 = \mu_2$ が正しいもとで，$\boxed{\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim N(0,1)}$ となる $(\mu_1 - \mu_2 = 0$ を代入する$)$。

このとき，検定統計量 $\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$。検定統計量の値 $\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$。

1. 対立仮説 H_1 : $\mu_1 < \mu_2$ (片側検定)

$P\left(\dfrac{\overline{X}_1-\overline{X}_2}{\sqrt{S_1^2/n_1+S_2^2/n_2}}<-z_\alpha\right)\approx\alpha$ **なので，** $\dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{s_1^2/n_1+s_2^2/n_2}}<-z_\alpha$ **のとき，有意水準** α **で** H_0 **を棄却する。**

2. 対立仮説 H_1 ： $\mu_1>\mu_2$ (片側検定)

$P\left(\dfrac{\overline{X}_1-\overline{X}_2}{\sqrt{S_1^2/n_1+S_2^2/n_2}}>z_\alpha\right)\approx\alpha$ **なので，** $\dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{s_1^2/n_1+s_2^2/n_2}}>z_\alpha$ **のとき，有意水準** α **で** H_0 **を棄却する。**

3. 対立仮説 H_1 ： $\mu_1\neq\mu_2$ (両側検定)

$P\left(\left|\dfrac{\overline{X}_1-\overline{X}_2}{\sqrt{S_1^2/n_1+S_2^2/n_2}}\right|>z_{\alpha/2}\right)\approx\alpha$ **なので，** $\left|\dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{s_1^2/n_1+s_2^2/n_2}}\right|>z_{\alpha/2}$ **のとき，有意水準** α **で** H_0 **を棄却する。**

母比率 p の検定

非正規母集団 (大標本，母分散が未知のとき) に相当

$\overline{X} = \hat{p}$，$\overline{x} = \hat{p}$ **と考える。**

中心極限定理により，近似的に， $\dfrac{\overline{X} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$ **を得るので，帰無仮説** H_0： $p = p_0$ **が正しいもとで，** $\boxed{\dfrac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)}$ **となる (**p **を** p_0 **で置き換える)。このとき，検定統計量** $\dfrac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)/n}}$**。検定統計量の値** $\dfrac{\overline{x} - p_0}{\sqrt{p_0(1-p_0)/n}}$**。**

1. 対立仮説 H_1 **：** $\mu < \mu_0$ **(片側検定)**

$P\Bigl(\dfrac{\overline{X}-p_0}{\sqrt{p_0(1-p_0)/n}} < -z_\alpha\Bigr) \approx \alpha$ なので，$\dfrac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)/n}} < -z_\alpha$ のとき，有意水準 α で H_0 : $\mu = \mu_0$ を棄却する。

2. 対立仮説 H_1 : $\mu > \mu_0$ (片側検定)

$P\Bigl(\dfrac{\overline{X}-p_0}{\sqrt{p_0(1-p_0)/n}} > z_\alpha\Bigr) \approx \alpha$ なので，$\dfrac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)/n}} > z_\alpha$ のとき，有意水準 α で H_0 : $\mu = \mu_0$ を棄却する。

3. 対立仮説 H_1 : $\mu \neq \mu_0$ (両側検定)

$P\Bigl(\left|\dfrac{\overline{X}-p_0}{\sqrt{p_0(1-p_0)/n}}\right| > z_{\alpha/2}\Bigr) \approx \alpha$ なので，$\left|\dfrac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)/n}}\right| > z_{\alpha/2}$ のとき，有意水準 α で H_0 を棄却する。

9　最小二乗法について

経済理論に基づいた線型モデルの係数の値をデータから求める時に用いられる手法 $\Longrightarrow$ 最小二乗法

9.1　最小二乗法と回帰直線

$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ のように n 組のデータがあり，X_i と Y_i との間に以下の線型関係を想定する。

$$Y_i = \alpha + \beta X_i,$$

X_i は説明変数，Y_i は被説明変数，α, β はパラメータとそれぞれ呼ばれる。

上の式は回帰モデル (または，回帰式) と呼ばれる。目的は，切片 α と傾き β をデータ $\{(X_i, Y_i), i = 1, 2, \cdots, n\}$ から推定すること，

データについて：

1. タイム・シリーズ (時系列)・データ：i が時間を表す (第 i 期)。
2. クロス・セクション (横断面)・データ：i が個人や企業を表す (第 i 番目の家計，第 i 番目の企業)。

9.2　切片 α と傾き β の推定

次のような関数 $S(\alpha,\beta)$ を定義する。

$$S(\alpha,\beta) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2$$

このとき，

$$\min_{\alpha,\beta} S(\alpha,\beta)$$

となるような α, β を求める (最小自乗法)。このときの解を $\hat{\alpha}, \hat{\beta}$ とする。

最小化のためには，

$$\frac{\partial S(\alpha,\beta)}{\partial \alpha} = 0$$

$$\frac{\partial S(\alpha,\beta)}{\partial \beta} = 0$$

を満たす α, β が $\hat{\alpha}, \hat{\beta}$ となる。

すなわち，$\hat{\alpha}, \hat{\beta}$ は，

$$\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0, \tag{1}$$

$$\sum_{i=1}^{n} X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0, \tag{2}$$

を満たす。

さらに，

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} X_i,$$
$$\sum_{i=1}^{n} X_i Y_i = \hat{\alpha} \sum_{i=1}^{n} X_i + \hat{\beta} \sum_{i=1}^{n} X_i^2, \tag{3}$$

行列表示によって，

$$\begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix},$$

逆行列の公式：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$\hat{\alpha}, \hat{\beta}$ について，まとめて，

$$
\begin{aligned}
&\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \\
&= \begin{pmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{pmatrix} \\
&= \frac{1}{n \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2} \\
&\quad \times \begin{pmatrix} \sum_{i=1}^{n} X_i^2 & -\sum_{i=1}^{n} X_i \\ -\sum_{i=1}^{n} X_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{pmatrix}
\end{aligned}
$$

さらに，$\hat{\beta}$ について解くと，

$$
\hat{\beta} = \frac{n \sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}
$$

$$= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

連立方程式の (3) 式から，

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

となる。ただし，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

とする。

数値例：　**以下の数値例を使って，回帰式 $Y_i = \alpha + \beta X_i$ の α，β の推定値 $\hat{\alpha}$，$\hat{\beta}$ を求める。**

i	Y_i	X_i
1	**6**	**10**
2	**9**	**12**
3	**10**	**14**
4	**10**	**16**

$\hat{\alpha}$，$\hat{\beta}$ **を求めるための公式は**

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

なので，必要なものは $\overline{X}$，$\overline{Y}$，$\sum_{i=1}^{n} X_i^2$，$\sum_{i=1}^{n} X_i Y_i$ である。

i	Y_i	X_i	$X_i Y_i$	X_i^2
1	**6**	**10**	**60**	**100**
2	**9**	**12**	**108**	**144**
3	**10**	**14**	**140**	**196**
4	**10**	**16**	**160**	**256**
合計	$\sum Y_i$	$\sum X_i$	$\sum X_i Y_i$	$\sum X_i^2$
	35	**52**	**468**	**696**
平均	$\overline{Y}$	$\overline{X}$		
	8.75	**13**		

よって，

$$
\begin{aligned}
\hat{\beta} &= \frac{468 - 4 \times 13 \times 8.75}{696 - 4 \times 13^2} = \frac{13}{20} = 0.65 \\
\hat{\alpha} &= 8.75 - 0.65 \times 13 = 0.3
\end{aligned}
$$

となる。

注意事項：

1. α, β は真の値で未知
2. $\hat{\alpha}, \hat{\beta}$ は α, β の推定値でデータから計算される

回帰直線は

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i,$$

として与えられる。

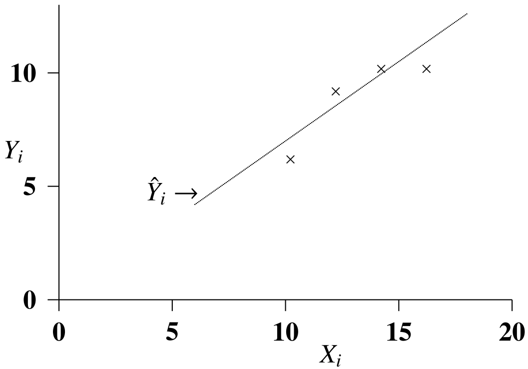
上の数値例では，

$$\hat{Y}_i = 0.3 + 0.65 X_i$$

となる。

i	Y_i	X_i	X_iY_i	X_i^2	$\hat{Y}_i$
1	**6**	**10**	**60**	**100**	**6.8**
2	**9**	**12**	**108**	**144**	**8.1**
3	**10**	**14**	**140**	**196**	**9.4**
4	**10**	**16**	**160**	**256**	**10.7**
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \hat{Y}_i$
	35	**52**	**468**	**696**	**35.0**
平均	$\overline{Y}$	$\overline{X}$			
	8.75	**13**			

図 2： Y_i, X_i, $\hat{Y}_i$



$\hat{Y}_i$ を実績値 Y_i の予測値または理論値と呼ぶ。

$$\hat{u}_i = Y_i - \hat{Y}_i,$$

$\hat{u}_i$ を残差と呼ぶ。

$$Y_i = \hat{Y}_i + \hat{u}_i = \hat{\alpha} + \hat{\beta}X_i + \hat{u}_i,$$

さらに，$\overline{Y}$ を両辺から引いて，

$$(Y_i - \overline{Y}) = (\hat{Y}_i - \overline{Y}) + \hat{u}_i,$$

9.3　残差 $\hat{u}_i$ の性質について

$\hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta}X_i$ に注意して，(1) 式から，

$$\sum_{i=1}^{n} \hat{u}_i = 0,$$