(Material cedido por completo pelo professor Alessandro do DEST)

Inferência estatística frequentista

Prof. Bartolomeu Zamprogno

Departamento de Estatística
Centro de Ciências Exatas
Universidade Federal do Espírito Santo

Estrutura

Estrutura

Introdução

- Interesse: como uma variável se comporta em determinada população;

Introdução

- Interesse: como uma variável se comporta em determinada população;
- Problema: o acesso a população por completo é inviável ou impossível;

Introdução

- Interesse: como uma variável se comporta em determinada população;
- Problema: o acesso a população por completo é inviável ou impossível;
- A distribuição verdadeira da variável é desconhecida;

Introdução

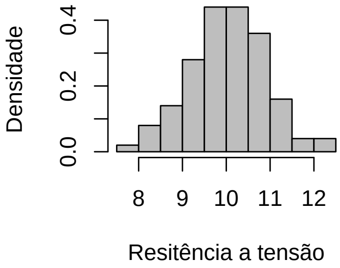
- Interesse: como uma variável se comporta em determinada população;
- Problema: o acesso a população por completo é inviável ou impossível;
- A distribuição verdadeira da variável é desconhecida;
- Solução:
  - propor um modelo probabilístico dependente de um, ou mais, parâmetros;

Introdução

- Interesse: como uma variável se comporta em determinada população;
- Problema: o acesso a população por completo é inviável ou impossível;
- A distribuição verdadeira da variável é desconhecida;
- Solução:
  - propor um modelo probabilístico dependente de um, ou mais, parâmetros;
  - através de uma amostra **estimar** e **testar hipóteses** com respeito a esse(s) parâmetro(s).

Exemplo 1

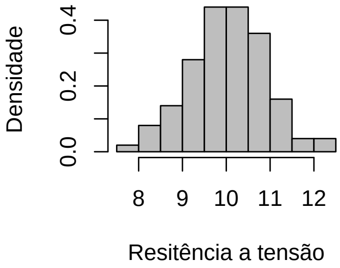
- Seja X a resitência a tensão de um componente utilizado em um chassi de automóvel;
- Suponha que n observações de X foram feitas e forneceram os valores $x_1, x_2, \dots, x_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?

Exemplo 1

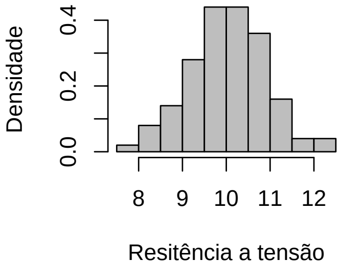
- Seja X a resitência a tensão de um componente utilizado em um chassi de automóvel;
- Suponha que n observações de X foram feitas e forneceram os valores $x_1, x_2, \ldots, x_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?
- O modelo $N(\mu, \sigma^2)$ parece ser adequado!

Exemplo 1

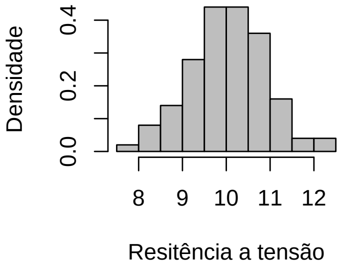
- Seja X a resitência a tensão de um componente utilizado em um chassi de automóvel;
- Suponha que n observações de X foram feitas e forneceram os valores $x_1, x_2, \ldots, x_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?
- O modelo $N(\mu, \sigma^2)$ parece ser adequado!
- Quais valores de μ e σ^2 são apropriados?

Exemplo 1

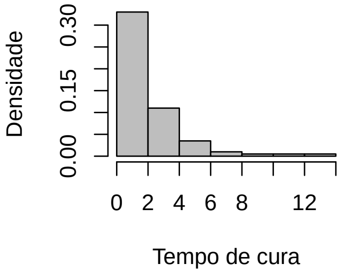
- Seja X a resitência a tensão de um componente utilizado em um chassi de automóvel;
- Suponha que n observações de X foram feitas e forneceram os valores $x_1, x_2, \ldots, x_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?
- O modelo $N(\mu, \sigma^2)$ parece ser adequado!
- Quais valores de μ e σ^2 são apropriados?
- Estimar μ e σ^2!

Exemplo 2

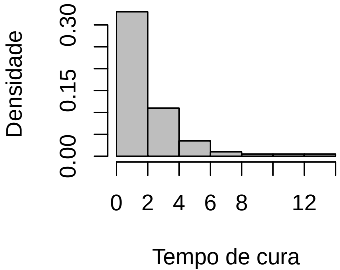
- Seja T o tempo de cura de um paciente com determinada doença após a administração de determinado medicamento.;
- Suponha que n observações de T foram feitas e forneceram os valores $t_1, t_2, \ldots, t_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?

Exemplo 2

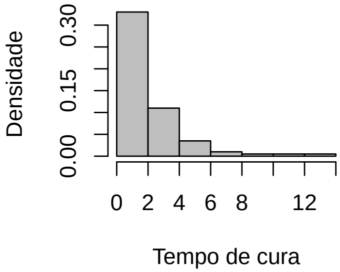
- Seja T o tempo de cura de um paciente com determinada doença após a administração de determinado medicamento.;
- Suponha que n observações de T foram feitas e forneceram os valores $t_1, t_2, \ldots, t_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?
- O modelo $Exp(\lambda)$ parece ser adequado!

Exemplo 2

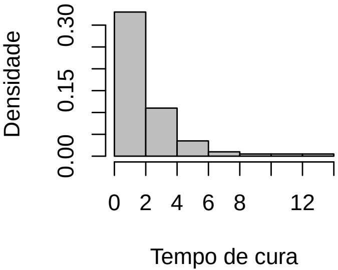
- Seja T o tempo de cura de um paciente com determinada doença após a administração de determinado medicamento.;
- Suponha que n observações de T foram feitas e forneceram os valores $t_1, t_2, \ldots, t_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?
- O modelo $Exp(\lambda)$ parece ser adequado!
- Qual valor de λ é apropriado?

Exemplo 2

- Seja T o tempo de cura de um paciente com determinada doença após a administração de determinado medicamento.;
- Suponha que n observações de T foram feitas e forneceram os valores $t_1, t_2, \ldots, t_n$;
- O histograma das observações realizadas é apresentado abaixo:



- Qual modelo probabilístico conjecturar para esses dados?
- O modelo $Exp(\lambda)$ parece ser adequado!
- Qual valor de λ é apropriado?
- Estimar λ!

Diferenças entre estatística descritiva e inferencial

Estatística descritiva

- Os dados já foram observados (pós-experimental);

Diferenças entre estatística descritiva e inferencial

Estatística descritiva

- Os dados já foram observados (pós-experimental);
- As ferramentas não permitem a extrapolação dos resultados para toda a população.

Diferenças entre estatística descritiva e inferencial

Estatística descritiva

- Os dados já foram observados (pós-experimental);
- As ferramentas não permitem a extrapolação dos resultados para toda a população.

Inferência estatística

- Raciocínio pré-experimental, estabelecendo um modelo para as observações (como se os dados não tivessem sido observados);

Diferenças entre estatística descritiva e inferencial

Estatística descritiva

- Os dados já foram observados (pós-experimental);
- As ferramentas não permitem a extrapolação dos resultados para toda a população.

Inferência estatística

- Raciocínio pré-experimental, estabelecendo um modelo para as observações (como se os dados não tivessem sido observados);
- Os resultados obtidos podem ser extrapolados e fornecerem conclusões para toda a população.

Estrutura

Estrutura

Parâmetros

- Seja X v.a. com distribuição de probabilidades f (função de probabilidades se X é discreta e função de densidade se X é contínua);

Parâmetros

- Seja X v.a. com distribuição de probabilidades f (função de probabilidades se X é discreta e função de densidade se X é contínua);
- Um **parâmetro** é qualquer quantidade **constante** que dependa de f.

Exemplos:

- Seja X v.a. com distribuição de probabilidades f (função de probabilidades se X é discreta e função de densidade se X é contínua);
- Um **parâmetro** é qualquer quantidade **constante** que dependa de f.

Exemplos:

- A probabilidade de sucesso p de um ensaio de Bernoulli;

Parâmetros

- Seja X v.a. com distribuição de probabilidades f (função de probabilidades se X é discreta e função de densidade se X é contínua);
- Um **parâmetro** é qualquer quantidade **constante** que dependa de f.

Exemplos:

- A probabilidade de sucesso p de um ensaio de Bernoulli;
- O valor esperado $\mu = E(X)$ de uma v.a. X;

Parâmetros

- Seja X v.a. com distribuição de probabilidades f (função de probabilidades se X é discreta e função de densidade se X é contínua);
- Um **parâmetro** é qualquer quantidade **constante** que dependa de f.

Exemplos:

- A probabilidade de sucesso p de um ensaio de Bernoulli;
- O valor esperado $\mu = E(X)$ de uma v.a. X;
- A variância $\sigma^2 = Var(X)$ de uma v.a. X;

Parâmetros

- Seja X v.a. com distribuição de probabilidades f (função de probabilidades se X é discreta e função de densidade se X é contínua);
- Um **parâmetro** é qualquer quantidade **constante** que dependa de f.

Exemplos:

- A probabilidade de sucesso p de um ensaio de Bernoulli;
- O valor esperado $\mu = E(X)$ de uma v.a. X;
- A variância $\sigma^2 = Var(X)$ de uma v.a. X;
- A taxa de ocorrências λ de uma Poisson;

Parâmetros

- Seja X v.a. com distribuição de probabilidades f (função de probabilidades se X é discreta e função de densidade se X é contínua);
- Um **parâmetro** é qualquer quantidade **constante** que dependa de f.

Exemplos:

- A probabilidade de sucesso p de um ensaio de Bernoulli;
- O valor esperado $\mu = E(X)$ de uma v.a. X;
- A variância $\sigma^2 = Var(X)$ de uma v.a. X;
- A taxa de ocorrências λ de uma Poisson;
- O mínimo a e o máximo b de uma v.a. $X \sim U[a, b]$.

Estrutura

Amostra

- A população não pode ser completamente observada;

Amostra

- A população não pode ser completamente observada;
- Embora constantes, os parâmetros são desconhecidos!

Amostra

- A população não pode ser completamente observada;
- Embora constantes, os parâmetros são desconhecidos!
- Uma **amostra** pode ser utilizada para aproximar os parâmetros;

Amostra

- A população não pode ser completamente observada;
- Embora constantes, os parâmetros são desconhecidos!
- Uma **amostra** pode ser utilizada para aproximar os parâmetros;
- Uma amostra selecionada por conveniência pode acarretar os seguintes problemas:

Amostra

- A população não pode ser completamente observada;
- Embora constantes, os parâmetros são desconhecidos!
- Uma **amostra** pode ser utilizada para aproximar os parâmetros;
- Uma amostra selecionada por conveniência pode acarretar os seguintes problemas:
    - resultados tendenciosos;

Amostra

- A população não pode ser completamente observada;
- Embora constantes, os parâmetros são desconhecidos!
- Uma **amostra** pode ser utilizada para aproximar os parâmetros;
- Uma amostra selecionada por conveniência pode acarretar os seguintes problemas:
  - resultados tendenciosos;
  - o comportamento probabilístico desconhecido;

- A população não pode ser completamente observada;
- Embora constantes, os parâmetros são desconhecidos!
- Uma **amostra** pode ser utilizada para aproximar os parâmetros;
- Uma amostra selecionada por conveniência pode acarretar os seguintes problemas:
  - resultados tendenciosos;
  - o comportamento probabilístico desconhecido;
- Necessidade de uma amostra representativa;

Amostra

- A população não pode ser completamente observada;
- Embora constantes, os parâmetros são desconhecidos!
- Uma **amostra** pode ser utilizada para aproximar os parâmetros;
- Uma amostra selecionada por conveniência pode acarretar os seguintes problemas:
  - resultados tendenciosos;
  - o comportamento probabilístico desconhecido;
- Necessidade de uma amostra representativa;
- Principal solução $\rightarrow$ mecanismo aleatório para selecionar a amostra.

Estrutura

Amostra aleatória

Definição

Seja X uma v.a. Uma amostra aleatória (a.a.) de X é um conjunto de observações independentes entre si e com a mesma distribuição de X.

Amostra aleatória

Definição

Seja X uma v.a. Uma amostra aleatória (a.a.) de X é um conjunto de observações independentes entre si e com a mesma distribuição de X.

- Uma a.a. é caracterizada pré-experimentalmente;

Amostra aleatória

Definição

Seja X uma v.a. Uma amostra aleatória (a.a.) de X é um conjunto de observações independentes entre si e com a mesma distribuição de X.

- Uma a.a. é caracterizada pré-experimentalmente;
- As observações ainda não foram coletadas;

Amostra aleatória

Definição

Seja X uma v.a. Uma amostra aleatória (a.a.) de X é um conjunto de observações independentes entre si e com a mesma distribuição de X.

- Uma a.a. é caracterizada pré-experimentalmente;
- As observações ainda não foram coletadas;
- Portanto, uma a.a. é um conjunto de **v.a.**;

Amostra aleatória

Definição

Seja X uma v.a. Uma amostra aleatória (a.a.) de X é um conjunto de observações independentes entre si e com a mesma distribuição de X.

- Uma a.a. é caracterizada pré-experimentalmente;
- As observações ainda não foram coletadas;
- Portanto, uma a.a. é um conjunto de **v.a.**;
- Uma a.a. de X de tamanho n é denotada $X_1, \ldots, X_n$.

Amostra aleatória

Definição

Seja X uma v.a. Uma amostra aleatória (a.a.) de X é um conjunto de observações independentes entre si e com a mesma distribuição de X.

- Uma a.a. é caracterizada pré-experimentalmente;
- As observações ainda não foram coletadas;
- Portanto, uma a.a. é um conjunto de **v.a.**;
- Uma a.a. de X de tamanho n é denotada $X_1, \ldots, X_n$.

Exercício: Em uma população de N indivíduos, K possuem determinada característica. Suponha que desejamos selecionar n indíviduos para inferir sobre a verdadeira proporção K/N de indivíduos com essa característica. Uma a.a. seria obtida se realizassemos o sorteio dos indíviduos com ou sem reposição?

Estrutura

Exemplo

Considere um experimento de Bernoulli com probabilidade de sucesso p (desconhecida) e seja X a v.a. que associa o valor 1 se esse experimento resulta sucesso e 0 caso contrário. Dizemos que X é uma v.a. de Bernoulli com parâmetro p e escrevemos $X \sim Ber(p)$. Note que também temos $X \sim B(1, p)$.

Exemplo

Considere um experimento de Bernoulli com probabilidade de sucesso p (desconhecida) e seja X a v.a. que associa o valor 1 se esse experimento resulta sucesso e 0 caso contrário. Dizemos que X é uma v.a. de Bernoulli com parâmetro p e escrevemos $X \sim Ber(p)$. Note que também temos $X \sim B(1, p)$.

- Como tirar conclusões sobre a probabilidade de sucesso p?

Exemplo

Considere um experimento de Bernoulli com probabilidade de sucesso p (desconhecida) e seja X a v.a. que associa o valor 1 se esse experimento resulta sucesso e 0 caso contrário. Dizemos que X é uma v.a. de Bernoulli com parâmetro p e escrevemos $X \sim Ber(p)$. Note que também temos $X \sim B(1, p)$.

- Como tirar conclusões sobre a probabilidade de sucesso p?
- Podemos coletar uma a.a. $X_1, \ldots, X_n$ de X;

Exemplo

Considere um experimento de Bernoulli com probabilidade de sucesso p (desconhecida) e seja X a v.a. que associa o valor 1 se esse experimento resulta sucesso e 0 caso contrário. Dizemos que X é uma v.a. de Bernoulli com parâmetro p e escrevemos $X \sim Ber(p)$. Note que também temos $X \sim B(1, p)$.

- Como tirar conclusões sobre a probabilidade de sucesso p?
- Podemos coletar uma a.a. $X_1, \ldots, X_n$ de X;
- Após coletar a amostra podemos calcular a **proporção amostral** $\hat{p}$;

Exemplo

Considere um experimento de Bernoulli com probabilidade de sucesso p (desconhecida) e seja X a v.a. que associa o valor 1 se esse experimento resulta sucesso e 0 caso contrário. Dizemos que X é uma v.a. de Bernoulli com parâmetro p e escrevemos $X \sim Ber(p)$. Note que também temos $X \sim B(1, p)$.

- Como tirar conclusões sobre a probabilidade de sucesso p?
- Podemos coletar uma a.a. $X_1, \ldots, X_n$ de X;
- Após coletar a amostra podemos calcular a **proporção amostral** $\hat{p}$;
- Podemos usar o valor que $\hat{p}$ **fornecerá** como aproximação de p;

Exemplo

Considere um experimento de Bernoulli com probabilidade de sucesso p (desconhecida) e seja X a v.a. que associa o valor 1 se esse experimento resulta sucesso e 0 caso contrário. Dizemos que X é uma v.a. de Bernoulli com parâmetro p e escrevemos $X \sim Ber(p)$. Note que também temos $X \sim B(1, p)$.

- Como tirar conclusões sobre a probabilidade de sucesso p?
- Podemos coletar uma a.a. $X_1, \ldots, X_n$ de X;
- Após coletar a amostra podemos calcular a **proporção amostral** $\hat{p}$;
- Podemos usar o valor que $\hat{p}$ **fornecerá** como aproximação de p;
- Mas, antes de observar a a.a., $\hat{p}$ é v.a., pois é função de v.a.'s;

Exemplo

Considere um experimento de Bernoulli com probabilidade de sucesso p (desconhecida) e seja X a v.a. que associa o valor 1 se esse experimento resulta sucesso e 0 caso contrário. Dizemos que X é uma v.a. de Bernoulli com parâmetro p e escrevemos $X \sim Ber(p)$. Note que também temos $X \sim B(1, p)$.

- Como tirar conclusões sobre a probabilidade de sucesso p?
- Podemos coletar uma a.a. $X_1, \ldots, X_n$ de X;
- Após coletar a amostra podemos calcular a **proporção amostral** $\hat{p}$;
- Podemos usar o valor que $\hat{p}$ **fornecerá** como aproximação de p;
- Mas, antes de observar a a.a., $\hat{p}$ é v.a., pois é função de v.a.'s;
- Como dizer se $\hat{p}$ será uma boa aproximação para p sem conhecer p (e antes de obervar a amostra)?

Estimador

Definição

Sejam X v.a. cuja distribuição depende de um parâmetro θ e $X_1, \ldots, X_n$ a.a. de tamanho n de X. Um **estimador** $\hat{\theta}$ *de θ é qualquer função da amostra $X_1, \ldots, X_n$ utilizada para aproximar (estimar) o valor de θ. Isto é,*

$$\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n).$$

- Suponha que a amostra observada (ponto amostral) foi $x_1, \ldots, x_n$, isto é, que

$$X_1 = x_1, \ldots, X_n = x_n.$$

A realização de $\hat{\theta}$, $\hat{\theta}(x_1, \ldots, x_n)$, é denominada **estimativa**.

Exemplos de estimadores

Dada uma a.a. $X_1, \ldots, X_n$, alguns exemplos de estimadores são:

- O total $\sum_{i=1}^{n} X_i$;

Exemplos de estimadores

Dada uma a.a. $X_1, \ldots, X_n$, alguns exemplos de estimadores são:

- O total $\sum_{i=1}^{n} X_i$;
- A média $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$;

Exemplos de estimadores

Dada uma a.a. $X_1, \ldots, X_n$, alguns exemplos de estimadores são:

- O total $\sum_{i=1}^{n} X_i$;
- A média $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$;
- A variância (viciada) $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$;

Exemplos de estimadores

Dada uma a.a. $X_1, \ldots, X_n$, alguns exemplos de estimadores são:

- O total $\sum_{i=1}^{n} X_i$;
- A média $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$;
- A variância (viciada) $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$;
- A variância (não viciada) $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$;

Exemplos de estimadores

Dada uma a.a. $X_1, \ldots, X_n$, alguns exemplos de estimadores são:

- O total $\sum_{i=1}^{n} X_i$;
- A média $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$;
- A variância (viciada) $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$;
- A variância (não viciada) $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$;
- O mínimo $X_{(1)} = \min\{X_1, \ldots, X_n\}$;

Exemplos de estimadores

Dada uma a.a. $X_1, \dots, X_n$, alguns exemplos de estimadores são:

- O total $\sum_{i=1}^{n} X_i$;
- A média $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$;
- A variância (viciada) $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$;
- A variância (não viciada) $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$;
- O mínimo $X_{(1)} = \min\{X_1, \dots, X_n\}$;
- O máximo $X_{(n)} = \max\{X_1, \dots, X_n\}$;

Exemplos de estimadores

Dada uma a.a. $X_1, \ldots, X_n$, alguns exemplos de estimadores são:

- O total $\sum_{i=1}^{n} X_i$;
- A média $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$;
- A variância (viciada) $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$;
- A variância (não viciada) $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$;
- O mínimo $X_{(1)} = \min\{X_1, \ldots, X_n\}$;
- O máximo $X_{(n)} = \max\{X_1, \ldots, X_n\}$;
- A proporção $\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$ (se a a.a. for de uma v.a. Bernoulli).

Estrutura

Distribuição amostral

- Estimadores $\rightarrow$ pré-observacional $\rightarrow$ são v.a.'s;

Distribuição amostral

- Estimadores $\rightarrow$ pré-observacional $\rightarrow$ são v.a.'s;
- Estimadores tem distribuição de probabilidades denominada **distribuição amostral**.

Distribuição amostral

- Estimadores $\rightarrow$ pré-observacional $\rightarrow$ são v.a.'s;
- Estimadores tem distribuição de probabilidades denominada **distribuição amostral**.

Definição

Seja X v.a. com distribuição de probabilidades dependente de um parâmetro θ. Seja $\hat{\theta}$ um estimador de θ. A distribuição amostral de $\hat{\theta}$ é a distribuição de probabilidades da v.a. (estimador) $\hat{\theta}$.

Distribuição amostral

- Estimadores → pré-observacional → são v.a.'s;
- Estimadores tem distribuição de probabilidades denominada **distribuição amostral**.

Definição

Seja X v.a. com distribuição de probabilidades dependente de um parâmetro θ. Seja $\hat{\theta}$ um estimador de θ. A distribuição amostral de $\hat{\theta}$ é a distribuição de probabilidades da v.a. (estimador) $\hat{\theta}$.

- A distribuição amostral depende principalmente:

Distribuição amostral

- Estimadores → pré-observacional → são v.a.'s;
- Estimadores tem distribuição de probabilidades denominada **distribuição amostral**.

Definição

Seja X v.a. com distribuição de probabilidades dependente de um parâmetro θ. Seja $\hat{\theta}$ um estimador de θ. A distribuição amostral de $\hat{\theta}$ é a distribuição de probabilidades da v.a. (estimador) $\hat{\theta}$.

- A distribuição amostral depende principalmente:
  - da distribuição da v.a. X cuja amostra $X_1, \ldots, X_n$ foi coletada;

Distribuição amostral

- Estimadores → pré-observacional → são v.a.'s;
- Estimadores tem distribuição de probabilidades denominada **distribuição amostral**.

Definição

Seja X v.a. com distribuição de probabilidades dependente de um parâmetro θ. Seja $\hat{\theta}$ um estimador de θ. A distribuição amostral de $\hat{\theta}$ é a distribuição de probabilidades da v.a. (estimador) $\hat{\theta}$.

- A distribuição amostral depende principalmente:
  - da distribuição da v.a. X cuja amostra $X_1, \ldots, X_n$ foi coletada;
  - do tipo de amostragem realizada.

Exemplo

Suponha nove indivíduos: dois com nenhum filho; três com um filho; e quatro com dois filhos. Denote X como o número de filhos de um indivíduo selecionado ao acaso dessa população.

Exemplo

Suponha nove indivíduos: dois com nenhum filho; três com um filho; e quatro com dois filhos. Denote X como o número de filhos de um indivíduo selecionado ao acaso dessa população. Temos que a função de probabilidades de X é dada por:

X	0	1	2
P_X	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{4}{9}$

Exemplo

Suponha nove indivíduos: dois com nenhum filho; três com um filho; e quatro com dois filhos. Denote X como o número de filhos de um indivíduo selecionado ao acaso dessa população. Temos que a função de probabilidades de X é dada por:

X	0	1	2
P_X	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{4}{9}$

O valor esperado de X é $\mu = E(X) = \frac{11}{9} = 1.22$.

Exemplo

Suponha nove indivíduos: dois com nenhum filho; três com um filho; e quatro com dois filhos. Denote X como o número de filhos de um indivíduo selecionado ao acaso dessa população. Temos que a função de probabilidades de X é dada por:

X	0	1	2
P_X	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{4}{9}$

O valor esperado de X é $\mu = E(X) = \frac{11}{9} = 1.22$.
Um pesquisador desconhece a distribuição de X e o valor de μ. Para inferir sobre μ, esse pesquisador:

- coletará uma amostra **com reposição** de X de tamanho $n = 2$: X_1, X_2;
- utilizará $\bar{X} = \frac{X_1+X_2}{2}$ para estimar μ.

Encontremos a distribuição amostral de $\bar{X}$.

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0, 0)$	$4/81$	0

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0,0)$	$4/81$	0
$(0,1)$	$2/27$	0.5

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0, 0)$	$4/81$	0
$(0, 1)$	$2/27$	0.5
$(0, 2)$	$8/81$	1

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0,0)$	$4/81$	0
$(0,1)$	$2/27$	0.5
$(0,2)$	$8/81$	1
$(1,0)$	$2/27$	0.5

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0, 0)$	$4/81$	0
$(0, 1)$	$2/27$	0.5
$(0, 2)$	$8/81$	1
$(1, 0)$	$2/27$	0.5
$(1, 1)$	$1/9$	1

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0, 0)$	$4/81$	0
$(0, 1)$	$2/27$	0.5
$(0, 2)$	$8/81$	1
$(1, 0)$	$2/27$	0.5
$(1, 1)$	$1/9$	1
$(1, 2)$	$4/27$	1.5

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0, 0)$	$4/81$	0
$(0, 1)$	$2/27$	0.5
$(0, 2)$	$8/81$	1
$(1, 0)$	$2/27$	0.5
$(1, 1)$	$1/9$	1
$(1, 2)$	$4/27$	1.5
$(2, 0)$	$8/81$	1

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0, 0)$	$4/81$	0
$(0, 1)$	$2/27$	0.5
$(0, 2)$	$8/81$	1
$(1, 0)$	$2/27$	0.5
$(1, 1)$	$1/9$	1
$(1, 2)$	$4/27$	1.5
$(2, 0)$	$8/81$	1
$(2, 1)$	$4/27$	1.5

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0,0)$	$4/81$	0
$(0,1)$	$2/27$	0.5
$(0,2)$	$8/81$	1
$(1,0)$	$2/27$	0.5
$(1,1)$	$1/9$	1
$(1,2)$	$4/27$	1.5
$(2,0)$	$8/81$	1
$(2,1)$	$4/27$	1.5
$(2,2)$	$16/81$	2
Total	1	-

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0,0)$	$4/81$	0
$(0,1)$	$2/27$	0.5
$(0,2)$	$8/81$	1
$(1,0)$	$2/27$	0.5
$(1,1)$	$1/9$	1
$(1,2)$	$4/27$	1.5
$(2,0)$	$8/81$	1
$(2,1)$	$4/27$	1.5
$(2,2)$	$16/81$	2
Total	1	-

Distribuição de $\bar{X}$ é:

$\bar{X}$	0	0.5	1	1.5	2
$P_{\bar{X}}$	$\frac{4}{81}$	$\frac{4}{27}$	$\frac{25}{81}$	$\frac{8}{27}$	$\frac{16}{81}$

Exemplo - cont.

A tabela abaixo mostra os possíveis **pontos amostrais**; suas respectivas probabilidades e as estimativas geradas por esses pontos.

(x_1, x_2)	$P_{X_1,X_2}(x_1, x_2)$	$\bar{x}$
$(0, 0)$	$4/81$	0
$(0, 1)$	$2/27$	0.5
$(0, 2)$	$8/81$	1
$(1, 0)$	$2/27$	0.5
$(1, 1)$	$1/9$	1
$(1, 2)$	$4/27$	1.5
$(2, 0)$	$8/81$	1
$(2, 1)$	$4/27$	1.5
$(2, 2)$	$16/81$	2
Total	1	-

Distribuição de $\bar{X}$ é:

$\bar{X}$	0	0.5	1	1.5	2
$P_{\bar{X}}$	$\frac{4}{81}$	$\frac{4}{27}$	$\frac{25}{81}$	$\frac{8}{27}$	$\frac{16}{81}$

O valor esperado de $\bar{X}$ é:

$$\begin{aligned} \mu_{\bar{X}} &= \frac{2}{27} + \frac{25}{81} + \frac{12}{27} + \frac{32}{81} \\ &= \frac{1}{81}(6 + 25 + 36 + 32) \\ &= \frac{99}{81} = \frac{11}{9} = \mu \end{aligned}$$

Exemplo - cont.

Exercício: Refaça o exemplo anterior com amostragem **sem** reposição para verificar que a distribuição amostral de $\bar{X}$ é modificada. Verifique a distribuição amostral de S^2.

- A diferença na distribuição amostral provocada por amostragem com e sem reposição é menos evidente quando o tamanho da população é grande;
- Esse efeito é nulo para populações infinitas;
- No exemplo anterior a amostragem com reposição fornece uma a.a;
- De agora em diante consideraremos apenas amostras aleatórias.

Estrutura

Consistência e distribuição amostral assintótica

Teorema

(Populações infinitas) Seja $X_1, \ldots, X_n$ *a.a. de* X*, onde* $E(X) = \mu$ *e* $Var(X) = \sigma^2 < \infty$*. Então* $E(\bar{X}) = \mu$ *e* $Var(\bar{X}) = \frac{\sigma^2}{n}$*, onde* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *é a média amostral.*

Teorema

(Teorema Central do Limite - TCL) Seja $X_1, \ldots, X_n$ *a.a. de* X*, onde* $E(X) = \mu$ *e* $Var(X) = \sigma^2 < \infty$*. Então a v.a.*

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

tem distribuição aproximadamente normal padrão. A aproximação fica melhor com o aumento de n*.*

Corolário

(Populações infinitas) Seja $X_1, \ldots, X_n$ *a.a. de* $X \sim Ben(p)$*. Então, se* $\hat{p}$ *for a proporção amostral,*

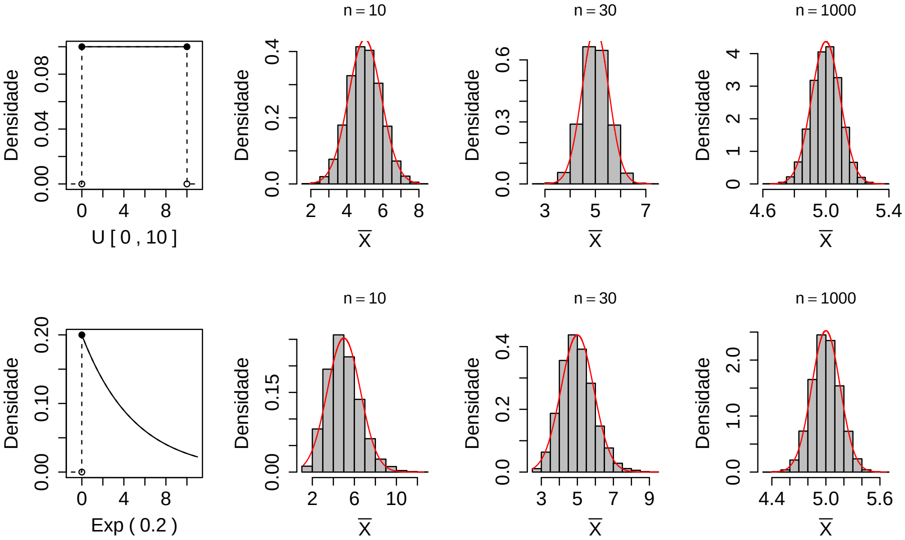
$$Z_n = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

tem distribuição aproximadamente normal-padrão. A aproximação fica melhor com o aumento de n*.*

Comentários

- A média amostral tem distribuição aproximadamente normal com parâmetros μ e $\frac{\sigma^2}{n}$;
- A aproximação se torna melhor com o aumento do tamanho amostral n;
- O valor de n necessário para que a aproximação seja boa depende da distribuição dos dados;
- Dados provenientes de distribuições assimétricas exigem o valor de n maior.

Número de amostras coletadas: 10000.



Exemplo

Seja T o tempo em meses que uma lâmpada produzida por determinada empresa dura até queimar. Suponha que $T \sim Exp(\lambda)$, onde $\lambda = 0.25$. Uma sala é iluminada por essa lâmpada. Caso queime uma lâmpada ela é imediatamente trocada. Se há 36 dessas lâmpadas no estoque. Qual a probabilidade que não haja mais lâmpadas em estoque em menos de 9 anos?

Primeiramente:

- $\mu = E(T) = \frac{1}{\lambda} = \frac{1}{0.25} = 4;$
- $\sigma^2 = Var(T) = \frac{1}{\lambda^2} = 16.$

Portanto, temos que

$$\begin{aligned} P\left(\sum_{i=1}^{36} T_i < 9 \times 12\right) &= P(\bar{T} < 3) = P\left(\frac{\bar{T} - \mu}{\sigma/\sqrt{n}} < \frac{3-4}{4/\sqrt{36}}\right) \\ &= P(Z_n < -1.5) \approx P(Z < -1.5) \\ &= 0.05591 \end{aligned}$$

A aproximação acima se deve ao TCL.

Exemplo

Seja T o tempo em meses que uma lâmpada produzida por determinada empresa dura até queimar. Suponha que $T \sim Exp(\lambda)$, onde $\lambda = 0.25$. Uma sala é iluminada por essa lâmpada. Caso queime uma lâmpada ela é imediatamente trocada. Se há 36 dessas lâmpadas no estoque. Qual a probabilidade que não haja mais lâmpadas em estoque em menos de 9 anos?

Primeiramente:

- $\mu = E(T) = \frac{1}{\lambda} = \frac{1}{0.25} = 4$;
- $\sigma^2 = Var(T) = \frac{1}{\lambda^2} = 16$.

Portanto, temos que

$$\begin{aligned} P\left(\sum_{i=1}^{36} T_i < 9 \times 12\right) &= P(\bar{T} < 3) = P\left(\frac{\bar{T} - \mu}{\sigma/\sqrt{n}} < \frac{3-4}{4/\sqrt{36}}\right) \\ &= P(Z_n < -1.5) \approx P(Z < -1.5) \\ &= 0.05591 \end{aligned}$$

A aproximação acima se deve ao TCL. A probabilidade exata é 0.05588.

Estrutura

Introdução

Podemos dividir a tarefa de estimação de parâmetros em:

- Estimação pontual:
  - lida com a avaliação do desempenho de um estimador;
- Estimação intervalar:
  - lida com a construção de intervalos que englobam a variabilidade do estimador no processo de estimação.

Estrutura

Estimação pontual

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;
- Um estimador de θ é qualquer função de $X_1, \ldots, X_n$ utilizada para aproximar o valor de θ;

Estimação pontual

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;
- Um estimador de θ é qualquer função de $X_1, \ldots, X_n$ utilizada para aproximar o valor de θ;
- Assim sendo, é possível definir diversos estimadores para um determinado parâmetro;

Estimação pontual

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;
- Um estimador de θ é qualquer função de $X_1, \ldots, X_n$ utilizada para aproximar o valor de θ;
- Assim sendo, é possível definir diversos estimadores para um determinado parâmetro;
- Como avaliar a qualidade desses estimadores?

Estimação pontual

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;
- Um estimador de θ é qualquer função de $X_1, \ldots, X_n$ utilizada para aproximar o valor de θ;
- Assim sendo, é possível definir diversos estimadores para um determinado parâmetro;
- Como avaliar a qualidade desses estimadores?
- Resposta:

Estimação pontual

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;
- Um estimador de θ é qualquer função de $X_1, \ldots, X_n$ utilizada para aproximar o valor de θ;
- Assim sendo, é possível definir diversos estimadores para um determinado parâmetro;
- Como avaliar a qualidade desses estimadores?
- Resposta:
  - Vício;

Estimação pontual

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;
- Um estimador de θ é qualquer função de $X_1, \ldots, X_n$ utilizada para aproximar o valor de θ;
- Assim sendo, é possível definir diversos estimadores para um determinado parâmetro;
- Como avaliar a qualidade desses estimadores?
- Resposta:
  - Vício;
  - Erro-padrão;

Estimação pontual

- Sejam X v.a., $X_1, \ldots, X_n$ uma a.a. de X e θ um parâmetro;
- Um estimador de θ é qualquer função de $X_1, \ldots, X_n$ utilizada para aproximar o valor de θ;
- Assim sendo, é possível definir diversos estimadores para um determinado parâmetro;
- Como avaliar a qualidade desses estimadores?
- Resposta:
  - Vício;
  - Erro-padrão;
  - Erro Quadrático Médio (EQM).

Estrutura

Vício

- Deve ser **esperado** de um bom estimador que ele forneça o verdadeiro valor do parâmetro;

Vício

- Deve ser **esperado** de um bom estimador que ele forneça o verdadeiro valor do parâmetro;
- O vício de um estimador mede, em média, quanto o estimador se distância do parâmetro.

Vício

- Deve ser **esperado** de um bom estimador que ele forneça o verdadeiro valor do parâmetro;
- O vício de um estimador mede, em média, quanto o estimador se distância do parâmetro.

Definição

Seja θ um parâmetro e $\hat{\theta}$ um estimador de θ. O vício (ou viés) de $\hat{\theta}$ é definido por

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta = \mu_{\hat{\theta}} - \theta,$$

onde $\mu_{\hat{\theta}} = E(\hat{\theta})$ é o valor esperado de $\hat{\theta}$.

Vício

- Deve ser **esperado** de um bom estimador que ele forneça o verdadeiro valor do parâmetro;
- O vício de um estimador mede, em média, quanto o estimador se distância do parâmetro.

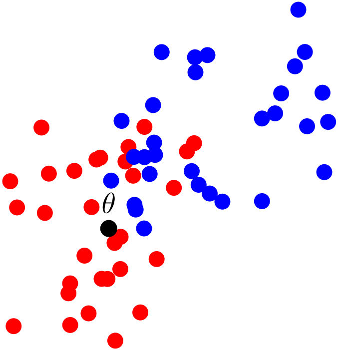
Definição

Seja θ um parâmetro e $\hat{\theta}$ um estimador de θ. O vício (ou viés) de $\hat{\theta}$ é definido por

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta = \mu_{\hat{\theta}} - \theta,$$

onde $\mu_{\hat{\theta}} = E(\hat{\theta})$ é o valor esperado de $\hat{\theta}$.

- Dizemos que $\hat{\theta}$ é não viciado (ou não viesado) se seu vício $B(\hat{\theta}) = 0$.



$\hat{\theta}_V$ não viciado

$\hat{\theta}_A$ viciado

Exemplo 1

Suponha que serão observadas $n-1 \geq 1$ observações $X_1, \ldots, X_{n-1}$ de X, com $\mu = E(X)$. Defina a média dessas $n-1$ observações por $\bar{X}_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i$. Suponha que uma nova observação X_n ficará disponível. Considere os seguintes estimadores de μ:

$$\hat{\mu}_1 = \frac{n-1}{n}\bar{X}_{n-1} + \frac{1}{n}X_n \quad \text{e} \quad \hat{\mu}_2 = \frac{\bar{X}_{n-1} + X_n}{2}.$$

Qual deles é não viciado?
Temos que

$$E(\hat{\mu}_1) = \frac{n-1}{n}E(\bar{X}_{n-1}) + \frac{1}{n}E(X_n) = \frac{n-1}{n}\mu + \frac{1}{n}\mu = \mu$$

e

$$E(\hat{\mu}_2) = \frac{E(\bar{X}_{n-1} + X_n)}{2} = \frac{E(\bar{X}_{n-1}) + E(X_n)}{2} = \frac{\mu + \mu}{2} = \mu.$$

Portanto, ambos são não viesados.

Exemplo 2

O estimador $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ é viciado para estimar σ^2! De fato,

$$
\begin{aligned}
E(\hat{\sigma}^2) &= \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \bar{X})^2] = \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \mu - (\bar{X} - \mu))^2] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\{ \sigma^2 - 2E\left[(X_i - \mu)\left(\frac{1}{n}\sum_{j}(X_j - \mu)\right)\right] + E[(\bar{X} - \mu)^2] \right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\{ \sigma^2 - \frac{2}{n}\sum_{j} E[(X_i - \mu)(X_j - \mu)] + \frac{\sigma^2}{n} \right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\{ \sigma^2 - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n} \right\} = \sigma^2 - \frac{\sigma^2}{n}
\end{aligned}
$$

Exercício

Tendo em vista o exemplo anterior, mostre que S^2 é um estimador **não** viesado de σ^2.

Estrutura

Erro-padrão

- Estimadores são v.a.;
- Além do vício, é importante saber a sua variabilidade;
- Podemos comparar dois estimadores não viesados pelo seu **Erro-padrão**.

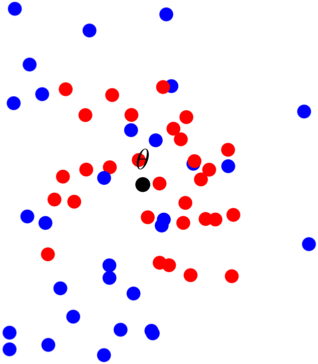
Definição

Seja $\hat{\theta}$ *estimador de* θ. *Denominamos erro-padrão de* $\hat{\theta}$ *o desvio-padrão de* $\hat{\theta}$. *Notação:* $EP(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$.

Definição

Sejam $\hat{\theta}_1$ *e* $\hat{\theta}_2$ *estimadores não viesados de* θ. *Dizemos que* $\hat{\theta}_1$ *é mais eficiente que* $\hat{\theta}_2$ *se* $EP(\hat{\theta}_1) < EP(\hat{\theta}_2)$.

$\hat{\theta}_V$ mais eficiente que $\hat{\theta}_A$



Exemplo 1 - cont.

Verifiquemos o Erro-padrão de $\hat{\mu}_1$ e de $\hat{\mu}_2$. Temos que

$$Var(\hat{\mu}_1) = Var\left(\frac{n-1}{n}\bar{X}_{n-1} + \frac{1}{n}X_n\right) = \left(\frac{n-1}{n}\right)^2 \frac{\sigma^2}{n-1} + \frac{1}{n^2}\sigma^2$$
$$= \frac{n-1+1}{n^2}\sigma^2 = \frac{1}{n}\sigma^2$$

e

$$Var(\hat{\mu}_2) = Var\left(\frac{\bar{X}_{n-1} + x_n}{2}\right) = \frac{1}{4}(Var(\bar{X}_{n-1}) + Var(X_n))$$
$$= \frac{1}{4}\left(\frac{\sigma^2}{n-1} + \sigma^2\right) = \frac{n}{4(n-1)}\sigma^2$$

Vejamos quando $\hat{\mu}_1$ é mais eficiente do que $\hat{\mu}_2$. De fato,

$$Var(\hat{\mu}_1) < Var(\hat{\mu}_2) \Rightarrow \frac{\sigma^2}{n} < \frac{n\sigma^2}{4(n-1)} \Rightarrow n^2 - 4n + 4 > 0$$
$$\Rightarrow (n-2)^2 > 0 \Rightarrow n \neq 2 \Rightarrow n > 2.$$

Estrutura

Erro Quadrático Médio

- Como comparar estimadores que tenham vícios diferentes?

Erro Quadrático Médio

- Como comparar estimadores que tenham vícios diferentes?
- Reposta: Erro quadrático médio!

Definição

Seja $\hat{\theta}$ estimador de θ. Definimos o Erro Quadrático Médio (EQM) de $\hat{\theta}$ por

$$EQM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

Exercício: Mostre que o EQM pode ser escrito como a soma do vício e da variância do estimador, isto é,

$$EQM(\hat{\theta}) = [B(\hat{\theta})]^2 + Var(\hat{\theta})$$

Exemplo 2 - cont.

No Exemplo 2, mostramos que $B(\hat{\sigma}^2) = -\frac{\sigma^2}{n}$. É possível mostrar que $B(S^2) = 0$ e que (se $X \sim N(\mu, \sigma^2)$):

$$Var(S^2) = \frac{2\sigma^4}{n-1} \text{ e } Var(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1}.$$

Logo, os EQM's de S^2 e de $\hat{\sigma}^2$ são, respectivamente,

$$EQM(S^2) = \frac{2\sigma^4}{n-1} \text{ e } EQM(\hat{\sigma}^2) = \left(\frac{\sigma^2}{n}\right)^2 + \frac{(n-1)2\sigma^4}{n^2} = \frac{2\sigma^4}{n} - \frac{\sigma^4}{n^2}$$

Logo, em termos de EQM, vemos que $\hat{\sigma}^2$ é melhor do que S^2.

Definição

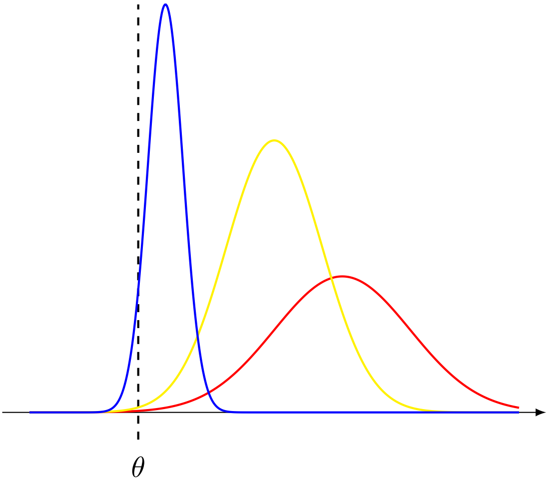
Seja $\hat{\theta}$ um estimador do parâmetro θ. Dizemos que $\hat{\theta}$ é **consistente** *se seu EQM tende a zero quando o tamanho da amostra tende a infinito. Isto é, se*

$$\lim_{n\to\infty} EQM(\hat{\theta}) = 0$$

- A definição acima implica que $\hat{\theta}$ é consistente se seu vício e sua variância tendem a 0;
- Em poucas palavras, quando um estimador é consistente suas estimativas tendem a ser próximas do verdadeiro valor do parâmetro

Distribuição amostral de um estimador consistente com

$$n < n < n$$



Estrutura

Estimação intervalar

- Mesmo que o estimador tenha boas características, é possível que a estimativa fornecida seja muito diferente do parâmetro;

Estimação intervalar

- Mesmo que o estimador tenha boas características, é possível que a estimativa fornecida seja muito diferente do parâmetro;
- Esse problema pode ser catastrófico caso o estimador tenha grande variabilidade;

Estimação intervalar

- Mesmo que o estimador tenha boas características, é possível que a estimativa fornecida seja muito diferente do parâmetro;
- Esse problema pode ser catastrófico caso o estimador tenha grande variabilidade;
- Idéia: através de intervalos incorporar a variabilidade do estimador no processo de estimação.

Estimação intervalar

- Mesmo que o estimador tenha boas características, é possível que a estimativa fornecida seja muito diferente do parâmetro;
- Esse problema pode ser catastrófico caso o estimador tenha grande variabilidade;
- Idéia: através de intervalos incorporar a variabilidade do estimador no processo de estimação.

Definição

Seja θ um parâmetro e $\hat{\theta}$ um estimador de θ. Um Intervalo de Confiança (IC) para θ com coeficiente de confiança γ é qualquer intervalo $[Li(\hat{\theta}); Ls(\hat{\theta})]$ tal que

$$P\left([Li(\hat{\theta}); Ls(\hat{\theta})] \supset \{\theta\}\right) = P\left(Li(\hat{\theta}) \leq \theta \leq Ls(\hat{\theta})\right) = \gamma.$$

Notação: $IC(\theta; \gamma) = [Li(\hat{\theta}); Ls(\hat{\theta})]$.

Comentários

- O intervalo de confiança é aleatório!

Comentários

- O intervalo de confiança é aleatório!
- A probabilidade descrita na definição anterior se refere ao intervalo conter (ou cobrir) o parâmetro;

Comentários

- O intervalo de confiança é aleatório!
- A probabilidade descrita na definição anterior se refere ao intervalo conter (ou cobrir) o parâmetro;
- A probabilidade anterior **não** se refere ao parâmetro pertencer ao intervalo;

Comentários

- O intervalo de confiança é aleatório!
- A probabilidade descrita na definição anterior se refere ao intervalo conter (ou cobrir) o parâmetro;
- A probabilidade anterior **não** se refere ao parâmetro pertencer ao intervalo;
- O parâmetro é **constante**, portanto não faz sentido associar probabilidades a ele;

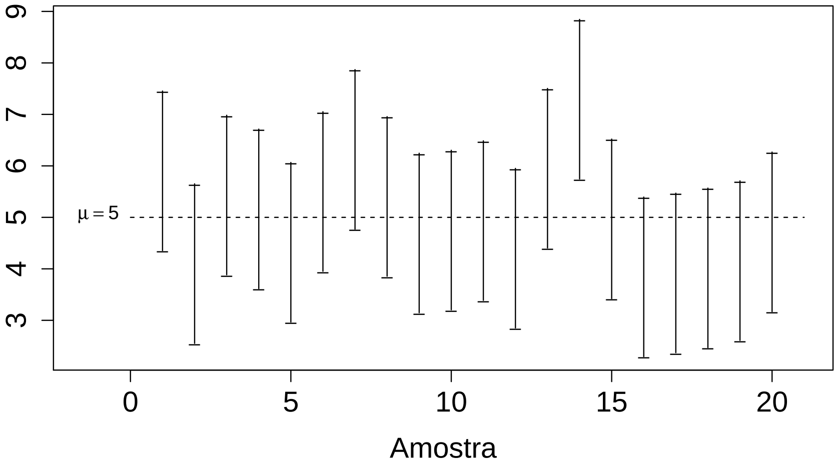
Comentários

- O intervalo de confiança é aleatório!
- A probabilidade descrita na definição anterior se refere ao intervalo conter (ou cobrir) o parâmetro;
- A probabilidade anterior **não** se refere ao parâmetro pertencer ao intervalo;
- O parâmetro é **constante**, portanto não faz sentido associar probabilidades a ele;
- Após observada a amostra o intervalo não é mais aleatório, assim a probabilidade de ele conter o parâmetro é 0 ou 1;

- Se coletássemos várias amostras e, para cada uma, calculássemos o IC, esperaríamos que, $100\gamma\%$ dos IC's contivessem o parâmetro;

- Se coletássemos várias amostras e, para cada uma, calculássemos o IC, esperaríamos que, $100\gamma\%$ dos IC's contivessem o parâmetro;
- Por esse motivo, dizemos que temos uma **confiança** γ, isto é, uma interpretação frequentista;

- Se coletássemos várias amostras e, para cada uma, calculássemos o IC, esperaríamos que, $100\gamma\%$ dos IC's contivessem o parâmetro;
- Por esse motivo, dizemos que temos uma **confiança** γ, isto é, uma interpretação frequentista;
- Em outras palavras, após coletar uma amostra não saberemos se o intervalo calculado cobrirá o parâmetro. Entretanto, a construção do intervalo garante que isso ocorre em $100\gamma\%$ das vezes.



3
4
5
6
7
8
9
μ = 5
0
5
10
15
20
Amostra

Estrutura

IC para μ com σ^2 conhecida

Encontremos um IC para μ em função de $\bar{X}$. Seja $z_{\frac{\gamma}{2}}$ o número tal que $P(-z_{\frac{\gamma}{2}} < Z < z_{\frac{\gamma}{2}}) = \gamma$, onde $Z \sim N(0,1)$. De fato, pelo TCL

$$
\begin{aligned}
\gamma &= P\left(-z_{\frac{\gamma}{2}} < Z < z_{\frac{\gamma}{2}}\right) \approx P\left(-z_{\frac{\gamma}{2}} < Z_n < z_{\frac{\gamma}{2}}\right) \\
&= P\left(-z_{\frac{\gamma}{2}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\frac{\gamma}{2}}\right) = P\left(-z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(\bar{X} - z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}}\right)
\end{aligned}
$$

Logo um IC para μ com coeficiente de confiança γ é dado por

$$
IC(\mu;\gamma) = \left[\bar{X} - z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}}\right].
$$

Exemplo

Uma máquina estava regulada para encher pacotes de café com peso médio de $500g$ e variância de $196g^2$. Uma a.a. de 49 pacotes forneceu média de $485g$. O que podemos dizer sobre a verdadeira média com confiança de 95%?

Exemplo

Uma máquina estava regulada para encher pacotes de café com peso médio de $500g$ e variância de $196g^2$. Uma a.a. de 49 pacotes forneceu média de $485g$. O que podemos dizer sobre a verdadeira média com confiança de 95%?

Temos que $\bar{x} = 485$, $n = 49$, $\sigma^2 = 196$ e $\gamma = 0.95$. Primeiramente, $z_{\frac{\gamma}{2}} = z_{0.475} = 1.96$. Logo,

$$IC(\mu; 0.95) = \left[485 - 1.96 \times \frac{14}{7}; 485 + 1.96 \times \frac{14}{7}\right] = [481.08; 488.92].$$

Exemplo

Uma máquina estava regulada para encher pacotes de café com peso médio de $500g$ e variância de $196g^2$. Uma a.a. de 49 pacotes forneceu média de $485g$. O que podemos dizer sobre a verdadeira média com confiança de 95%?

Temos que $\bar{x} = 485$, $n = 49$, $\sigma^2 = 196$ e $\gamma = 0.95$. Primeiramente, $z_{\frac{\gamma}{2}} = z_{0.475} = 1.96$. Logo,

$$IC(\mu; 0.95) = \left[485 - 1.96 \times \frac{14}{7}; 485 + 1.96 \times \frac{14}{7}\right] = [481.08; 488.92].$$

Com confiança de 95% podemos afirmar que a máquina está desregulada.

Comentários

- No exemplo anterior, supomos que a desregulagem afeta apenas a média com que os pacotes são preenchidos;

Comentários

- No exemplo anterior, supomos que a desregulagem afeta apenas a média com que os pacotes são preenchidos;
- E se afetasse a variância?

Comentários

- No exemplo anterior, supomos que a desregulagem afeta apenas a média com que os pacotes são preenchidos;
- E se afetasse a variância?
- Em outras palavras, e se σ^2 for **desconhecida**?

IC para μ com σ^2 desconhecida (grandes amostras)

Temos primeiramente que S^2 é um estimador consistente de σ^2. Portanto, em grandes amostras (n muito grande), $S^2 \approx \sigma^2$.

IC para μ com σ^2 desconhecida (grandes amostras)

Temos primeiramente que S^2 é um estimador consistente de σ^2. Portanto, em grandes amostras (n muito grande), $S^2 \approx \sigma^2$. Daí podemos utilizar o seguinte IC para μ quando σ^2 é desconhecida:

$$IC(\mu;\gamma) = \left[\bar{X} - z_{\frac{\gamma}{2}}\frac{S}{\sqrt{n}}; \bar{X} + z_{\frac{\gamma}{2}}\frac{S}{\sqrt{n}}\right].$$

IC para μ com σ^2 desconhecida (grandes amostras)

Temos primeiramente que S^2 é um estimador consistente de σ^2. Portanto, em grandes amostras (n muito grande), $S^2 \approx \sigma^2$. Daí podemos utilizar o seguinte IC para μ quando σ^2 é desconhecida:

$$IC(\mu;\gamma) = \left[\bar{X} - z_{\frac{\gamma}{2}}\frac{S}{\sqrt{n}}; \bar{X} + z_{\frac{\gamma}{2}}\frac{S}{\sqrt{n}}\right].$$

Se o tamanho amostral é suficientemente grande, o IC acima fornece aproximadamente confiança γ.

IC para μ com σ^2 desconhecida (amostras normais)

- Caso o valor de n seja pequeno, o valor fornecido por S^2 pode estar distante de σ^2;

IC para μ com σ^2 desconhecida (amostras normais)

- Caso o valor de n seja pequeno, o valor fornecido por S^2 pode estar distante de σ^2;
- Portanto, a distribuição de $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ pode não ser muito próxima da de v.a. $Z \sim N(0,1)$;

IC para μ com σ^2 desconhecida (amostras normais)

- Caso o valor de n seja pequeno, o valor fornecido por S^2 pode estar distante de σ^2;
- Portanto, a distribuição de $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ pode não ser muito próxima da de v.a. $Z \sim N(0,1)$;
- Entretanto, supondo que **os dados** são observados de uma v.a. normal, pode-se mostrar que a v.a.

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

IC para μ com σ^2 desconhecida (amostras normais)

- Caso o valor de n seja pequeno, o valor fornecido por S^2 pode estar distante de σ^2;
- Portanto, a distribuição de $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ pode não ser muito próxima da de v.a. $Z \sim N(0,1)$;
- Entretanto, supondo que **os dados** são observados de uma v.a. normal, pode-se mostrar que a v.a.

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

- O símbolo $T \sim t_r$ significa que a v.a. T tem uma distribuição t-student com r graus de liberdade (g.l.).

IC para μ com σ^2 desconhecida (amostras normais)

- Caso o valor de n seja pequeno, o valor fornecido por S^2 pode estar distante de σ^2;
- Portanto, a distribuição de $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ pode não ser muito próxima da de v.a. $Z \sim N(0,1)$;
- Entretanto, supondo que **os dados** são observados de uma v.a. normal, pode-se mostrar que a v.a.

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}.$$

- O símbolo $T \sim t_r$ significa que a v.a. T tem uma distribuição t-student com r graus de liberdade (g.l.).

Comentário

Em resumo se a v.a. X *que será observada satisfizer* $X \sim N(\mu, \sigma^2)$. *Então*

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tem distribuição t-student com $n-1$ *g.l.*

Utilizando o resultado anterior, encontremos um IC para μ em função de $\bar{X}$. Seja $t_{(\frac{\gamma}{2};n-1)}$ o número tal que $P\left(-t_{(\frac{\gamma}{2};n-1)} < T < t_{(\frac{\gamma}{2};n-1)}\right) = \gamma$, onde $T \sim t_{n-1}$. De fato,

$$\gamma = P\left(-t_{(\frac{\gamma}{2};n-1)} < T < t_{(\frac{\gamma}{2};n-1)}\right) = P\left(-t_{(\frac{\gamma}{2};n-1)} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{(\frac{\gamma}{2};n-1)}\right)$$
$$= P\left(\bar{X} - t_{(\frac{\gamma}{2};n-1)}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{(\frac{\gamma}{2};n-1)}\frac{S}{\sqrt{n}}\right)$$

Logo quando σ^2 é desconhecida e o tamanho amostral não é grande, se os dados vierem de uma v.a. normal, um IC para μ com coeficiente de confiança γ é dado por

$$IC(\mu;\gamma) = \left[\bar{X} - t_{(\frac{\gamma}{2};n-1)}\frac{S}{\sqrt{n}}; \bar{X} + t_{(\frac{\gamma}{2};n-1)}\frac{S}{\sqrt{n}}\right].$$

Exemplo 1 - cont.

No exemplo dos sacos de café, suponha que em vez de 49, apenas 9 sacos tivessem sido coletados.

Exemplo 1 - cont.

No exemplo dos sacos de café, suponha que em vez de 49, apenas 9 sacos tivessem sido coletados. Entretanto, suponha que os pesos dos sacos de café seguem uma distribuição normal.

Exemplo 1 - cont.

No exemplo dos sacos de café, suponha que em vez de 49, apenas 9 sacos tivessem sido coletados. Entretanto, suponha que os pesos dos sacos de café seguem uma distribuição normal. Há suspeita de que a variância também pode ter sido alterada.

Exemplo 1 - cont.

No exemplo dos sacos de café, suponha que em vez de 49, apenas 9 sacos tivessem sido coletados. Entretanto, suponha que os pesos dos sacos de café seguem uma distribuição normal. Há suspeita de que a variância também pode ter sido alterada. Por este motivo, a variância amostral foi calculada e forneceu $s^2 = 225$. O que podemos dizer sobre a situação da máquina.

Exemplo 1 - cont.

No exemplo dos sacos de café, suponha que em vez de 49, apenas 9 sacos tivessem sido coletados. Entretanto, suponha que os pesos dos sacos de café seguem uma distribuição normal. Há suspeita de que a variância também pode ter sido alterada. Por este motivo, a variância amostral foi calculada e forneceu $s^2 = 225$. O que podemos dizer sobre a situação da máquina. Primeiramente, temos que $t_{(\frac{\gamma}{2};n-1)} = t_{(0.475;8)} = 2,306$.

Exemplo 1 - cont.

No exemplo dos sacos de café, suponha que em vez de 49, apenas 9 sacos tivessem sido coletados. Entretanto, suponha que os pesos dos sacos de café seguem uma distribuição normal. Há suspeita de que a variância também pode ter sido alterada. Por este motivo, a variância amostral foi calculada e forneceu $s^2 = 225$. O que podemos dizer sobre a situação da máquina. Primeiramente, temos que $t_{(\frac{\gamma}{2};n-1)} = t_{(0.475;8)} = 2,306$. Logo,

$$IC(\mu;\gamma) = \left[485 - 2.306 \times \frac{15}{3}; 485 + 2.306 \times \frac{15}{3}\right] = [473.47; 496.53].$$

Exemplo 1 - cont.

No exemplo dos sacos de café, suponha que em vez de 49, apenas 9 sacos tivessem sido coletados. Entretanto, suponha que os pesos dos sacos de café seguem uma distribuição normal. Há suspeita de que a variância também pode ter sido alterada. Por este motivo, a variância amostral foi calculada e forneceu $s^2 = 225$. O que podemos dizer sobre a situação da máquina. Primeiramente, temos que $t_{(\frac{\gamma}{2};n-1)} = t_{(0.475;8)} = 2,306$. Logo,

$$IC(\mu;\gamma) = \left[485 - 2.306 \times \frac{15}{3}; 485 + 2.306 \times \frac{15}{3}\right] = [473.47; 496.53].$$

Portanto, percebemos que a amplitude do intervalo aumentou significativamente. Entretanto, as conclusões obtidas não foram alteradas.

Propriedades de IC

- $\uparrow \gamma \Rightarrow \uparrow$ amplitude;

Propriedades de IC

- $\uparrow \gamma \Rightarrow \uparrow$ amplitude;
- $\uparrow \sigma^2 \Rightarrow \uparrow$ amplitude;

Propriedades de IC

- $\uparrow \gamma \Rightarrow \uparrow$ amplitude;
- $\uparrow \sigma^2 \Rightarrow \uparrow$ amplitude;
- $\uparrow n \Rightarrow \downarrow$ amplitude.

Exercício: Criar um IC para a probabilidade de sucesso p utilizando como estimador a proporção amostral $\hat{p}$. Note que a variância dependerá de p (que é desconhecido)! Qual solução você sugere?

Estrutura

Introdução

- Objetivo: Testar afirmações sobre parâmetros populacionais.

Introdução

- Objetivo: Testar afirmações sobre parâmetros populacionais.

Definição

Seja θ um parâmetro da distribuição de um v.a. X. Uma hipótese é qualquer afirmação feita com relação a θ que pode ser submetida a teste.

- Objetivo: Testar afirmações sobre parâmetros populacionais.

Definição

Seja θ um parâmetro da distribuição de um v.a. X. Uma hipótese é qualquer afirmação feita com relação a θ que pode ser submetida a teste.

Exemplo: Seja θ um parâmetro populacional. Exemplos mais usuais de hipóteses com respeito a θ são:

- Simples: $\theta = \theta_0$;
- Compostas:
  - $\theta \geq \theta_0$ e $\theta > \theta_0$ (unilateral);
  - $\theta \leq \theta_0$ e $\theta < \theta_0$ (unilateral); e
  - $\theta \neq \theta_0$ (bilateral).

- Para realizar um teste de hipóteses estatístico devemos estabelecer corretamente as hipóteses:
  - nula - denotada por H_0, é aquela que será submetida a teste. Pode ser, ou não rejeitada; e
  - alternativa - denotada por H_1, é aquela que é aceita quando rejeitamos H_0.
- Geralmente, mas nem sempre, H_1 representa o complementar de H_0;
- Entretanto, H_0 e H_1 devem ser mutuamente excludentes.

Definição

Um teste de hipósteses é o procedimento de tomada de decisão sobre rejeitar, ou não, H_0.

Definição

Um teste de hipósteses é o procedimento de tomada de decisão sobre rejeitar, ou não, H_0.

- Como testar entre H_0 e H_1?

Definição

Um teste de hipósteses é o procedimento de tomada de decisão sobre rejeitar, ou não, H_0.

- Como testar entre H_0 e H_1?
- Utilizar um estimador $\hat{\theta}$ do parâmetro θ para a tomada de decisão;

Definição

Um teste de hipósteses é o procedimento de tomada de decisão sobre rejeitar, ou não, H_0.

- Como testar entre H_0 e H_1?
- Utilizar um estimador $\hat{\theta}$ do parâmetro θ para a tomada de decisão;
- O primeiro passo é estabelecer a Região Crítica (RC) do teste.

Definição

Um teste de hipósteses é o procedimento de tomada de decisão sobre rejeitar, ou não, H_0.

- Como testar entre H_0 e H_1?
- Utilizar um estimador $\hat{\theta}$ do parâmetro θ para a tomada de decisão;
- O primeiro passo é estabelecer a Região Crítica (RC) do teste.

Definição

A região crítica de um teste de hipóteses é um conjunto de valores que, se fornecidos pelo estimador, levam a rejeição de H_0. Formalmente,

$$RC_{H_0} = \{x : \text{ se } \hat{\theta} = x,\ H_0 \text{ é rejeitada}\}.$$

Definição

Um teste de hipósteses é o procedimento de tomada de decisão sobre rejeitar, ou não, H_0.

- Como testar entre H_0 e H_1?
- Utilizar um estimador $\hat{\theta}$ do parâmetro θ para a tomada de decisão;
- O primeiro passo é estabelecer a Região Crítica (RC) do teste.

Definição

A região crítica de um teste de hipóteses é um conjunto de valores que, se fornecidos pelo estimador, levam a rejeição de H_0. Formalmente,

$$RC_{H_0} = \{x : \text{ se } \hat{\theta} = x, \ H_0 \text{ é rejeitada}\}.$$

Exemplo: Suponha que desejamos testar $H_0 : \mu \geq \mu_0$. Uma RC natural para testar H_0 seria $RC_{H_0} = \{\bar{X} < x_c\}$. Isto é, valores muito pequenos de $\bar{X}$ indicariam que H_0 deve ser rejeitada. O problema agora consistiria em determinar o **valor crítico** x_c, ou seja, determinar o que seria muito pequeno.

Tipos de erros

- Decidir entre rejeitar, ou não, H_0 pode acarretar em dois tipos de erros:

Tipos de erros

- Decidir entre rejeitar, ou não, H_0 pode acarretar em dois tipos de erros:
  - erro tipo 1 - ao rejeitarmos H_0, quando H_0 é verdadeira;
  - erro tipo 2 - ao não rejeitamos H_0, quando H_0 é falsa.

Tipos de erros

- Decidir entre rejeitar, ou não, H_0 pode acarretar em dois tipos de erros:
  - erro tipo 1 - ao rejeitarmos H_0, quando H_0 é verdadeira;
  - erro tipo 2 - ao não rejeitamos H_0, quando H_0 é falsa.
- Pela decisão ser tomada com base em um estimador (v.a.) cada tipo de erro tem uma probabilidade de ocorrer:
  - $P(\text{erro tipo 1}) = P(\text{Rejeitar } H_0|H_0 \text{ é verdadeira}) := \alpha(\theta)$. A função $\alpha(\theta)$ é denominada tamanho do teste;
  - $P(\text{erro tipo 2}) = P(\text{Não Rejeitar } H_0|H_0 \text{ é falsa}) := \beta(\theta)$. A função $1 - \beta(\theta)$ é denominada poder do teste.

Exemplo

Uma indústria utiliza um parafuso que deve satisfazer determinadas especificações. Uma dessas é a resistência à tração X (em kg/m^2). Dependendo da localidade em que os parafusos são fabricados, X pode ter diferentes distribuições. Suponha que apenas as localidades A e B produzam os parafusos. Se o parafuso vem:

- de A, então $E(X) = 145$ e $\sigma^2 = 12^2$;
- de B, então $E(X) = 155$ e $\sigma^2 = 18^2$.

Exemplo

Uma indústria utiliza um parafuso que deve satisfazer determinadas especificações. Uma dessas é a resistência à tração X (em kg/m^2). Dependendo da localidade em que os parafusos são fabricados, X pode ter diferentes distribuições. Suponha que apenas as localidades A e B produzam os parafusos. Se o parafuso vem:

- de A, então $E(X) = 145$ e $\sigma^2 = 12^2$;
- de B, então $E(X) = 155$ e $\sigma^2 = 18^2$.

Um lote grande de parafusos será leiloado e a empresa deseja comprar esse lote somente se o mesmo vier da localidade B. No dia do leilão o leiloeiro divulgará a resistência à tração média de 36 parafusos escolhidos desse lote.

Exemplo

Uma indústria utiliza um parafuso que deve satisfazer determinadas especificações. Uma dessas é a resistência à tração X (em kg/m^2). Dependendo da localidade em que os parafusos são fabricados, X pode ter diferentes distribuições. Suponha que apenas as localidades A e B produzam os parafusos. Se o parafuso vem:

- de A, então $E(X) = 145$ e $\sigma^2 = 12^2$;
- de B, então $E(X) = 155$ e $\sigma^2 = 18^2$.

Um lote grande de parafusos será leiloado e a empresa deseja comprar esse lote somente se o mesmo vier da localidade B. No dia do leilão o leiloeiro divulgará a resistência à tração média de 36 parafusos escolhidos desse lote. Podemos estabelecer as hipóteses nula e alternativa como sendo:

$$H_0 : \mu = 155, \ \sigma = 18 \quad \times \quad H_1 : \mu = 145, \ \sigma = 12.$$

Se decidirmos rejeitar H_0 quando (por exemplo) $\bar{X} < 150$, cometeriamos os erros tipos 1 e 2 com quais probabilidades?

Exemplo - cont.

Como neste caso H_0 e H_1 são ambas hipóteses simples, $\alpha = \alpha(\mu = 155, \sigma = 18)$ e $\beta = \beta(\mu = 145, \sigma = 12)$ são de fácil obtenção.

Exemplo - cont.

Como neste caso H_0 e H_1 são ambas hipóteses simples, $\alpha = \alpha(\mu = 155, \sigma = 18)$ e $\beta = \beta(\mu = 145, \sigma = 12)$ são de fácil obtenção. De fato,

$$
\begin{aligned}
\alpha &= P(\text{erro tipo 1}) = P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}) \\
&= P(\bar{X} < 150 | \mu = 155, \sigma = 18) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{150 - 155}{18/\sqrt{36}}\right) \\
&\approx P(Z < -1.67) = 0.0475
\end{aligned}
$$

Exemplo - cont.

Como neste caso H_0 e H_1 são ambas hipóteses simples, $\alpha = \alpha(\mu = 155, \sigma = 18)$ e $\beta = \beta(\mu = 145, \sigma = 12)$ são de fácil obtenção. De fato,

$$\begin{aligned}
\alpha &= P(\text{erro tipo 1}) = P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}) \\
&= P(\bar{X} < 150 | \mu = 155, \sigma = 18) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{150 - 155}{18/\sqrt{36}}\right) \\
&\approx P(Z < -1.67) = 0.0475
\end{aligned}$$

e, analogamente,

$$\begin{aligned}
\beta &= P(\text{erro tipo 2}) = P(\text{Não Rejeitar } H_0 | H_0 \text{ é falsa}) \\
&= P(\bar{X} \geq 150 | \mu = 145, \sigma = 12) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{150 - 145}{12/\sqrt{36}}\right) \\
&\approx P(Z \geq 2.5) = 0.0062.
\end{aligned}$$

Exemplo - cont.

Embora tenhamos fixado a RC e depois verificado α e β, o usual é encontrar a RC que fornece um valor pré-estabelecido para α.

Exemplo - cont.

Embora tenhamos fixado a RC e depois verificado α e β, o usual é encontrar a RC que fornece um valor pré-estabelecido para α. Primeiramente, devemos encontrar o "formato" da RC.

Exemplo - cont.

Embora tenhamos fixado a RC e depois verificado α e β, o usual é encontrar a RC que fornece um valor pré-estabelecido para α. Primeiramente, devemos encontrar o "formato" da RC. Perceba que teremos evidências para rejeitar H_0 se o valor de $\bar{X}$ for "suficientemente" menor do que 155, isto é, se $\bar{X} < x_c$, para algum $x_c < 155$.

Exemplo - cont.

Embora tenhamos fixado a RC e depois verificado α e β, o usual é encontrar a RC que fornece um valor pré-estabelecido para α. Primeiramente, devemos encontrar o "formato" da RC. Perceba que teremos evidências para rejeitar H_0 se o valor de $\bar{X}$ for "suficientemente" menor do que 155, isto é, se $\bar{X} < x_c$, para algum $x_c < 155$. Podemos encontrar o valor de x_c que forneça probabilidade de erro tipo 1 igual a 0.1 por exemplo.

Exemplo - cont.

Embora tenhamos fixado a RC e depois verificado α e β, o usual é encontrar a RC que fornece um valor pré-estabelecido para α. Primeiramente, devemos encontrar o "formato" da RC. Perceba que teremos evidências para rejeitar H_0 se o valor de $\bar{X}$ for "suficientemente" menor do que 155, isto é, se $\bar{X} < x_c$, para algum $x_c < 155$. Podemos encontrar o valor de x_c que forneça probabilidade de erro tipo 1 igual a 0.1 por exemplo. Temos que

$$\begin{aligned}
0.1 = \alpha = P(\text{erro tipo 1}) &= P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}) \\
&= P(\bar{X} < x_c | \mu = 155, \sigma = 18) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{x_c - 155}{18/\sqrt{36}}\right) \\
&\approx P\left(Z < \frac{x_c - 155}{3}\right) = P(Z < z),
\end{aligned}$$

onde $z = \frac{x_c - 155}{3}$. Logo, devemos ter $\frac{x_c - 155}{3} = z = -1.28 \Rightarrow x_c = 151.16$.

Exemplo - cont.

Embora tenhamos fixado a RC e depois verificado α e β, o usual é encontrar a RC que fornece um valor pré-estabelecido para α. Primeiramente, devemos encontrar o "formato" da RC. Perceba que teremos evidências para rejeitar H_0 se o valor de $\bar{X}$ for "suficientemente" menor do que 155, isto é, se $\bar{X} < x_c$, para algum $x_c < 155$. Podemos encontrar o valor de x_c que forneça probabilidade de erro tipo 1 igual a 0.1 por exemplo. Temos que

$$\begin{aligned}
0.1 = \alpha = P(\text{erro tipo 1}) &= P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}) \\
&= P(\bar{X} < x_c | \mu = 155, \sigma = 18) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{x_c - 155}{18/\sqrt{36}}\right) \\
&\approx P\left(Z < \frac{x_c - 155}{3}\right) = P(Z < z),
\end{aligned}$$

onde $z = \frac{x_c - 155}{3}$. Logo, devemos ter $\frac{x_c - 155}{3} = z = -1.28 \Rightarrow x_c = 151.16$.
Exercício: Encontre β para a RC acima.

Tipos de erros

Definição

O valor que fixamos para a probabilidade do erro tipo 1 é denominado nível de significância.

Tipos de erros

Definição

O valor que fixamos para a probabilidade do erro tipo 1 é denominado nível de significância.

- No exemplo anterior foi possível fixar o nível de significância e determinar β de maneira exata, pois as duas hipóteses eram simples!

Tipos de erros

Definição

O valor que fixamos para a probabilidade do erro tipo 1 é denominado nível de significância.

- No exemplo anterior foi possível fixar o nível de significância e determinar β de maneira exata, pois as duas hipóteses eram simples!
- Como proceder no caso de hipóteses compostas?

Tipos de erros

Definição

O valor que fixamos para a probabilidade do erro tipo 1 é denominado nível de significância.

- No exemplo anterior foi possível fixar o nível de significância e determinar β de maneira exata, pois as duas hipóteses eram simples!
- Como proceder no caso de hipóteses compostas?
  - Se H_0 for composta (por exemplo $\theta \leq \theta_0$), a probabilidade do erro tipo 1 é uma função do verdadeiro valor de θ que é desconhecido sob H_0;

Tipos de erros

Definição

O valor que fixamos para a probabilidade do erro tipo 1 é denominado nível de significância.

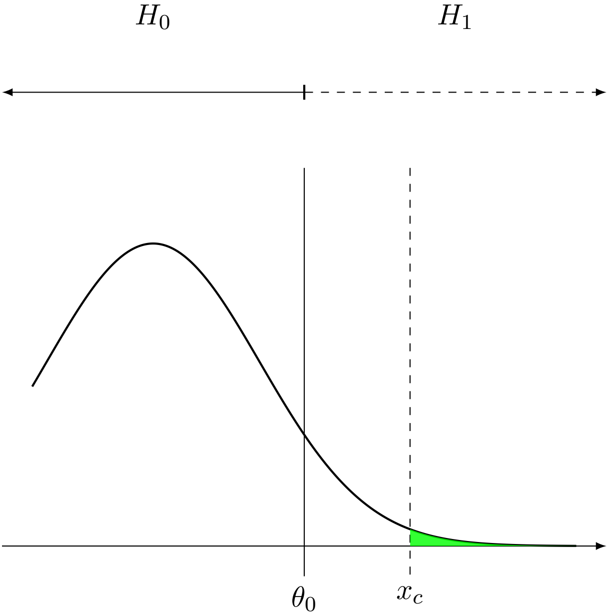
- No exemplo anterior foi possível fixar o nível de significância e determinar β de maneira exata, pois as duas hipóteses eram simples!
- Como proceder no caso de hipóteses compostas?
  - Se H_0 for composta (por exemplo $\theta \leq \theta_0$), a probabilidade do erro tipo 1 é uma função do verdadeiro valor de θ que é desconhecido sob H_0;
  - Neste caso podemos ser conservadores e tomar o nível de significância como sendo a maior das probabilidades de erro tipo 1, isto é podemos fazer o nível de significância como $\sup_\theta \alpha(\theta)$;

Tipos de erros

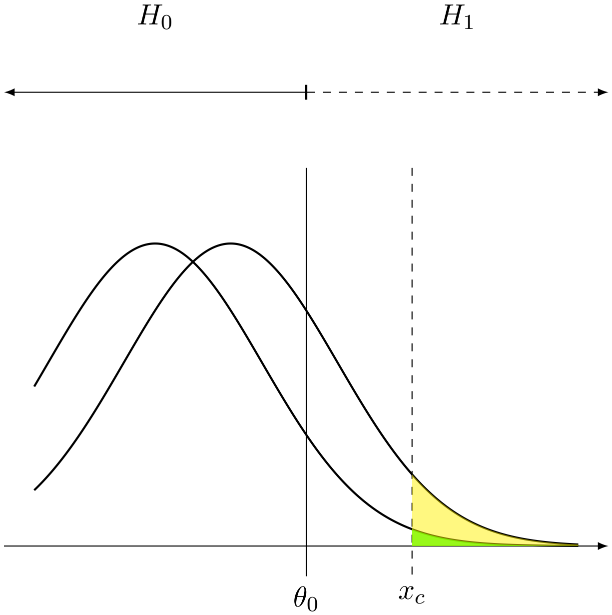
Definição

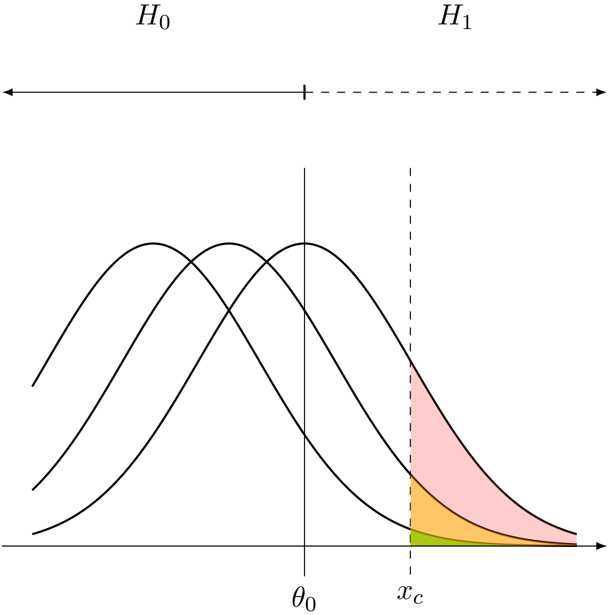
O valor que fixamos para a probabilidade do erro tipo 1 é denominado nível de significância.

- No exemplo anterior foi possível fixar o nível de significância e determinar β de maneira exata, pois as duas hipóteses eram simples!
- Como proceder no caso de hipóteses compostas?
  - Se H_0 for composta (por exemplo $\theta \leq \theta_0$), a probabilidade do erro tipo 1 é uma função do verdadeiro valor de θ que é desconhecido sob H_0;
  - Neste caso podemos ser conservadores e tomar o nível de significância como sendo a maior das probabilidades de erro tipo 1, isto é podemos fazer o nível de significância como $\sup_\theta \alpha(\theta)$;
  - Se o tamanho do teste $\alpha(\theta)$ tiver comportamento monótono, então $\sup_\theta \alpha(\theta) = \alpha(\theta_0)$ para ambos $H_0 : \theta = \theta_0$, $H_0 : \theta \leq \theta_0$ e $H_0 : \theta \geq \theta_0$.



H_0
H_1
θ_0
x_c


H_0
H_1
θ_0
x_c


H_0
H_1
θ_0
x_c

- Devido a construção do teste de hipóteses, H_0 sempre deve conter o sinal de igualdade;

- Devido a construção do teste de hipóteses, H_0 sempre deve conter o sinal de igualdade;
- O nível de significância pode ser enxergado como um *tradeoff* entre α e β;

- Devido a construção do teste de hipóteses, H_0 sempre deve conter o sinal de igualdade;
- O nível de significância pode ser enxergado como um *tradeoff* entre α e β;
- Pois:
  - $\uparrow \alpha \Rightarrow \downarrow \beta$;
  - $\downarrow \alpha \Rightarrow \uparrow \beta$;

- Devido a construção do teste de hipóteses, H_0 sempre deve conter o sinal de igualdade;
- O nível de significância pode ser enxergado como um *tradeoff* entre α e β;
- Pois:
  - $\uparrow \alpha \Rightarrow \downarrow \beta$;
  - $\downarrow \alpha \Rightarrow \uparrow \beta$;
- Se o custo do erro tipo 1 for mais alto que o do erro tipo 2, devemos adotar um nível de significância baixo;

- Devido a construção do teste de hipóteses, H_0 sempre deve conter o sinal de igualdade;
- O nível de significância pode ser enxergado como um *tradeoff* entre α e β;
- Pois:
  - $\uparrow \alpha \Rightarrow \downarrow \beta$;
  - $\downarrow \alpha \Rightarrow \uparrow \beta$;
- Se o custo do erro tipo 1 for mais alto que o do erro tipo 2, devemos adotar um nível de significância baixo;
- Se o custo do erro tipo 1 for mais baixo que o do erro tipo 2, devemos adotar um nível de significância alto (para indiretamente diminuir o valor de β);

- Devido a construção do teste de hipóteses, H_0 sempre deve conter o sinal de igualdade;
- O nível de significância pode ser enxergado como um *tradeoff* entre α e β;
- Pois:
  - $\uparrow \alpha \Rightarrow \downarrow \beta$;
  - $\downarrow \alpha \Rightarrow \uparrow \beta$;
- Se o custo do erro tipo 1 for mais alto que o do erro tipo 2, devemos adotar um nível de significância baixo;
- Se o custo do erro tipo 1 for mais baixo que o do erro tipo 2, devemos adotar um nível de significância alto (para indiretamente diminuir o valor de β);
- Depois de fixado o nível de significância e obtida a RC, se H_1 é composta, a probabilidade do erro tipo 2 não pode ser obtida de maneira exata e deve ser tratada como função de θ;

Etapas de um teste de hipóteses

1. Estabelecer H_0 e H_1;

Etapas de um teste de hipóteses

1. Estabelecer H_0 e H_1;
2. Observar o comportamento de H_1 e determinar o "formato" apropriado da RC;

Etapas de um teste de hipóteses

1. Estabelecer H_0 e H_1;
2. Observar o comportamento de H_1 e determinar o "formato" apropriado da RC;
3. Fixar o nível de significância;

Etapas de um teste de hipóteses

1. Estabelecer H_0 e H_1;
2. Observar o comportamento de H_1 e determinar o "formato" apropriado da RC;
3. Fixar o nível de significância;
4. Determinar a RC correspondente ao nível de significância estabelecido;

Etapas de um teste de hipóteses

1. Estabelecer H_0 e H_1;
2. Observar o comportamento de H_1 e determinar o "formato" apropriado da RC;
3. Fixar o nível de significância;
4. Determinar a RC correspondente ao nível de significância estabelecido;
5. Observar o valor obtido para $\hat{\theta}$ e verficar se pertence a RC (caso em que rejeitamos H_0 ao nível de significância fixado). Se $\hat{\theta}$ não pertencer a RC concluimos que a amostra não forneceu evidências suficientes para a rejeição de H_0.

Etapas de um teste de hipóteses

1. Estabelecer H_0 e H_1;
2. Observar o comportamento de H_1 e determinar o "formato" apropriado da RC;
3. Fixar o nível de significância;
4. Determinar a RC correspondente ao nível de significância estabelecido;
5. Observar o valor obtido para $\hat{\theta}$ e verficar se pertence a RC (caso em que rejeitamos H_0 ao nível de significância fixado). Se $\hat{\theta}$ não pertencer a RC concluimos que a amostra não forneceu evidências suficientes para a rejeição de H_0.

Por que não podemos dizer que aceitamos H_0?

Etapas de um teste de hipóteses

1. Estabelecer H_0 e H_1;
2. Observar o comportamento de H_1 e determinar o "formato" apropriado da RC;
3. Fixar o nível de significância;
4. Determinar a RC correspondente ao nível de significância estabelecido;
5. Observar o valor obtido para $\hat{\theta}$ e verficar se pertence a RC (caso em que rejeitamos H_0 ao nível de significância fixado). Se $\hat{\theta}$ não pertencer a RC concluimos que a amostra não forneceu evidências suficientes para a rejeição de H_0.

Por que não podemos dizer que aceitamos H_0? Pois não sabemos a probabilidade de erro ao fazermos tal afirmação!

Estrutura

Estrutura

- Um dos parâmetros de maior interesse é a média μ de uma população;
- Podemos considerar as seguintes hipóteses a serem testadas:
    1. $H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0$;
    2. $H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0$;
    3. $H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0$;
- Nas hipóteses acima, μ é um parâmetro, enquanto μ_0 é um número.

Caso 1: $H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0;$$

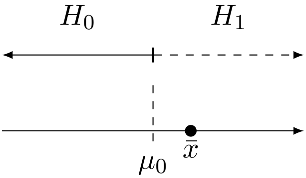
- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

Caso 1: $H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

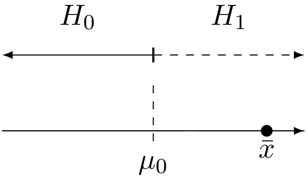


Caso 1: $H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

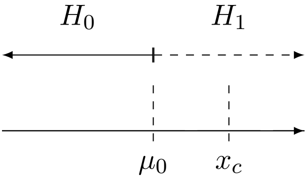


Caso 1: $H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

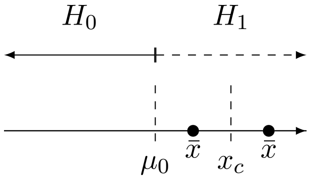


Caso 1: $H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

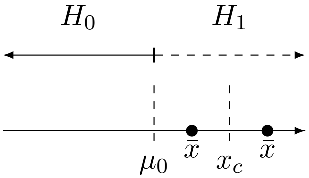


Caso 1: $H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \leq \mu_0 \times H_1 : \mu > \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.



- Portanto, $RC_{H_0} = \{\bar{X} > x_c\}$;
- Como determinar x_c?

- Fixemos o nível de significância em α;
- Seja z_α o valor tal que $P(Z > z_\alpha) = \alpha$;
- Dessa forma, temos que

$$
\begin{aligned}
\alpha = \sup \alpha(\mu) = \alpha(\mu_0) &= P(\text{Rejeitar } H_0 | \mu = \mu_0) \\
&= P(\bar{X} > x_c | \mu = \mu_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{x_c - \mu_0}{\sigma/\sqrt{n}} \right) \\
&\approx P\left(Z > \frac{x_c - \mu_0}{\sigma/\sqrt{n}} \right).
\end{aligned}
$$

- Logo, $\frac{x_c - \mu_0}{\sigma/\sqrt{n}} = z_\alpha \Rightarrow x_c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$;
- Assim, devemos rejeitar H_0 ao nível de significância α se

$$\bar{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Caso 2: $H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0;$$

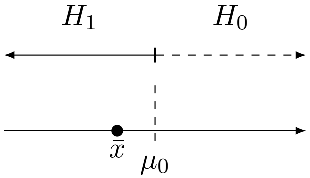
- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

Caso 2: $H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

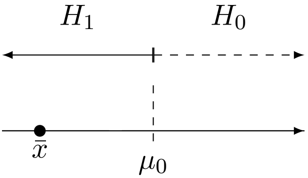


Caso 2: $H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

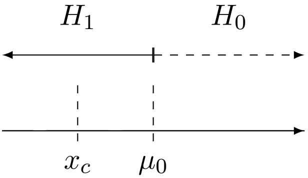


Caso 2: $H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

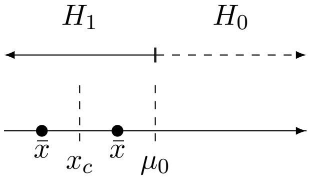


Caso 2: $H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

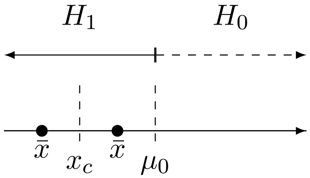


Caso 2: $H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu \geq \mu_0 \times H_1 : \mu < \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.



- Portanto, $RC_{H_0} = \{\bar{X} < x_c\}$;
- Como determinar x_c?

- Fixemos o nível de significância em α;
- Utilizando a mesma definição de z_α para o caso 1, isto é, z_α é o valor tal que $P(Z > z_\alpha) = \alpha$;
- Dessa forma, temos que

$$\begin{aligned}
\alpha = \sup \alpha(\mu) = \alpha(\mu_0) &= P(\text{Rejeitar } H_0 | \mu = \mu_0) \\
&= P(\bar{X} < x_c | \mu = \mu_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{x_c - \mu_0}{\sigma/\sqrt{n}}\right) \\
&\approx P\left(Z < \frac{x_c - \mu_0}{\sigma/\sqrt{n}}\right).
\end{aligned}$$

- Logo, $\frac{x_c - \mu_0}{\sigma/\sqrt{n}} = -z_\alpha \Rightarrow x_c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$;
- Assim, devemos rejeitar H_0 ao nível de significância α se

$$\bar{X} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Caso 3: $H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0;$$

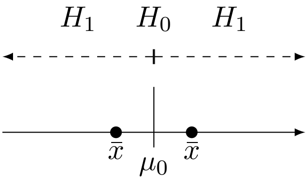
- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0 **ou** muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

Caso 3: $H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0 **ou** muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

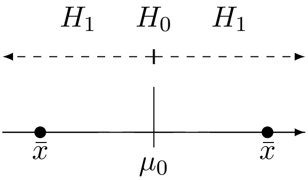


Caso 3: $H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0 **ou** muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

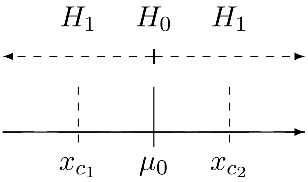


Caso 3: $H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0;$$

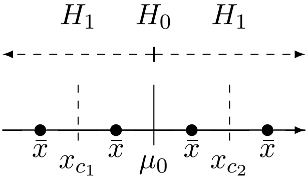
- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0 **ou** muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.



- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0 **ou** muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.

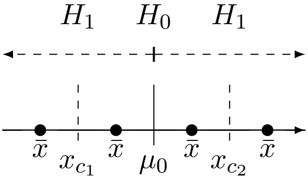


Caso 3: $H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0$

- Suponha que desejamos testar as hipóteses:

$$H_0 : \mu = \mu_0 \times H_1 : \mu \neq \mu_0;$$

- Podemos utilizar $\bar{X}$ para testar as hipósteses acima;
- Perceba que, se $\bar{X}$ fornecer um valor muito menor que μ_0 **ou** muito maior que μ_0, teremos evidências para rejeitar H_0 em favor de H_1.



- Portanto,

$$RC_{H_0} = \{\{\bar{X} < x_{c_1}\} \cup \{\bar{X} > x_{c_2}\}\};$$

- Como determinar x_{c_1} e x_{c_2}?

- Fixemos o nível de significância em α;
- Novamente, tomando z_α como o valor tal que $P(Z > z_\alpha) = \alpha$;
- Dessa forma, temos que

$$\begin{aligned}
\alpha &= P(\text{Rejeitar } H_0 | \mu = \mu_0) = P(\{\bar{X} < x_{c_1}\} \cup \{\bar{X} > x_{c_2}\} | \mu = \mu_0) \\
&= P(\bar{X} < x_{c_1} | \mu = \mu_0) + P(\bar{X} > x_{c_2} | \mu = \mu_0) \\
&= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{x_{c_1} - \mu_0}{\sigma/\sqrt{n}}\right) + P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{x_{c_2} - \mu_0}{\sigma/\sqrt{n}}\right) \\
&\approx \underbrace{P\left(Z < \frac{x_{c_1} - \mu_0}{\sigma/\sqrt{n}}\right)}_{\alpha/2} + \underbrace{P\left(Z > \frac{x_{c_2} - \mu_0}{\sigma/\sqrt{n}}\right)}_{\alpha/2}.
\end{aligned}$$

- Logo,

$$\frac{x_{c_1} - \mu_0}{\sigma/\sqrt{n}} = -z_{\frac{\alpha}{2}} \Rightarrow x_{c_1} = \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

  e

$$\frac{x_{c_2} - \mu_0}{\sigma/\sqrt{n}} = z_{\frac{\alpha}{2}} \Rightarrow x_{c_2} = \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}};$$

- Assim, devemos rejeitar H_0 ao nível de significância α se

$$\bar{X} < \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ ou } \bar{X} > \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Exemplo

Os sistemas de escapamento de uma aeronave funcionam devido a um propelente sólido. A taxa de queima é uma característica importante desse produto. As especificações exigem que a taxa de queima seja de 50 cm/s. Sabemos de experiências anteriores que essa variável tem desvio-padrão 2 cm/s. Uma amostra de tamanho 25 forneceu taxa média de queima de 51.3 cm/s. Quais conclusões podem ser tiradas ao nível de significância de 0.05?

Exemplo

Os sistemas de escapamento de uma aeronave funcionam devido a um propelente sólido. A taxa de queima é uma característica importante desse produto. As especificações exigem que a taxa de queima seja de 50 cm/s. Sabemos de experiências anteriores que essa variável tem desvio-padrão 2 cm/s. Uma amostra de tamanho 25 forneceu taxa média de queima de 51.3 cm/s. Quais conclusões podem ser tiradas ao nível de significância de 0.05?

- Temos que $\bar{x} = 51.3$, $\sigma = 2$ e $n = 25$;

Exemplo

Os sistemas de escapamento de uma aeronave funcionam devido a um propelente sólido. A taxa de queima é uma característica importante desse produto. As especificações exigem que a taxa de queima seja de 50 cm/s. Sabemos de experiências anteriores que essa variável tem desvio-padrão 2 cm/s. Uma amostra de tamanho 25 forneceu taxa média de queima de 51.3 cm/s. Quais conclusões podem ser tiradas ao nível de significância de 0.05?

- Temos que $\bar{x} = 51.3$, $\sigma = 2$ e $n = 25$;
- As hipóteses são

$$H_0 : \mu = 50 \;\times\; H_1 : \mu \neq 50.$$

  Logo, $\mu_0 = 50$;

Exemplo

Os sistemas de escapamento de uma aeronave funcionam devido a um propelente sólido. A taxa de queima é uma característica importante desse produto. As especificações exigem que a taxa de queima seja de 50 cm/s. Sabemos de experiências anteriores que essa variável tem desvio-padrão 2 cm/s. Uma amostra de tamanho 25 forneceu taxa média de queima de 51.3 cm/s. Quais conclusões podem ser tiradas ao nível de significância de 0.05?

- Temos que $\bar{x} = 51.3$, $\sigma = 2$ e $n = 25$;
- As hipóteses são

$$H_0 : \mu = 50 \ \times \ H_1 : \mu \neq 50.$$

  Logo, $\mu_0 = 50$;
- O nível de significância é 0.05, entao $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$;

Exemplo

Os sistemas de escapamento de uma aeronave funcionam devido a um propelente sólido. A taxa de queima é uma característica importante desse produto. As especificações exigem que a taxa de queima seja de 50 cm/s. Sabemos de experiências anteriores que essa variável tem desvio-padrão 2 cm/s. Uma amostra de tamanho 25 forneceu taxa média de queima de 51.3 cm/s. Quais conclusões podem ser tiradas ao nível de significância de 0.05?

- Temos que $\bar{x} = 51.3$, $\sigma = 2$ e $n = 25$;
- As hipóteses são

$$H_0 : \mu = 50 \ \times \ H_1 : \mu \neq 50.$$

  Logo, $\mu_0 = 50$;
- O nível de significância é 0.05, entao $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$;
- Os valores críticos são $x_{c_1} = 50 - 1.96\frac{2}{5} = 49.216$ e $x_{c_2} = 50 + 1.96\frac{2}{5} = 50.784$;

Exemplo

Os sistemas de escapamento de uma aeronave funcionam devido a um propelente sólido. A taxa de queima é uma característica importante desse produto. As especificações exigem que a taxa de queima seja de 50 cm/s. Sabemos de experiências anteriores que essa variável tem desvio-padrão 2 cm/s. Uma amostra de tamanho 25 forneceu taxa média de queima de 51.3 cm/s. Quais conclusões podem ser tiradas ao nível de significância de 0.05?

- Temos que $\bar{x} = 51.3$, $\sigma = 2$ e $n = 25$;
- As hipóteses são

$$H_0 : \mu = 50 \ \times \ H_1 : \mu \neq 50.$$

  Logo, $\mu_0 = 50$;
- O nível de significância é 0.05, entao $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$;
- Os valores críticos são $x_{c_1} = 50 - 1.96\frac{2}{5} = 49.216$ e $x_{c_2} = 50 + 1.96\frac{2}{5} = 50.784$;
- Como $\bar{x} > x_{c_2}$, chegamos a conclusão que devemos rejeitar H_0 ao nível de significância de 5%.

Estrutura

Tamanho amostral grande

- Suponha que n é grande o suficiente, para supormos, com base na consistência de S^2, que $S^2 \approx \sigma^2$;

Tamanho amostral grande

- Suponha que n é grande o suficiente, para supormos, com base na consistência de S^2, que $S^2 \approx \sigma^2$;
- Neste caso, todos os valores críticos vistos anteriormente podem ser aproximados substituindo σ^2 por S^2 em suas expressões;

Tamanho amostral grande

- Suponha que n é grande o suficiente, para supormos, com base na consistência de S^2, que $S^2 \approx \sigma^2$;
- Neste caso, todos os valores críticos vistos anteriormente podem ser aproximados substituindo σ^2 por S^2 em suas expressões;
- E se n não for suficientemente grande?

n pequeno (população normal)

- Considere o caso em que a variância σ^2 é desconhecida, e o tamanho amostral n é pequeno;

n pequeno (população normal)

- Considere o caso em que a variância σ^2 é desconhecida, e o tamanho amostral n é pequeno;
- Se pudermos supor que a v.a. X sendo observada segue distribuição normal;

n pequeno (população normal)

- Considere o caso em que a variância σ^2 é desconhecida, e o tamanho amostral n é pequeno;
- Se pudermos supor que a v.a. X sendo observada segue distribuição normal;
- Vimos anteriormente que

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1};$$

- Considere o caso em que a variância σ^2 é desconhecida, e o tamanho amostral n é pequeno;
- Se pudermos supor que a v.a. X sendo observada segue distribuição normal;
- Vimos anteriormente que

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1};$$

- Assim, todos os valores críticos anteriores podem ser considerados com σ^2 sendo substituido S^2. Entretanto, no lugar de z_α e $z_{\frac{\alpha}{2}}$, devemos utilizar $t_{(\alpha;n-1)}$ e $t_{(\frac{\alpha}{2};n-1)}$, respectivamente. Agora, $t_{(\alpha;n-1)}$ é o valor tal que $P\left(T > t_{(\alpha;n-1)}\right) = \alpha$.

Exemplo - cont.

Retornemos ao exemplo do propelente:

- Suponha que o desvio-padrão fosse desconhecido e que não podemos considerar $n = 25$ uma amostra grande o suficiente;

Exemplo - cont.

Retornemos ao exemplo do propelente:

- Suponha que o desvio-padrão fosse desconhecido e que não podemos considerar $n = 25$ uma amostra grande o suficiente;
- Além disso, suponha que a taxa de queima desse propelente tenha distribuição normal;

Exemplo - cont.

Retornemos ao exemplo do propelente:

- Suponha que o desvio-padrão fosse desconhecido e que não podemos considerar $n = 25$ uma amostra grande o suficiente;
- Além disso, suponha que a taxa de queima desse propelente tenha distribuição normal;
- Agora, considere que a amostra forneceu desvio-padrão 2;

Exemplo - cont.

Retornemos ao exemplo do propelente:

- Suponha que o desvio-padrão fosse desconhecido e que não podemos considerar $n = 25$ uma amostra grande o suficiente;
- Além disso, suponha que a taxa de queima desse propelente tenha distribuição normal;
- Agora, considere que a amostra forneceu desvio-padrão 2;
- Desejamos testar as mesmas hipóteses ao mesmo nível de significância. Logo, $\mu_0 = 50$;

Exemplo - cont.

Retornemos ao exemplo do propelente:

- Suponha que o desvio-padrão fosse desconhecido e que não podemos considerar $n = 25$ uma amostra grande o suficiente;
- Além disso, suponha que a taxa de queima desse propelente tenha distribuição normal;
- Agora, considere que a amostra forneceu desvio-padrão 2;
- Desejamos testar as mesmas hipóteses ao mesmo nível de significância. Logo, $\mu_0 = 50$;
- Assim, $t_{(\frac{\alpha}{2};n-1)} = t_{(0.025;24)} = 2.0639$;

Exemplo - cont.

Retornemos ao exemplo do propelente:

- Suponha que o desvio-padrão fosse desconhecido e que não podemos considerar $n = 25$ uma amostra grande o suficiente;
- Além disso, suponha que a taxa de queima desse propelente tenha distribuição normal;
- Agora, considere que a amostra forneceu desvio-padrão 2;
- Desejamos testar as mesmas hipóteses ao mesmo nível de significância. Logo, $\mu_0 = 50$;
- Assim, $t_{(\frac{\alpha}{2};n-1)} = t_{(0.025;24)} = 2.0639$;
- Os valores críticos são portanto

$$x_{c_1} = \mu_0 - t_{(\frac{\alpha}{2};n-1)} \frac{S}{\sqrt{n}} = 50 - 2.0639 \frac{2}{5} = 49.174$$

e

$$x_{c_2} = \mu_0 + t_{(\frac{\alpha}{2};n-1)} \frac{S}{\sqrt{n}} = 50 + 2.0639 \frac{2}{5} = 50.826;$$

Exemplo - cont.

Retornemos ao exemplo do propelente:

- Suponha que o desvio-padrão fosse desconhecido e que não podemos considerar $n = 25$ uma amostra grande o suficiente;
- Além disso, suponha que a taxa de queima desse propelente tenha distribuição normal;
- Agora, considere que a amostra forneceu desvio-padrão 2;
- Desejamos testar as mesmas hipóteses ao mesmo nível de significância. Logo, $\mu_0 = 50$;
- Assim, $t_{(\frac{\alpha}{2};n-1)} = t_{(0.025;24)} = 2.0639$;
- Os valores críticos são portanto

$$x_{c_1} = \mu_0 - t_{(\frac{\alpha}{2};n-1)}\frac{S}{\sqrt{n}} = 50 - 2.0639\frac{2}{5} = 49.174$$

e

$$x_{c_2} = \mu_0 + t_{(\frac{\alpha}{2};n-1)}\frac{S}{\sqrt{n}} = 50 + 2.0639\frac{2}{5} = 50.826;$$

- Como $\bar{x} > x_{c_2}$, rejeitar H_0 ao nível de significância de 5%.

FIM!!!