INFERÊNCIAS SOBRE QUALIDADE DO PROCESSO

Aula 05

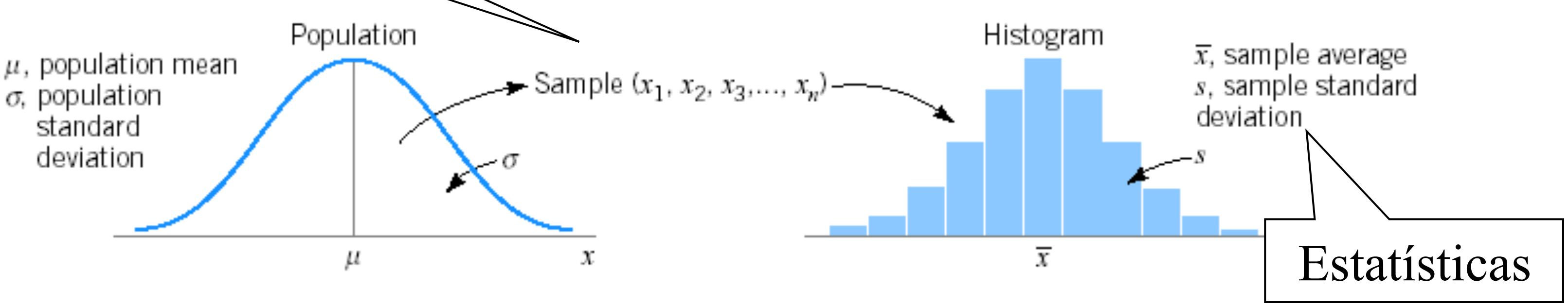
Profa. Carmela Maria Polito Braga – DELT/EEUFMG

ESTATÍSTICAS E DISTRIBUIÇÕES AMOSTRAIS

A inferência estatística usa quantidades calculadas a partir das observações na amostra.

Estatística é qualquer função dos dados amostrais que não contenha parâmetros desconhecidos.

Amostra aleatória



Se a distribuição de probabilidade da população de onde foi tirada a amostra é conhecida, pode-se determinar a distribuição de probabilidade de várias estatísticas calculadas a partir dos dados amostrais.

A distribuição de probabilidade de uma estatística é chamada de distribuição amostral.

AMOSTRANDO DE UMA DISTRIBUIÇÃO NORMAL

Distribuição Chi-quadrada (χ^2)

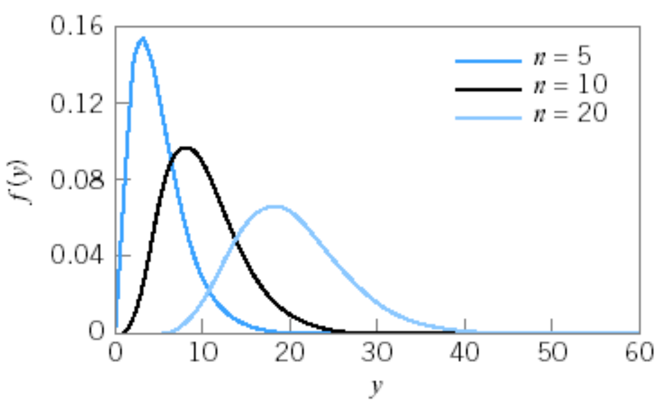
Do teorema do limite central sabemos que qualquer que seja a distribuição da população, a distribuição do somatório de x_i é aproximadamente normal com média nμ e variância nσ^2. Então a distribuição amostral da média amostral para qq distribuição de população é:

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

Ou seja, a média amostral, $\overline{X}$, tem distribuição Normal, N(μ,σ/ n)

$$\text{Se} \quad y = x_1^2 + x_2^2 + x_3^2 + \ldots x_n^2 \Rightarrow f(y) = \frac{1}{2^{n/2}\Gamma\left(n/2\right)} y^{\left(n/2\right)-1} e^{-y/2}$$

A distribuição χ2 (qui-quadrado) é uma importante distribuição amostral da distribuição normal.



Se x_1, x_2,...x_n são variáveis aleatórias independentes, com média zero e normalmente distribuidas N(0,1), então y tem distribuição χ2 com n graus de liberdade.
Distribuição assimétrica com média μ=n e variância σ^2=2n.

AMOSTRANDO DE UMA DISTRIBUIÇÃO NORMAL
Distribuição Chi-quadrada (χ^2)

Para ilustrar o uso da distribuição $\chi 2$, suponha que $x_1, x_2, \ldots x_n$ seja uma amostra aleatória de uma distribuição N(μ;σ^2). Então a variável aleatória

$$y = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}$$

tem distribuição $\chi 2$ com n-1 graus de liberdade. Entretanto, usando a equação que define a variância amostral,

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

pode-se reescrever a equação anterior como:

$$y = \frac{(n-1)S^2}{\sigma^2}$$

A variância amostral, devidamente normalizada, tem distribuição qui- quadrado com n-1 graus de liberdade.

Ou seja, a distribuição amostral de $(n-1)S^2 / \sigma^2$ é χ^2_{n-1}

quando a mostra é retirada de uma distribuição normal.

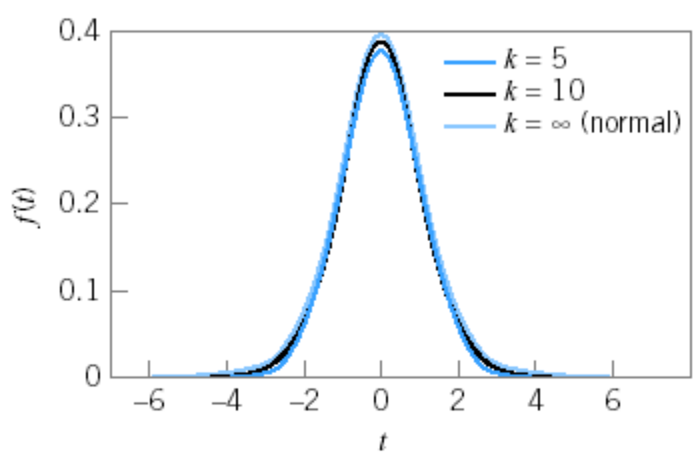
DISTRIBUIÇÃO T

Outra distribuição amostral importante é a distribuição *t*. Se x é uma variável aleatória normal padrão e se y é uma variável aleatória $\chi 2$ com k graus de liberdade, e x e y são independentes, então a variável aleatória

$$t = \frac{x}{\sqrt{y/k}}$$

tem distribuição *t* com k graus de liberdade. A distribuição de probabilidade de t é

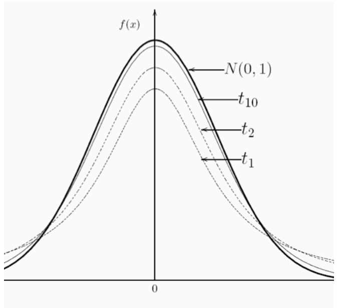
$$f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}\,\Gamma(k/2)}\left(\frac{t^2}{k}+1\right)^{-(k+1)/2} \qquad -\infty < t < \infty$$



A distribuição t de Student aparece no problema de determinação da média de uma população (que segue a distribuição normal) a partir de uma amostra. Neste problema, não se sabe qual é a média ou o desvio padrão da população, mas ela deve ser normal. Supondo que o tamanho da amostra *n* seja muito menor que o tamanho da população, temos que a amostra é dada por *n* variáveis aleatórias normais independentes $X_1, ..., X_n$, cuja média $\bar{X}$ é o melhor estimador para a média da população.

DISTRIBUIÇÃO T

- A distribuição *t* de Student varia de acordo com a dimensão da amostra que vai determinar o número de graus de liberdade.
- A curva da distribuição *t* de Student tem a mesma forma em sino da distribuição Normal, mas reflete a maior variabilidade (com curvas maisalargadas, caldas mais longas), esperado em amostras pequenas.
- A distribuição *t* de Student tem valor médio zero (tal como a distribuição Normal padrão).
- O desvio padrão da distribuição *t* de Student varia de acordo com o tamanho da amostra e é maior do que 1 (o que não acontece com a distribuição Normal standard, onde $\sigma = 1$).
- Quanto maior a dimensão da amostra, mais a distribuição *t* de Student se aproxima da distribuição Normal.



DISTRIBUIÇÃO F (DE FISHER)

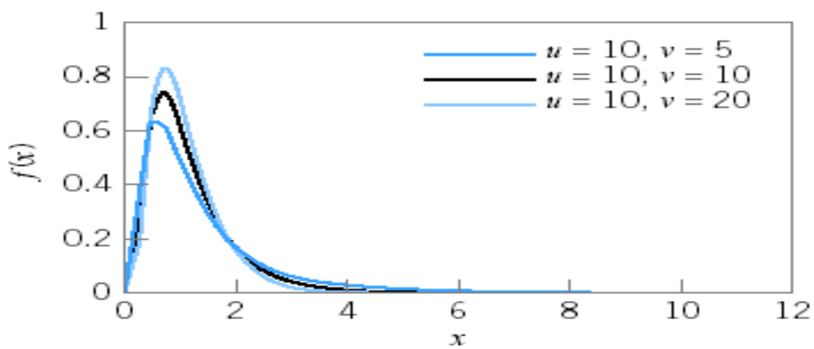
A outra distribuição amostral baseada em processo normal a ser considerada é a distribuição *F*. Se w e y são duas variáveis aleatórias independentes com distribuição χ2 com u e v graus de liberdade, respectivamente, então a razão

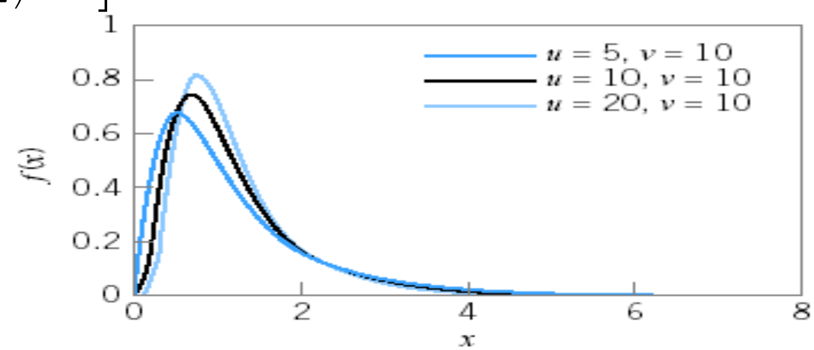
$$F_{u,v} = \frac{w/u}{y/v}$$

tem distribuição *F* com u graus de liberdade no numerador e y graus de liberdade no denominador. (Mede a razão entre duas qui quadrado. É importante para inferência sobre variâncias.)

Se x é uma variável aleatória F com u graus de liberdade no numerador e y graus de liberdade no denominador, então sua distribuição de probabilidade é:

$$f(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)} \frac{x^{(u/2)-1}}{\left[\left(\frac{u}{2}\right)x+1\right]^{(u+v)/2}} \qquad 0 < x < \infty$$





DISTRIBUIÇÃO F

Como exemplo de uma variável aleatória F, suponha que temos dois processos normais, independentes, tais como $x_1 \sim N(\mu_1;\sigma_1^2)$ e $x_2 \sim N(\mu_2;\sigma_2^2)$.

Seja $x_{11}, x_{12}, ..., x_{1n}$ uma amostra aleatória de n_1 observações do primeiro processo normal e seja $x_{21}, x_{22}, ..., x_{2n}$ uma amostra aleatória de tamanho n_2 do segundo processo. Se S_1^2 e S_2^2 são as respectivas variâncias amostrais, então:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

Este resultado provém diretamente da distribuição amostral de S^2, discutida anteriormente. A distribuição *F* será usada na inferência sobre as variâncias de duas distribuições normais *(distribuição importante na análise de CEP multivariado).*

INFERÊNCIA ESTATÍSTICA PARA UMA AMOSTRA

As técnicas de inferência estatística podem ser classificadas em duas amplas categorias: estimação de parâmetros e teste de hipóteses.

Uma hipótese estatística é uma afirmativa sobre os valores dos parâmetros de uma distribuição de probabilidade. P.ex., se achamos que o diâmetro de um mancal é 1,5in, podemos expressar esta afirmativa como

Hipótese Nula

Hipótese Alternativa

$$H_o : \mu = 1,5$$

$$H_1 : \mu \neq 1,5$$

Neste exemplo, H_1 é uma hipótese alternativa bilateral, pois este especifica o diâmetro médio que pode ser maior ou menor que este valor. Em outros problemas hipóteses unilaterais podem ser mais apropriadas.

INFERÊNCIA ESTATÍSTICA PARA UMA AMOSTRA

- Os procedimentos de teste de hipótese são muito úteis em vários problemas de CEP. Também formam a base para a maioria das técnicas de CEP que estudaremos a partir de agora.
- Uma parte importante no problema de teste de hipótese é a determinação dos valores dos parâmetros especificados nas hipóteses nula e alternativa. Eles podem ser obtidos:
    - ✓ de evidência ou conhecimento anterior. Freqüente em CEP, quando usamos dados históricos para projeto e periodicamente testamos a hipótese de que este valor não mudou;
    - ✓ de alguma teoria ou modelo do processo;
    - ✓ de uma especificação contratual ou de projeto, o que é freqüente também.

INFERÊNCIA ESTATÍSTICA PARA UMA AMOSTRA

- Para testar uma hipótese, toma-se uma amostra aleatória da população em estudo, calcula-se uma estatística de teste apropriada e, então, rejeita-se ou não a hipótese nula **Ho**.
- O conjunto de valores da estatística de teste que levam `a rejeição de **Ho** é chamado de **região crítica** ou **região de rejeição** do teste.
- Dois tipos de erros podem ser cometidos quando testamos hipóteses:
  - a rejeição da hipótese nula quando ela é verdadeira, o que é chamado de **erro tipo I**;
  - a não-rejeição da mesma quando ela é falsa, o que é chamado de **erro tipo II**.

INFERÊNCIA ESTATÍSTICA PARA UMA AMOSTRA

As probabilidades de ocorrência destes tipos de erro são denotadas como:

$$\alpha = P\{\text{erro tipo I}\} = P\{\text{rejeitar } H_o \mid H_o \text{ é verdadeira}\}$$

$$\beta = P\{\text{erro tipo II}\} = P\{\text{não rejeitar } H_o \mid H_o \text{ é falsa}\}$$

Pode ser conveniente trabalhar com o poder de um teste, onde:

$$\text{Poder} = 1 - \beta = P\{\text{rejeitar } H_o \mid H_o \text{ é falsa}\}$$

- Em outras palavras, o poder é a probabilidade de rejeitar corretamente Ho.
  - Em controle de qualidade **α** é às vezes chamado de **Risco do Fabricante**, por denotar a probabilidade de um lote bom ser rejeitado ou um processo produzindo com qualidade ser rejeitado como produzindo insatisfatoriamente.
  - Já **β** é às vezes chamado de **Risco do Consumidor**, por denotar a probabilidade de aceitação de um lote de baixa qualidade ou aceitar que um processo produzindo fora da especificação continue em operação.