Introdução à Probabilidade e Estatística

Distribuições Amostrais
Estimação pontual e intervalar

Departamento de Matemática
Universidade de Évora

Ano lectivo de 2016/17

Patrícia Filipe

Distribuições Amostrais

Distribuição Amostral da Média: População normal e σ^2 conhecida

Se a distribuição da população for normal e a variância (σ^2) for conhecida então, a Distribuição amostral da média é

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

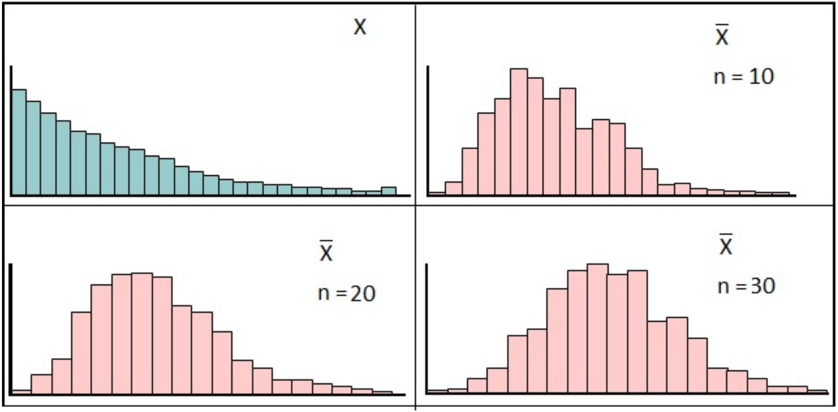
Diz-se que a distribuição por amostragem da média $\overline{X}$ segue uma distribuição normal com valor esperado μ e variância $\frac{\sigma^2}{n}$. Ao desvio padrão da média amostral ($\frac{\sigma}{\sqrt{n}}$) dá-se o nome de erro padrão.

Distribuição Amostral da Média: Teorema Limite Central

Teorema Limite Central (TLC)

Suponhamos que se recolhe uma amostra de dimensão n de uma dada população X (muito grande), com valor esperado μ e variância σ^2. Então, se a amostra for suficientemente grande, $\overline{X}$ pode ser modelada, aproximadamente, por um modelo **normal** com valor esperado μ e variância $\frac{\sigma^2}{n}$.

Distribuição Amostral da Média: Teorema Limite Central



Distribuição Amostral da Média: Teorema Limite Central

Teorema Limite Central

Seja $(X_1, X_2, \ldots, X_n)$ uma sucessão de v.a.'s i.i.d., tal que $E[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2$, $\forall i = 1, \ldots, n$, finita.

Considere-se $S_n = X_1 + \cdots + X_n$. Para valores grandes de n tem-se que:

$$\frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \dot{\sim} N(0, 1)$$

Consequentemente, quando $n \to \infty$,

$$\sqrt{n}\frac{\overline{X} - \mu}{\sigma} \to Z \dot{\sim} N(0, 1)$$

Na prática, considera-se amostra suficientemente grande quando $n \geqslant 30$.

Distribuição Amostral da Média (cont.)

Se a distribuição populacional não for normal, a variância for conhecida mas a amostra for suficientemente grande então (pelo TLC),

$$\overline{X} \dot{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Se a distribuição populacional for normal, a variância não for conhecida então,

$$\sqrt{n}\frac{\overline{X}-\mu}{S} \sim t_{(n-1)}$$

Distribuição Amostral da Média (cont.)

Se a distribuição populacional for normal, a variância não for conhecida mas a dimensão da amostra for suficientemente grande então (pela aprox. da t-Student pela Normal),

$$\overline{X} \dot{\sim} N\left(\mu, \frac{S}{\sqrt{n}}\right)$$

Distribuição Amostral da Proporção

Seja $X =$ "Número de sucessos do acontecimento A" tal que $X \sim B(n, p)$. Então,

$$\overline{P} = \frac{X}{n} \Rightarrow E[\overline{P}] = p \text{ e } \mathrm{Var}[\overline{P}] = \frac{np(1-p)}{n^2}$$

Para uma a.a. de dimensão elevada

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

$$\boxed{\overline{P} \dot{\sim} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)} \Rightarrow$$ Distribuição amostral da proporção.

Distribuição Amostral da Variância

Seja $(X_1, X_2, \ldots, X_n)$ uma sucessão de v.a.'s i.i.d. de uma população com variância σ^2, finita. Se a distribuição da população for normal então

$$\boxed{\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}} \Rrightarrow \quad \text{Distribuição amostral da variância}$$

Nota: Seja $Z_1, Z_2, \ldots, Z_n$ uma sucessão de v.a.'s i.i.d. Se $Z_i \sim N(0,1),\ i = 1, \ldots, n$, então, $\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$

Distribuições Amostrais duas populações

Estatística	Condições	Distribuição amostral
	σ_1^2, σ_2^2 conh., Pop. Normais	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
	σ_1^2, σ_2^2 conh. Pop. quaisquer, $n_1, n_2 > 30$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \dot{\sim} N(0,1)$
$\bar{X}_1 - \bar{X}_2$		
	σ_1^2, σ_2^2 desconh., $\sigma_1^2 = \sigma_2^2$ Pop. Normais, $n_1, n_2 \leqslant 30$	$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$
	σ_1^2, σ_2^2 desconh., $\sigma_1^2 = \sigma_2^2$ Pop. quaisquer, $n_1, n_2 > 30$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \dot{\sim} N(0,1)$

Distribuições Amostrais duas populações (cont.)

Estatística	Condições	Distribuição amostral
$\frac{S_1^2}{S_2^2}$	Pop. Normais	$F = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1,n_2-1}$
$\bar{P}_1 - \bar{P}_2$	$n_1, n_2 > 30$ Pop. Bernoulli	$Z = \frac{(\bar{P}_1 - \bar{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \dot{\sim} N(0,1)$

Estimação

Estimação Pontual

Um Estimador Pontual é uma qualquer função que depende unicamente dos dados. Designa-se usualmente por $T(X_1, X_2, \cdots, X_n) \equiv T_n$ o estimador pontual de θ.

Assim (genericamente) podemos dizer que qualquer estatística é um estimador pontual!

Estimador *versus* Estimativa

Estimador

Um estimador é uma função da amostra aleatória $(X_1, X_2, ..., X_n)$. Esta função não envolve parâmetros desconhecidos e o seu contradomínio é o espaço de estados dos parâmetros. Um estimador é uma variável aleatória.

Estimativa

Uma estimativa é o valor concreto assumido pelo estimador para uma amostra particular $(x_1, x_2, ..., x_n)$. É um número.

Em termos formais:

- Represente-se uma dada população por X.
- Digamos que esse modelo possui função de distribuição (f.d.) $F(x) = P(X \leqslant x)$. Mais concretamente, possui f.d. dada por $F(x;\theta) = F(x|\theta)$, onde θ representa o parâmetro (ou o vector de parâmetros) do modelo.
- O objectivo consiste em estimar o parâmetro $\theta \in \Theta$ (com Θ o espaço de parâmetros).

Por exemplo:

$$X \sim N(\mu, \sigma)$$

$\theta = (\theta_1, \theta_2) = (\mu, \sigma)$ e $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$.

$\hat{\theta}_1 \equiv \hat{\mu}$=?
e/ou
$\hat{\theta}_2 \equiv \hat{\sigma}$=?

Alguns Estimadores Pontuais Mais Usuais

1. Estimador da média μ

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

onde $(x_1, x_2, \cdots, x_n)$ é uma realização da amostra aleatória $(X_1, X_2, \cdots, X_n)$.

2. Estimador da variância σ^2

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

ou?

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

Alguns Estimadores Pontuais Mais Usuais

3. Estimador da proporção p
   Seja $X =$ "Número de sucessos do acontecimento A"
   tal que $X \sim B(n, p)$.
   Então,
   $$\overline{P} = \frac{X}{n}$$
   é o estimador pontual do parâmetro p, quando a dimensão da amostra (n) é conhecida.

Métodos de Estimação: Método dos momentos (Karl Pearson, 1902)

Seja θ o parâmetro (ou vector de parâmetros) desconhecido do modelo ao qual temos associada uma a.a. de dimensão n.
Para este modelo considere-se:

$F(x|\theta)$ Função de distribuição

$f(x|\theta)$ Função densidade de probabilidade (f.d.p.) (ou f.m.p.)

- Considerar todos os momentos populacionais (supondo que existem) até à k-ésima ordem.
- Escrever os seus correspondentes momentos amostrais ($M_r' = n^{-1}\sum_{i=1}^{n} X_i^r,\ 1 \leqslant r \leqslant k$)
- Obter um sistema de k equações em ordem aos k parâmetros do modelo.

Um Exemplo

E, por último,

- Resolver este sistema de equações, de onde se obtêm os k estimadores dos parâmetros do modelo.

Seja $(X_1, X_2, \cdots, X_n)$ uma a.a. (i.i.d.) de $N(\mu, \sigma)$.

$$\theta = (\mu, \sigma^2) \text{ e } \Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in (0, \infty)\}$$

$$(\theta_1 = \mu \text{ e } \theta_2 = \sigma^2)$$

$$E[X] = \mu \quad \text{e} \quad E[X^2] = \mu^2 + \sigma^2$$

$$\begin{cases} \mu = \frac{1}{n}\sum_{i=1}^{n} X_i \\ \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{cases}$$

$$\boxed{\hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = \frac{n-1}{n}S^2}$$

Outros Métodos de Estimação

- Método dos Momentos;
- Método dos Mínimos Quadrados (usado, por exemplo, na Regressão Linear);
- Método da Máxima Verosimilhança;
- etc...

Estimadores Pontuais: Quais?

Existindo mais do que um estimador para um dado parâmetro, como escolher o mais indicado?

O mais indicado é aquele que possui as melhores propriedades $\Rrightarrow$ O estudo das Propriedades dos Estimadores

Estimativas do valor esperado — Exemplo

1. À medida que se aumenta a dimensão da amostra espera-se que a estimativa do valor esperado seja cada vez melhor (no sentido, de cada vez mais próxima do verdadeiro valor)
2. Se para a mesma dimensão da amostra se calcular a correspondente média amostral esta vai, portanto, variar de amostra para amostra mas apresentar um comportamento que é característico de uma distribuição simétrica e com com variabilidade pequena.

Propriedades dos Estimadores

1. O estimador de θ deverá ser uma estatística. Ou seja, ser uma função unicamente dos dados.
2. O estimador deverá ter como valor esperado o parâmetro que se pretende estimar. Neste caso, diz-se que o estimador é centrado ou não enviesado.
3. De entre todos os estimadores centrados de θ deverá escolher-se aquele que tiver a menor variância.

Propriedades dos Estimadores

Seja X uma população caracterizada por um parâmetro θ, $(X_1, X_2, ..., X_n)$ uma amostra aleatória dessa população e $\hat{\theta} = \hat{\theta}\,(X_1, X_2, ..., X_n)$ um estimador de θ

- $\hat{\theta}$ diz-se centrado ou não enviesado se $E\left[\hat{\theta}\right] = \theta$.
  Se $\hat{\theta}$ não satisfaz a propriedade anterior diz-se não centrado ou enviesado, e chama-se viés do estimador $\hat{\theta}$ a $Viés\left[\hat{\theta}\right] = E\left[\hat{\theta}\right] - \theta$.
- Se $\hat{\theta}$ e $\tilde{\theta}$ são dois estimadores centrados de θ.
  O estimador $\hat{\theta}$ diz-se mais eficiente do que $\tilde{\theta}$ se $Var\left[\hat{\theta}\right] < Var\left[\tilde{\theta}\right]$.

Propriedades dos Estimadores

- Um estimador $\hat{\theta}$ diz-se consistente quando, para qualquer número real $\varepsilon > 0$, se verifica:
  $\lim_{n\to\infty} P\left[\left|\hat{\theta}-\theta\right| < \varepsilon\right] = 1.$
  As condições

  $$i)\ \lim_{n\to\infty} E\left[\hat{\theta}\right] = \theta;$$
  $$ii)\ \lim_{n\to\infty} Var\left[\hat{\theta}\right] = 0.$$

  são suficientes para que $\hat{\theta}$ seja estimador consistente.

Estimação Intervalar

Na estimação pontual, pretende-se obter um valor que, com base na informação amostral, seja o melhor valor para estimar um determinado parâmetro populacional.

Na estimação intervalar constrói-se um intervalo de valores que, com um certo grau de certeza, previamente estipulado, contenha o verdadeiro valor do parâmetro.

Estimação Intervalar

Seja $(X_1, X_2, \cdots, X_n)$ uma a.a. (i.i.d.) de uma população com f.d.p. (ou f.m.p.) $f(x|\theta)$, com $\theta \in \Theta$ parâmetro desconhecido. Se $\Theta \subset \mathbb{R}$, fala-se em *estimação intervalar*

Uma estimativa intervalar de $\theta \in \mathbb{R}$ é um qualquer par de funções (estatísticas) de $\overrightarrow{x} = (x_1, x_2, \cdots, x_n)$, $L(x_1, x_2, \cdots, x_n)$ e $U(x_1, x_2, \cdots, x_n)$
tais que $L(\overrightarrow{x}) \leqslant U(\overrightarrow{x})$, $\forall \overrightarrow{x}$ e

$$I(\overrightarrow{x}) = L(\overrightarrow{x}) \leqslant \theta \leqslant U(\overrightarrow{x})$$

Ao intervalo $[L(\overrightarrow{X}), U(\overrightarrow{X})]$ chama-se *estimador intervalar* de θ.

Métodos de construção do Intervalo de Confiança

Construção através da Variável Pivot

Construção do intervalo de confiança para o parâmetro θ: consiste em encontrar uma v.a. $Z(\overrightarrow{X}, \theta)$ cuja distribuição seja independente de θ.

A obtenção de um intervalo de confiança a $(1-\alpha) \times 100\%$ para θ através do método da variável *pivot* consiste na obtenção da variável *pivot* $Z(\overrightarrow{X}, \theta)$ e na obtenção de pontos a e b (independentes de θ) tal que

$$P(a < Z(\overrightarrow{X}, \theta) < b) \geqslant 1-\alpha$$

e, posteriormente, na inversão destas desigualdades em ordem ao parâmetro θ.

1 - α é o grau ou nível de confiança.

Intervalo de Confiança para a Média

Seja $(X_1, X_2, \cdots, X_n)$ uma a.a. i.i.d. $N(\mu, \sigma^2)$, com σ^2 conhecida. Dado $\alpha \in (0, 1)$, pretende-se construir um intervalo de confiança a $(1-\alpha) \times 100\%$ para μ.

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

$$Z = \sqrt{n}\frac{\overline{X} - \mu}{\sigma} \sim N(0, 1)$$ Z é uma variável *pivot* para μ

$$P(a < Z < b) = 1 - \alpha \Rightarrow a = -Z_{1-\frac{\alpha}{2}} \quad \text{e} \quad b = Z_{1-\frac{\alpha}{2}}$$

$$P(\overline{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Intervalo de Confiança para a Média

Por outras palavras,

$$\left]\overline{X} - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right[$$

é o intervalo de confiança a $(1-\alpha)\times 100\%$ para μ.

Obs. Quanto maior for a confiança com que se queira estimar o intervalo, maior será a amplitude deste.
À metade da amplitude de um intervalo de confiança designa-se por margem de erro

Interpretação de Nível de Confiança

Como exemplo, considere-se um nível de confiança de 95%.

1. Recolha-se uma amostra aleatória de dimensão n, estime-se a média e o desvio padrão amostrais e construa-se o respectivo intervalo de confiança para o valor médio.
2. Repitam-se estes passos para 100 amostras aleatórias da mesma população.

O que se pode dizer é que cerca de 95% desses intervalos contêm o valor médio μ, enquanto que os restantes 5% não contêm esse parâmetro.

Um intervalo de confiança deverá ser estimado de modo a possuir uma amplitude pequena.
Como diminuir a amplitude do intervalo?

1. Ao diminuir a confiança, diminui-se a amplitude do intervalo. Pouco recomendável

2. Ao aumentar a dimensão da amostra, para uma confiança fixa, diminui-se a amplitude do intervalo.

Intervalo de Confiança para a Média

Seja $(X_1, X_2, \cdots, X_n)$ uma a.a. i.i.d. $N(\mu, \sigma^2)$, com σ^2 desconhecida. Dado $\alpha \in (0, 1)$, pretende-se construir um intervalo de confiança a $(1-\alpha) \times 100\%$ para μ.

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

$$P(\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}) = 1-\alpha$$

$$\left]\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right[$$

é o intervalo de confiança a $(1-\alpha) \times 100\%$ para μ.

- Se $n > 30$, $Z = \sqrt{n}\frac{\overline{X}-\mu}{S} \dot{\sim} N(0, 1)$

$$\left]\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right[$$

é o intervalo de confiança a $(1-\alpha) \times 100\%$ para μ.

Intervalo de Confiança para a Proporção

Seja $(X_1, X_2, \cdots, X_n)$ uma a.a. i.i.d. $Ber(p)$, com p parâmetro desconhecido. Dado $\alpha \in (0,1)$, pretende-se construir um intervalo de confiança a $(1-\alpha)\times 100\%$ para p.

$$Z = \frac{\bar{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \dot{\sim} N(0,1), \quad n > 30$$

$$P\left(\bar{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} < p < \bar{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right) = 1-\alpha$$

$$\left]\bar{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

é o intervalo de confiança a $(1-\alpha)\times 100\%$ para p.
O intervalo anterior resulta apropriado para $n\bar{p} \geqslant 10$ e $n(1-\bar{p}) \geqslant 10$.

Margem de Erro dos Intervalo de Confiança

Margem de Erro do IC para μ

- σ^2 conhecida: $Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$
- σ^2 desconhecida: $t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$
- σ^2 desconhecida e $n > 30$: $Z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$

Margem de Erro do IC para p

$$Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

Margem de Erro dos Intervalo de Confiança

Questões:

1. Qual a dimensão da amostra a considerar de modo a que a margem de erro associada ao IC para μ (considerando σ^2 conhecida e desconhecida) seja inferior a β?

2. A mesma questão mas agora para o IC para p (considerando $\bar{p}$ conhecido e desconhecido)?

Intervalo de Confiança para a Variância

Seja $(X_1, X_2, \cdots, X_n)$ uma a.a. i.i.d. proveniente de uma população Normal.

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

$$P(a < \frac{(n-1)S^2}{\sigma^2} < b) = 1 - \alpha$$

Resolvendo as desigualdades em ordem ao parâmetro de interesse resulta,

$$\left] \frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}; \frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}} \right[$$

é o intervalo de confiança a $(1-\alpha) \times 100\%$ para σ^2.

Intervalo de Confiança para a Diferença de Médias

Considere-se duas amostras de dimensões n_1 e n_2 de duas populações independentes X_1 e X_2 com valores médios μ_1 e μ_2 e variâncias σ_1^2 e σ_2^2.
Admitamos que as duas populações são normais e que σ_1^2 e σ_2^2 são conhecidas.

$$\left]\overline{X_1} - \overline{X_2} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; \overline{X_1} - \overline{X_2} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right[$$

é o intervalo de confiança a $(1-\alpha)\times 100\%$ para $\mu_1 - \mu_2$ com σ_1^2 e σ_2^2 são conhecidas.

Intervalo de Confiança para a Diferença de Médias

Considere-se duas amostras de dimensões n_1 e n_2 de duas populações independentes X_1 e X_2 com valores médios μ_1 e μ_2 e variâncias σ_1^2 e σ_2^2.

Admitamos que as duas populações são normais e que σ_1^2 e σ_2^2 são desconhecidas, mas $\sigma_1^2 = \sigma_2^2$.

$$\left]\overline{X_1} - \overline{X_2} - t_{n_1+n_2-2;1-\frac{\alpha}{2}}\, S^*; \overline{X_1} - \overline{X_2} + t_{n_1+n_2-2;1-\frac{\alpha}{2}}\, S^*\right[$$

com

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

é o intervalo de confiança a $(1-\alpha)\times 100\%$ para $\mu_1 - \mu_2$ com σ_1^2 e σ_2^2 são desconhecidas.

Intervalo de Confiança para a Diferença de Médias

Considere-se duas amostras de dimensões n_1 e n_2 de duas populações independentes X_1 e X_2 com valores médios μ_1 e μ_2 e variâncias σ_1^2 e σ_2^2.
que σ_1^2 e σ_2^2 são desconhecidas, mas $\sigma_1^2 = \sigma_2^2$ e $n_1 > 30$ e $n_2 > 30$

$$\left]\overline{X_1} - \overline{X_2} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}; \overline{X_1} - \overline{X_2} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right[$$

é o intervalo de confiança a $(1-\alpha) \times 100\%$ para $\mu_1 - \mu_2$ com σ_1^2 e σ_2^2 são desconhecidas.

Intervalo de Confiança para a Diferença de Proporções

Considere-se duas amostras de dimensões n_1 e n_2 de duas populações Bernoulli independentes X_1 e X_2, com p_1 e p_2 as probabilidades de sucesso associadas às populações X_1 e X_2, respectivamente.
Admitamos que $n_1 > 30$ e $n_2 > 30$.

$$\left]\overline{p_1} - \overline{p_2} - Z_{1-\frac{\alpha}{2}} P^*; \overline{p_1} - \overline{p_2} + Z_{1-\frac{\alpha}{2}} P^*\right[$$

com

$$P^* \sqrt{\frac{\overline{p_1}(1-\overline{p_1})}{n_1} + \frac{\overline{p_2}(1-\overline{p_2})}{n_2}}$$

é o intervalo de confiança a $(1-\alpha) \times 100\%$ para $p_1 - p_2$.

Intervalo de Confiança para a Diferença de Variâncias

Considere-se duas amostras de dimensões n_1 e n_2 de duas populações normais independentes X_1 e X_2, com σ_1^2 e σ_2^2 as variâncias populacionais (desconhecidas).

$$\left] \frac{1}{F_{n_1-1,n_2-1;1-\frac{\alpha}{2}}} \frac{S_1^2}{S_2^2}; \; \frac{1}{F_{n_1-1,n_2-1;\frac{\alpha}{2}}} \frac{S_1^2}{S_2^2} \right[$$

é o intervalo de confiança a $(1-\alpha)\times 100\%$ para $\frac{\sigma_1^2}{\sigma_2^2}$.