Testes para Média da População

Assumiremos que os dados são amostras de uma distribuição normal.

Apresentaremos procedimentos para testar hipóteses em μ (média da normal), utilizamos para isso o valor médio amostral $\bar{X}$ observado em uma amostra casual simples de tamanho n.

Exemplo

Em períodos de pico, os clientes de um banco são obrigados a enfrentar longas filas para sacar dinheiro nos caixas eletrônicos. Dados históricos de vários anos de operação indicam que o tempo de transação nesses caixas tem distribuição normal com média $\mu = 270$ segundos.

Para melhorar essa situação (diminuindo a permanência dos clientes nas filas) o banco resolve instalar, em caráter experimental, alguns caixas eletrônicos de concepção mais avançada. Após o período de experiência, o banco pretende examinar o tempo médio obtido em uma amostra casual simples das transações realizadas nesses caixas.

Exemplo

- Procedimento Geral:
  - Formular H_0 e H_1
  - Adotar critério de reconhecimento na amostra (nível α) de evidência contra H_0
  - Coletar dados e determinar o $\pi - valor$ "evidência contida nos dados"
  - Comparar se $\pi \leq \alpha$ então rejeitamos H_0 ao nível α

Exemplo. Teste de média com variância desconhecida

- Cálculo de π: supondo H_0 verdadeira, corresponde à probabilidade de acontecer um desvio entre o valor esperado da média amostral (μ) e o valor observado da média amostral ($\bar{X}$) tão grande quanto o observado.
- $H_0 : \mu = 270 \quad vs \quad H_1 : \mu < 270$

  (os novos caixas prometem diminuir o tempo de execução)
- dados: 240, 245, 286, 288, 238, 239, 278, 287, 291, 248, 257, 225, 257, 264, 282, 252, 243, 260, 248, 259, 262, 271, 234, 250
- Valor esperado $E(\bar{X}) = \mu_0 = 270$
- Valor observado $\bar{X} = 258.5$

Exemplo. Teste de média com variância desconhecida

- Metodologia: no lugar de medir a diferença entre μ e $\bar{X}$, $\bar{X} - \mu_0 = 258.2 - 270 = -11.5$, medimos a diferença registrada por $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$
- $S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$
- Resultado: se $X_1, ..., X_n$ é uma amostra de $N(\mu, \sigma^2)$, então $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ possui distribuição t-student com n-1 graus de liberdade
- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$
- Esse resultado permite garantir que se H_0 é verdadeira ($\mu_0 = 270$), logo $T_{\mu_0} = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$

Exemplo. Teste de média com variância desconhecida

- Cálculo do $\pi - valor$: $n = 24$, $\bar{X} = 258.5$, $\mu_0 = 270$, $S = 18.95$
  - $\pi = P\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \leq \frac{258.5-270}{18.95/\sqrt{24}}\right) = P(T \leq -2.97) = P(T \geq 2.97) = 1 - P(T \leq 2.97) = 1 - 0.996 = 0.004$
  - Seja $\alpha = 0.01$, como $\pi < \alpha$, então rejeitamos H_0 ao nível $\alpha = 0.01$

Generalizando. Teste de média com variância desconhecida

- $X_1, ..., X_n$ a.a. $N(\mu, \sigma^2)$
- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$

$$\begin{aligned}
\pi &= P\left(\frac{|\bar{X} - \mu_0|}{S/\sqrt{n}} \geq \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}}\right) \\
&= P\left(|T| \geq \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}}\right) \\
&= 2 \times P\left(T \geq \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}}\right) \\
&= 2 \times \left[1 - P\left(T \leq \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}}\right)\right]
\end{aligned}$$

Generalizando. Teste de média com variância desconhecida

- $X_1, ..., X_n$ a.a. $N(\mu, \sigma^2)$
- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$

$$\begin{aligned}
\pi &= P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right) \\
&= P\left(T \geq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right) \\
&= 1 - P\left(T \leq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)
\end{aligned}$$

Generalizando. Teste de média com variância desconhecida

- $X_1, \ldots, X_n$ a.a. $N(\mu, \sigma^2)$
- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$

$$\begin{aligned} \pi &= P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \leq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right) \\ &= P\left(T \leq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right) \end{aligned}$$

Aproveitando a estatística T para a construção de um intervalo de confiança para μ.

- $X_1, ..., X_n$ a.a. $N(\mu, \sigma^2)$, variância desconhecida (por isso usamos S)
- fixe um grau de confiança γ
- determine para uma variável T-student com $n-1$ graus de liberdade o valor $t_\gamma : \gamma = P(-t_\gamma \leq T \leq t_\gamma)$
- $\gamma = P\left(\bar{X} - t_\gamma \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_\gamma \frac{S}{\sqrt{n}}\right)$

Teste de média com variância conhecida

- Situação Alternativa:
  - $X_1, ..., X_n$ a.a. $N(\mu, \sigma^2)$
  - σ^2 **valor conhecido**
  - μ valor desconhecido (de interesse para conduzir o teste)
  - $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Generalizando

- $X_1, ..., X_n$ a.a. $N(\mu, \sigma^2)$, σ^2 conhecido
- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$

$$\begin{aligned}
\pi &= P\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \geq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right) \\
&= P\left(|Z| \geq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right) \\
&= 2 \times P\left(Z \geq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right) \\
&= 2 \times \left[1 - P\left(Z \leq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right)\right] \\
&= 2 \times \left[1 - \Phi\left(\frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right)\right]
\end{aligned}$$

Generalizando

- $X_1, \ldots, X_n$ a.a. $N(\mu, \sigma^2)$, σ^2 conhecido
- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$

$$
\begin{aligned}
\pi &= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= P\left(Z \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= 1 - P\left(Z \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= 1 - \Phi\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)
\end{aligned}
$$

Generalizando

- $X_1, ..., X_n$ a.a. $N(\mu, \sigma^2)$, σ^2 conhecido
- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$

$$\begin{aligned}
\pi &= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= P\left(Z \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) \\
&= \Phi\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

Exemplo

Um fabricante de sistema contra incêndios afirma que a verdadeira temperatura de ativação do sistema é $130^o F$ $(72^o C)$. Uma amostra de $n = 9$ sistemas produz uma temperatura amostral de ativação de $131.08^o F$. Suponha que a distribuição das temperaturas é uma normal com $\sigma = 1.5^o F$.

- $H_0 : \mu = 130 \quad vs \quad H_1 : \mu \neq 130$
- $\pi = P\left(\frac{|\bar{X} - 130|}{\sigma/\sqrt{n}} \geq \frac{|131.08 - 130|}{1.5/\sqrt{9}}\right) = 0.0308$
- Como $\pi = 0.0308 > \alpha = 0.01$, então não há evidência para rejeitar a afirmação do fabricante

Aproveitando a estatística Z para a construção de um intervalo de confiança para μ.

- $X_1, ..., X_n$ a.a. $N(\mu, \sigma^2)$, variância conhecida (por isso usamos σ^2)
- fixe um grau de confiança γ
- determine para uma variável $Z \sim$ N(0,1), o valor $z_\gamma : \gamma = P(-z_\gamma \leq Z \leq z_\gamma)$
- $\gamma = P\left(\bar{X} - z_\gamma \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_\gamma \frac{\sigma}{\sqrt{n}}\right)$

Teste para Comparar a Média entre 2 Populações Normais

Comparando dois vírus: O tempo de incubação do vírus 1 segue uma distribuição normal de desvio padrão $\sigma_1 = \sqrt{2}$. Por outro lado, o tempo de incubação do vírus 2 segue uma distribuição normal com desvio padrão $\sigma_2 = 1$. Os tempos de incubação de ambos os vírus são considerados independentes, e afirma-se que $\mu_1 = 3 + \mu_2$, onde μ_i é o verdadeiro tempo médio de incubação do vírus i, $i = 1, 2$. É realizado um estudo de controle e os tempos de incubação registrados foram (tempo em meses):

Teste para Comparar a Média entre 2 Populações Normais

- Vírus 1 - X (20 observações)

4.557275	3.722736	3.448939	2.861428	4.026696
4.083330	6.557597	4.306554	0.418356	5.562828
5.922411	2.654981	4.544995	4.044079	4.228879
6.238453	6.159126	5.465733	3.221166	2.275712

- Vírus 2 - Y (22 observações)

2.4425295	1.4914287	2.6786907	2.6040701	1.5131508
1.5990686	1.4671296	3.6954911	2.2256972	1.7821843
2.3612049	1.5617520	2.9785126	3.3317346	2.2222777
0.5768514	2.2649723	2.2604456	1.9195684	0.4961384
1.1703792	1.6995109			

Teste para Comparar a Média entre 2 Populações Normais

- Suspeita-se que há evidência contra $\mu_1 = 3 + \mu_2$
- Os dados suportam a rejeição de $\mu_1 = 3 + \mu_2$?
  - $\bar{x} = 4.216564$
  - $\bar{y} = 2.015581$

Teste para Comparar a Média entre 2 Populações Normais

- Em geral:
  - $X_1, ..., X_m$ população 1 - $N(\mu_1, \sigma_1^2)$, σ_1^2 conhecido
  - $Y_1, ..., Y_n$ população 2 - $N(\mu_2, \sigma_2^2)$, σ_2^2 conhecido
  - $H_0 : \mu_1 - \mu_2 = \Delta_0 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq \Delta_0$
  - Metodologia: o representante natural da diferença $\mu_1 - \mu_2$ é $\bar{X} - \bar{Y}$

$$Z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

  - Resultado: sob as condições acima, se a população de X é independente da população de Y, $Z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1)$

Teste para Comparar a Média entre 2 Populações Normais

- Retomando o exemplo dos vírus:
  - $H_0 : \mu_1 - \mu_2 = 3 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq 3$
  - $m = 20,\ n = 22,\ \bar{x} = 4.216564,\ \bar{y} = 2.015581,\ \sigma_1 = \sqrt{2},\ \sigma_2 = 1$

$$
\begin{aligned}
\pi &= P\left(\frac{|\bar{X} - \bar{Y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \geq \frac{|4.216564 - 2.015581 - 3|}{\sqrt{\frac{2}{20} + \frac{1}{22}}}\right) \\
&= P\left(|Z| \geq 2.095\right) = 2 \times P\left(Z \geq 2.095\right) \\
&= 2 \times \left[1 - \Phi\left(2.095\right)\right] = 2 \times \left[1 - 0.9817\right] = 0.0366
\end{aligned}
$$

para $\alpha = 0.01$, como $\pi = 0.0366 > \alpha = 0.01$, não temos evidência para rejeitar $H_0 : \mu_1 = 3 + \mu_2$

Teste para Comparar a Média entre 2 Populações Normais

- Seja $\{X_i\}_{i=1}^{m}$ independente de $\{Y_i\}_{i=1}^{n}$
- $X_i \sim N(\mu_1, \sigma_1^2)$, σ_1 conhecido
- $\bar{x}$ a média amostral
- $Y_i \sim N(\mu_2, \sigma_2^2)$, σ_2 conhecido
- $\bar{y}$ a média amostral

Teste para Comparar a Média entre 2 Populações Normais

- $H_0 : \mu_1 - \mu_2 = \Delta_0 \quad vs \quad H_1 : \mu_1 - \mu_2 > \Delta_0$

$$
\begin{aligned}
\pi &= P\left(\frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \geq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right) \\
&= P\left(Z \geq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right) \\
&= 1 - \Phi\left(\frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right)
\end{aligned}
$$

Teste para Comparar a Média entre 2 Populações Normais

- $H_0 : \mu_1 - \mu_2 = \Delta_0 \quad vs \quad H_1 : \mu_1 - \mu_2 < \Delta_0$

$$
\begin{aligned}
\pi &= P\left(\frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \leq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right) \\
&= P\left(Z \leq \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right) \\
&= \Phi\left(\frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right)
\end{aligned}
$$

Teste para Comparar a Média entre 2 Populações Normais

- $H_0 : \mu_1 - \mu_2 = \Delta_0 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq \Delta_0$

$$
\begin{aligned}
\pi &= P\left(\frac{|\bar{X} - \bar{Y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \geq \frac{|\bar{x} - \bar{y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \right) \\
&= P\left(|Z| \geq \frac{|\bar{x} - \bar{y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \right) \\
&= 2 \times \left[1 - \Phi\left(\frac{|\bar{x} - \bar{y} - \Delta_0|}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \right) \right]
\end{aligned}
$$

Aproveitando a estatística Z para a construção de um intervalo de confiança para Δ_0.

- Seja $\{X_i\}_{i=1}^{m}$ independente de $\{Y_i\}_{i=1}^{n}$
- $X_i \sim N(\mu_1, \sigma_1^2)$, σ_1 conhecido $\bar{x}$ a média amostral
- $Y_i \sim N(\mu_2, \sigma_2^2)$, σ_2 conhecido $\bar{y}$ a média amostral
- $\Delta_0 = \mu_1 - \mu_2$
- fixe um grau de confiança γ e determine para $Z \sim$ N(0,1), o valor $z_\gamma : \gamma = P(-z_\gamma \leq Z \leq z_\gamma)$
- $\gamma = P\left(\bar{X} - \bar{Y} - z_\gamma\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \leq \Delta_0 \leq \bar{X} - \bar{Y} + z_\gamma\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right)$