

## 1. Métodos de Amostragem

### 1.1 Amostragem probabilística

Dizemos que estamos perante um processo de **amostragem probabilística** quando qualquer elemento (ou parte) da população tem iguais hipóteses (mesma probabilidade) de ser incluído na amostra.

Qualquer processo de amostragem deve, à partida, ser um processo de amostragem probabilística.

Porquê? Porque a amostra deve ser **representativa** da população; ou seja, deve ser um “espelho” da mesma, com todos os seus grupos representados na mesma proporção que aí existem. Os métodos de amostragem probabilísticos são os que melhor evitam o **enviesamento** de selecção (**viés**, **bias**, erro sistemático); ou seja, afastam qualquer tendência sistemática para sub-representar e/ou sobre-representar na amostra alguns elementos da população, o que poderia falsear os resultados e impedir uma correcta e rigorosa inferência sobre a população. Resumindo, estes métodos são os que melhor afastam, na medida do que é possível, erros de origem sistemática (ver secções subsequentes).

*Exemplo:*      *População:*      *Opiniões dos telespectadores sobre um filme erótico exibido num canal aberto de televisão.*

*Amostra:*            *Opiniões expressas por telespectadores que telefonaram para a estação emissora.*

*Enviesamento de selecção:*      *Sobrerrepresentação das opiniões de telespectadores integrados em grupos de pressão auto denominados “defensores dos bons costumes”.*

*Este é um processo de amostragem que não é probabilístico !!*

*Exemplo:*      *População:*      *Opiniões dos alunos de certa Escola Privada sobre o aumento de propinas*

*Amostra:*            *Opiniões expressas por 50 alunos seleccionados pelo computador entre todos os inscritos.*

*Este é um processo de amostragem que é probabilístico !!*

Entre os processos de amostragem probabilística, os mais importantes são:

- **Amostragem aleatória** – “sorteio” entre todos os elementos da população, sem distinções nem excepções.

*Exemplo:*      *Num processo de controlo de segurança, seleccionam-se aleatoriamente 3 passageiros de um voo para que as suas bagagens sejam meticulosamente fiscalizadas.*

- **Amostragem estratificada** – só pode ser aplicada quando há um conhecimento perfeito da estratificação da população em relação a certo(s) factor(es) – a amostra é dividida em sub-amostras, com dimensões proporcionais ao estrato correspondente na população ; assim mantém-se a igual probabilidade de qualquer elemento da população ser incluído na amostra. No entanto, a amostra obtida terá uma distribuição semelhante à da população em relação ao factor conhecido.

*Exemplos:*

- O modo como são feitas as sondagens antes de uma eleição política.
- A recolha de solo ao longo num terreno, de acordo com a distribuição percentual dos vários tipos diferentes (zona cultivada, zona inerte, zona rochosa, perto do rio, zona mais seca, etc).

## 1.2 Amostragem não probabilística

Num processo de **amostragem não probabilística** não nos é permitido definir com rigor as probabilidades de inclusão dos diferentes elementos da população na amostra.

Estes processos conduzem normalmente a enviesamentos (erros sistemáticos). No entanto, em certos casos, esses enviesamentos poderão não ser importantes face às vantagens que estes métodos oferecem,

Os processos de amostragem não probabilística podem ser usados quando não haja preocupações de rigor na representatividade da amostra (por exemplo, na realização de estudos piloto ou de análises preliminares).

Quase sempre, estes processos são quase sempre mais expeditos e mais económicos do que os processos de amostragem probabilística.

Alguns processos de amostragem não probabilística comuns são:

- **Amostragem por conveniência** – a conveniência para o analista é o critério que preside à escolha dos elementos da amostra

*Exemplo:* Um teste de aceitabilidade de uma nova cerveja pelos consumidores, efectuado com base nas opiniões recolhidas entre um conjunto de empregados da empresa cervejeira.

- **Amostragem subjectiva** – os elementos da amostra são seleccionados pelo analista, com base num critério pessoal, eminentemente subjectivo, de representatividade.

*Exemplo:* Análise da clareza de um texto didáctico, com base na opinião de um conjunto de alunos que o autor julga serem típicos da população estudantil à qual se dirige.

## 2. Processos de recolha de dados

**Classificação dos dados no que respeita à sua fonte:**

- **Dados primários** – se forem obtidos directamente pelo analista ou pela sua organização. Os dois processos mais frequentemente utilizados na aquisição de dados primários são a observação (ou análise) e o questionário.

- **Dados secundários** – se forem compilados ou publicados por outra organização (tal como agências governamentais, associações empresariais ou sindicatos, empresas especializadas em estudos de mercado).

**Classificação dos dados no que respeita ao seu processo de recolha**

- **Processos experimentais** – quando é exercido um controlo directo sobre os factores que potencialmente afectam a característica ou o conjunto de características em análise.

*Exemplo:*      *No estudo da acção de um novo fármaco, este é ministrado a 10 doentes, enquanto que um placebo é ministrado a outros 10 doentes – grupo de controlo.*

- **Processos observacionais** – quando não há qualquer controlo dos factores que potencialmente afectam a(s) característica(s) em estudo, apenas há observação dos efeitos dos potenciais factores já existentes .

*Exemplos:*      *Um estudo da temperatura ambiente.*  
                     *Registo da evolução de cotações na Bolsa.*

### 3. Erros de amostragem

#### 3.1 Classificação das causas de erros de amostragem

Há dois tipos de causas principais de erros, que podem ser cometidos na selecção de uma amostra:

- **Causas aleatórias** – Estão relacionadas, quer com a selecção aleatória da amostra, quer com a aleatoriedade própria dos métodos / equipamentos / operador humano. Contribuem para aumentar a imprecisão dos resultados (ou seja, a sua incerteza), mas não influenciam a sua exactidão.

O efeito destas causas diminui com o aumento do tamanho da amostra. Na prática, o menor efeito possível consegue-se através do estudo de toda a população sem excepção de qualquer dos seus elementos. No entanto é muito complicado (para não dizer impossível) eliminar completamente as componentes aleatórias devidas a métodos / equipamentos / operador humano.

Apesar de tudo, quase sempre não é compensatório aumentar demasiadamente o número de elementos a estudar, dado o aumento do custo e do esforço necessário, face ao aumento do rigor dos resultados obtidos, que a partir de certa dimensão deixa de ter importância. É possível determinar-se uma dimensão amostral razoável de acordo com os recursos disponíveis e a precisão desejada nos resultados (capítulo 7).

- **Causas sistemáticas** – Também chamados de **enviesamentos**; ocorrem por aplicação de métodos de selecção tendenciosos ou então por descalibração do equipamento, erros sistemáticos do operador ou do método aplicado, etc. Enquanto que em muitos casos assumem uma importância demasiadamente relevante para serem ignoradas, vários casos há nos quais a sua importância é menor, sobretudo quando pensamos no esforço e custos que implicaria um método de selecção mais exigente.

### 3.2 Algumas causas sistemáticas de erros de amostragem

Por vezes importantes, por vezes desprezáveis, as causas sistemáticas de erro são sempre de ter em atenção, antes, durante e após a recolha de uma amostra.

Algumas causas sistemáticas mais comuns:

- De admissão ou de amostragem –** Estas causas ocorrem quando a amostra é seleccionada a partir de uma população especial, não representativa da realidade ou quando a amostragem é realizada por conveniência.

*Exemplo – Num inquérito realizado num aeroporto a utentes seleccionados casualmente pelo colaborador, este evita seleccionar famílias com filhos pequenos, pois é muito complicado manter as crianças sossegadas enquanto os pais respondem.*

*Exemplo – Determinação do teor médio de acidez dos frutos da espécie limoeiro, a partir de alguns limões seleccionados na zona Transmontana.*

*Exemplo - Determinação do teor médio de colesterol numa população, através dos resultados das análises clínicas de um Laboratório, durante uma semana – subrepresentação de pessoas sem problemas de saúde evidentes que não fazem exames periódicos.*

- De auto-selecção –** Ocorrem sobretudo quando a selecção dos elementos a incluir na amostra ocorre por auto-motivação ou voluntariado. Os elementos que integram na amostra são de algum modo tendenciosos na sua resposta, de acordo com o motivo pessoal que os levou à participação.

*Exemplo – Na fase experimental da utilização de um novo método de detecção de cancro da mama, voluntariam-se mais mulheres com história familiar desta doença, logo à partida com superior probabilidade de o contraírem.*

*Exemplo – Pretende-se prever a receptividade do público a um novo vinho destinado a conquistar mercado entre a população que habitualmente não consome. Para o teste realizado num centro comercial, voluntariam-se mais pessoas que são já à partida apreciadoras de vinho.*

- De não participação –** Ocorrem quando elementos da amostra se recusam a responder ou a participar no estudo.

*Exemplo – Inquérito presencial sobre práticas sexuais. Várias pessoas recusar-se-ão a participar.*

- De migração –** Quando os elementos de um grupo o abandonam e vão para outro grupo de estudo ou então saem do estudo.

*Exemplo – Durante o estudo comparativo da qualidade de vida dos funcionários de duas companhias aéreas rivais, alguns desses funcionários abandonam a sua companhia e vão trabalhar na outra.*

- De ausência de respostas –** Ocorrem quando não se consegue contactar/tratar todos os elementos da amostra.

*Exemplo – Um computador gera aleatoriamente 500 números de telefone fixo para a realização de um inquérito, mas alguns dos números sistematicamente não atendem a chamada.*

- **De resposta** – quando o elemento da amostra é induzido em erro, por exemplo por interpretação errónea de uma pergunta mal estruturada num questionário, ou por falta de atenção a regionalismos.

- **De tempo ganho** – Surgem quando os elementos que se submetem ao teste de detectam problemas mais precocemente e adquirem, portanto, mais hipóteses de superar o problema.

*Exemplo – Uma criança que participa num teste a um novo método de acompanhamento escolar terá os seus problemas detectados mais cedo e, por esse motivo, mais hipóteses de os tornar.*

*Exemplo - O criador cujos animais sejam sujeitos ao rastreio de determinada doença verá os seus problemas diagnosticados mais cedo e, por esse motivo, terá maiores hipóteses, tanto de recuperar os animais doentes, como de evitar a contaminação dos seus animais saudáveis.*

- **De não controlo de outras variáveis** – Quando não são controladas variáveis que de algum modo directo ou indirecto podem interferir na medição da grandeza que está a ser analisada.

*Exemplo – Uma companhia chega à conclusão de que tem em média mais clientes do que a sua rival, baseando o estudo próprio no mês de Agosto e o da outra companhia no mês de Fevereiro.*

- **De descalibração do equipamento de medição / análise**

*Etc.*

## 4. Qualidade das amostras

### 4.1 Exatidão e precisão

Os erros que se cometem em qualquer determinação classificam-se em duas categorias principais:

- **Erros aleatórios** – Ocorrem de forma aleatória; ao longo de muitas medições os erros negativos equilibram os erros positivos. São causados, por exemplo, a variação de critério do analista ou a oscilações internas do aparelho de medida

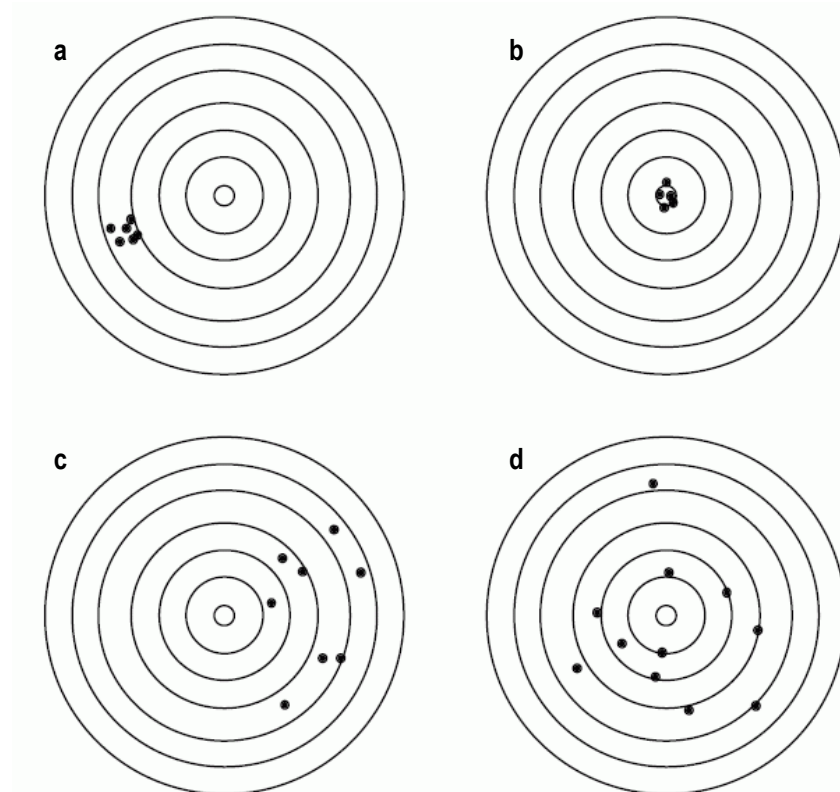
- **Erros sistemáticos** – Ocorrem de forma sistemática, ou seja, em todas as medições e sempre no mesmo sentido. São causados, por exemplo, pela descalibração do aparelho de medida, métodos errados ou incorretamente aplicados ou ainda técnica deficiente do analista.

As várias causas acuem de uma forma independente entre si e são acumuláveis. Isto é uma medida pode estar afetada de erros sistemático e/ou erros aleatório.

- **Exatidão** de uma medida ou um resultado é a sua capacidade de “acertar” em média no verdadeiro valor. Os erros sistemáticos contribuem para a falta de exatidão de uma medida.

- **Precisão** de uma medida ou um resultado é a variabilidade obtida em determinações sucessivas. A precisão de uma medida está relacionada com a quantidade de erros aleatórios.

*(Ver figura 2.1)*



**Figura 2.1** – Exatidão e precisão de vários atiradores ao alvo.

- a) Preciso. Pouco exato.      b) Preciso. Exato  
c) Pouco preciso. Pouco exato.      d) Pouco preciso. Exato.

Exatidão e precisão são duas características distintas, ambas desejáveis em quaisquer medidas ou resultados. A recolha das amostras deve ser realizada o mais corretamente possível de forma a maximizar estas características nos resultados obtidos.

*Exemplo –* Vamos admitir que é absolutamente verdade que o volume de 1 mol de um gás, considerado ideal, a 25°C e sujeito à pressão de uma atmosfera, é de 22,4 dm<sup>3</sup>. Qualquer método que determine que aquele volume é por exemplo, 23,1 ± 0,1 dm<sup>3</sup> é um método bastante preciso, mas pouco exato. Por outro lado, qualquer método que conclua que o volume referido é de, por exemplo, 22,5 ± 0,5 dm<sup>3</sup> é bastante mais exato, mas pouco preciso.

## 4.2 Conceito de Incerteza

O **erro** de uma medida foi definido com a diferença entre um resultado individual e o verdadeiro valor da grandeza medida. O verdadeiro erro não pode ser conhecido, pelo que o erro de uma medida também não pode ser conhecido. Mesmo que tenhamos consciência da existência do erro e utilizemos um método mais fiável para seu controlo, estaremos sempre sujeitos ao erro do próprio método de controlo.

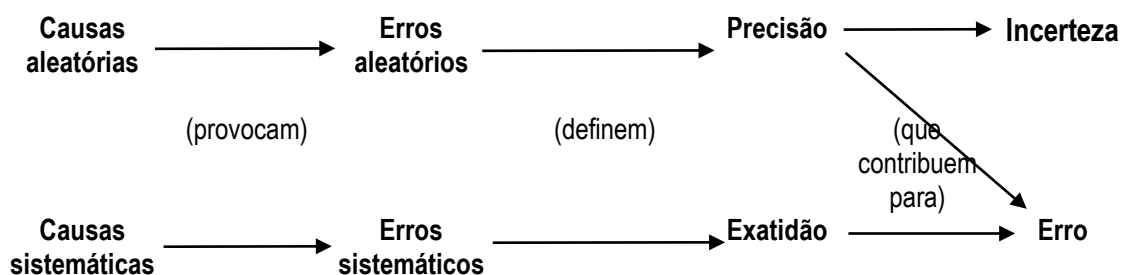
Por outro lado, a **incerteza** pode normalmente ser quantificada e define um intervalo de valores dentro do qual se estima poder encontrar o verdadeiro valor da grandeza medida. A incerteza pode ser determinada por um processo estatístico/matemático tendo por base um conjunto muito grande de valores medidos.

A incerteza contabiliza apenas os erros aleatórios, pelo que devemos dizer que o erro é uma medida de exatidão + precisão da medida enquanto que a incerteza apenas é uma medida da precisão da medida, nada nos informando acerca da sua exatidão.

A incerteza é um conceito fundamental em inferência estatística.

### Resumo de alguns conceitos:

Resume-se em seguida, de forma esquemática, a ligação e as implicações entre os conceitos expostos:



- FIM -