Amostragem

Cuidados a ter na amostragem

Tipos de amostragem

Distribuições de amostragem

Muito Importante!!

Em relação às amostras, deve assegurar-se a sua representatividade relativamente à população de onde foram retiradas.

Exemplo: amostras não representativas (I)

- Amostras inadequadas: em geral não acontece inadvertidamente. É intencional a generalização abusiva do que foi observado numa amostra muito diminuta para uma população de dimensão considerável.

  Por exemplo: 9 em cada 10 estrelas preferem...

Exemplo: amostras não representativas (II)

- Amostras auto-seleccionadas: apresentar uma questão e solicitar os espectadores que telefonem para um número se a sua opinião é "sim" e para outro número se a sua opinião é "não".
- Deve haver pagamento de propinas no ensino superior público?
- Deveríamos voltar a ter monarquia em Portugal?

Tipos de amostragem

- Amostragem aleatória (ou casual) – é possível calcular, a priori, a probabilidade de observar cada indivíduo da população na amostra.
- Amostragem determinística (ou dirigida)– não é possível calcular essa probabilidade.

Amostragem aleatória

- Alguns tipos de amostragem aleatória:
  - Amostragem aleatória simples: Cada elemento da amostra é retirado aleatoriamente de toda a população (com ou sem reposição). Assim, cada possível amostra tem a mesma probabilidade de ser recolhida.

    Este é o tipo de amostragem que iremos considerar ao longo da disciplina. Assim, daqui em diante, amostra aleatória designa um conjunto de variáveis aleatórias independentes e com a mesma distribuição.

Amostragem aleatória

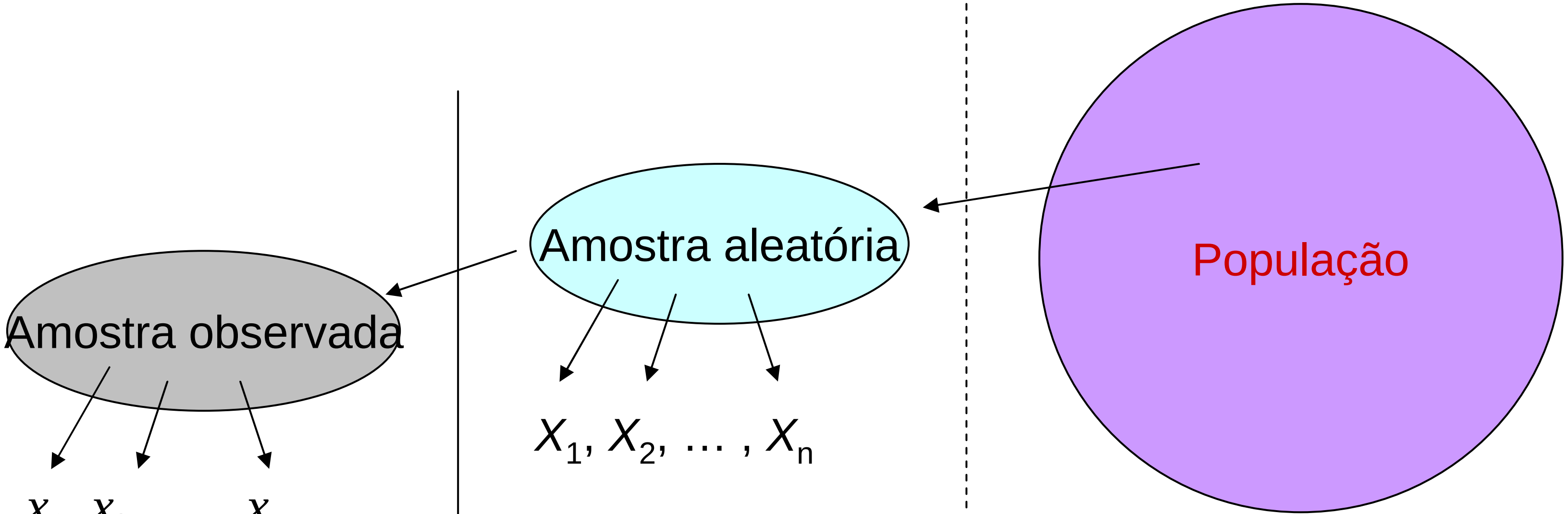
- Amostragem por clusters: dividir a população em secções (ou clusters); seleccionar aleatoriamente alguns desses clusters; escolher todos os membros dos clusters seleccionados.

- Amostragem estratificada: subdividir a população em, pelo menos, dois subgrupos distintos que partilham alguma característica e, em seguida, recolher uma amostra de cada um dos subgrupos (ou estratos).

Objectivos da recolha de amostras

- Estimar características desconhecidas de uma população
- Testar hipóteses sobre parâmetros de uma população

Esquema de amostragem



$x_1, x_2, \ldots, x_n$

x_i representa um número.

Medidas amostrais:

$\bar{x}$

s^2

$X_1, X_2, \ldots, X_n$

$X_i \sim F(x)$

$E[X_i] = \mu$

$Var[X_i] = \sigma^2$

$E[\bar{X}] = \mu$

$Var[\bar{X}] = \sigma^2/n$

Modelo probabilístico:

$F(x)$

Parâmetros:

μ

σ^2

Distribuições de amostragem

- Se X_1, X_2, … , X_n representa uma amostra aleatória com distribuição *F(x)*, qualquer função destas variáveis vai ser uma nova variável aleatória com uma certa distribuição. Chama-se a essa distribuição: distribuição de amostragem.

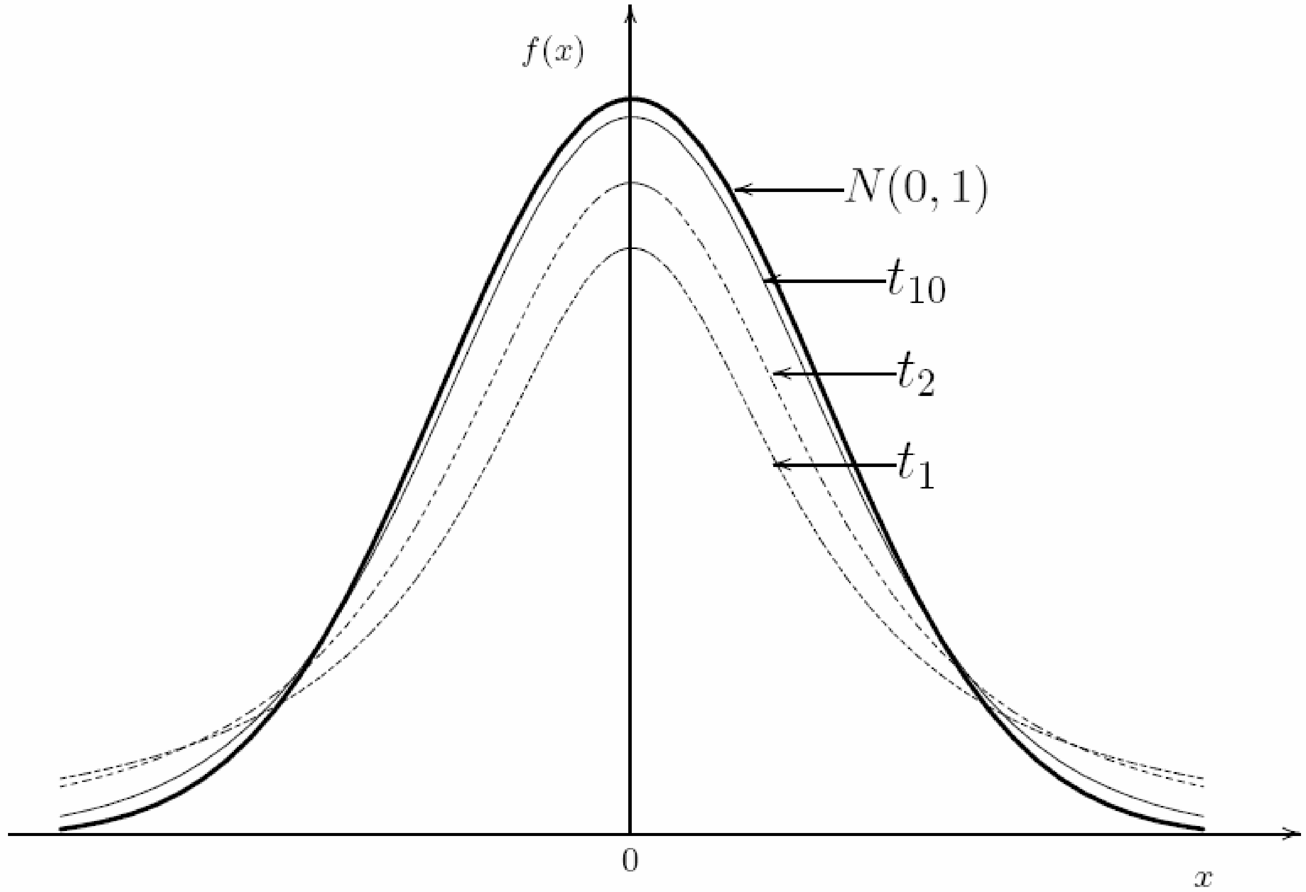
  Por exemplo: se X_1, X_2, … , X_n têm distribuição Normal, $N(\mu,\sigma^2)$, a média amostral, $\bar{X}$, tem distribuição Normal, $N(\mu,\sigma^2/n)$.

  Vamos ver outras distribuições de amostragem que surgem associadas a estatísticas calculadas em populações Normais.

Distribuição t (de Student)

- A distribuição *t* (de Student) é uma família de distribuições indexada por um parâmetro, que representa o número de graus de liberdade (g.l.). Quando *X* tem distribuição t com *n* graus de liberdade escreve-se $X \sim t_n$.

Distribuição t (de Student): curvas de densidade



Propriedades da 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 reflecte a maior variabilidade (com curvas mais alargadas) que é de esperar em amostras pequenas.
- A distribuição *t* de Student tem valor médio zero (tal como a distribuição Normal standard).
- 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 t

- Se X_1, X_2, … , X_n representa uma amostra aleatória com distribuição Normal, $N(\mu,\sigma^2)$, a seguinte variável tem distribuição Normal standard

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

e substituindo σ por S passamos a ter uma distribuição t com n-1 graus de liberdade.

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

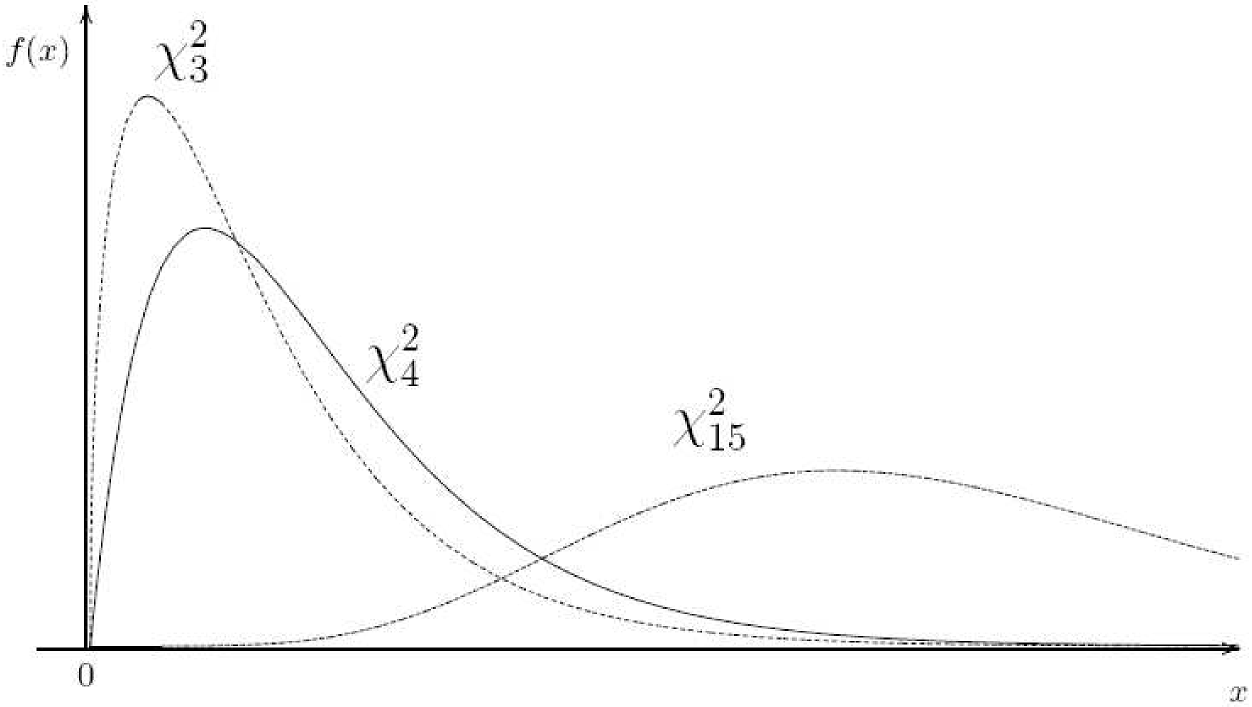
Distribuição t

- Para obter probabilidades e quantis da distribuição t podemos recorrer a software ou a tabelas.
- As tabelas fornecem habitualmente quantis da distribuição.
- No SPSS as funções associadas à distribuição t com n graus de liberdade são:
  - Cdf.t(x,n) para a função de distribuição no ponto x, $F(x)$;
  - Idf.t(p,n) para o quantil de ordem p, $t_{p,\, n}$.

Distribuição do qui-quadrado

- A distribuição do qui-quadrado é a uma família de distribuições indexada por um parâmetro, que representa o número de graus de liberdade (g.l.). Quando X tem distribuição do qui-quadrado n graus de liberdade escreve-se $X \sim \chi^2_n$.

Distribuição do qui-quadrado: curvas de densidade



Distribuição do qui-quadrado

- Se $X_1, X_2, \ldots, X_n$ representa uma amostra aleatória com distribuição Normal, $N(\mu,\sigma^2)$, a variância amostral devidamente normalizada tem distribuição do qui-quadrado com n-1 graus de liberdade,

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

- A distribuição do qui-quadrado é assimétrica positiva.
- A sua média aumenta com a dimensão da amostra.
- Quanto maior a dimensão da amostra, mais a distribuição do qui-quadrado se aproxima de uma distribuição Normal.

Distribuição do qui-quadrado

- Para obter probabilidades e quantis da distribuição do qui-quadrado podemos recorrer a software ou a tabelas.
- As tabelas fornecem habitualmente quantis da distribuição.
- No SPSS as funções associadas à distribuição do qui-quadrado com n graus de liberdade são:
  - Cdf.chisq (x,n) para a função de distribuição no ponto x, *F(x)*;
  - Idf.chisq(*p*,n) para o quantil de ordem *p*, $\chi^2_{p,n}$.

Distribuição F (de Fisher)

- A distribuição de Fisher está associada à razão de duas variáveis de qui-quadrado e é útil quando queremos analisar a razão de duas variâncias amostrais. Como cada qui-quadrado tem associado um certo número de graus de liberdade, a distribuição de Fisher vai ter associados dois números de graus de liberdade. Quando X tem distribuição de Fisher com n e m graus de liberdade escreve-se $X \sim F_{n,m}$.

Distribuição de Fisher

- Tal como para as distribuições referidas anteriormente, para obter probabilidades e quantis da distribuição de Fisher podemos recorrer a software ou a tabelas.
- As tabelas fornecem habitualmente quantis da distribuição.
- No SPSS as funções associadas à distribuição de Fisher com n,m graus de liberdade são:
  - Cdf.F (x,n,m) para a função de distribuição no ponto x, *F(x)*;
  - Idf.F(*p*,n,m) para o quantil de ordem *p*, $F_{p,n,m}$.