

ESTATÍSTICA II

Apresentação da UC. Revisão de Conceitos

Teoria da Probabilidade e Inferência Estatística

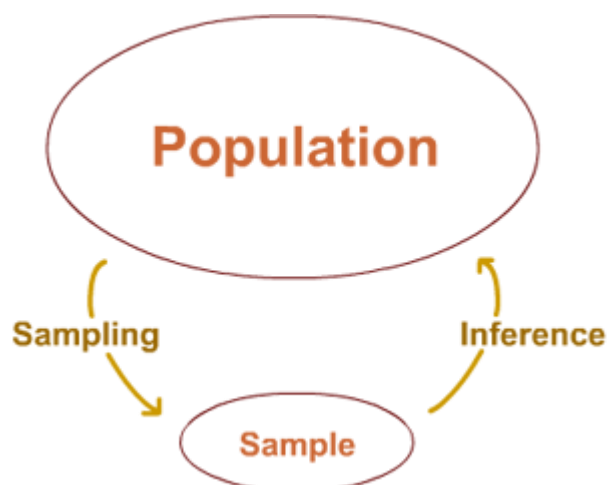
No campo da **Teoria da Probabilidade**, parte-se de determinado esquema ou modelo e calcula-se a probabilidade de certos resultados ou acontecimentos.

No campo da **Inferência Estatística**, parte-se de dados ou observações (amostra) e procura inferir-se algo sobre o esquema ou modelo.

Exemplo: Considere-se os alunos matriculados no ISEG entre os quais há uma proporção θ que tem conta no Facebook. Escolhe-se ao acaso e com reposição $n=20$ pessoas. Se θ fosse conhecido, seja $\theta=0.4$, podia haver interesse em calcular a probabilidade de encontrar x utilizadores do Facebook ($0 \leq x \leq \theta$) nesse grupo de 20 alunos. Trata-se de um problema de probabilidades.

Conceitos Fundamentais de Amostragem

- População e amostra
- População objetivo e população inquirida



Source: https://www.google.pt/search?q=sampling&rlz=1C1CHZL_pt-PTPT734PT734&source=Inms&tbn=isch&sa=X&ved=0ahUKEwiYrJ6f3eTcAhXlzYUKHTqdA3kQ_AUICigB&biw=1248&bih=769#imgrc=vfu1QP-CePmOwM:



Source:

https://www.google.pt/search?q=sampling&tbm=isch&tbs=rimg:CfNNGHfAvO68JiiZz3JVkb3bRRxUT9933IAZ0cgE3AbfxLhwBeFuYIVFNI5hSCIUM4YNI4cppd7ozFAEZXNEfagvJCoSCZnPclWRvdtFed8wwhw2XccCKhIJHFRP33feUBkRFOygn13pziYqEgnRyATcBt_1EuBgb7hrAokheRioSCXAF4W5iVUU2ERNH8AKq1kQ-KhIJXmFllhQzhgORIAK5EYwE1PcqEgmLhym13ujMUBFrE0AJ5VVJ9SoSCQRk1cR9qDIKEYS-HFZen6H&tbo=u&sa=X&ved=2ahUKewj79OX23eTcAhVPyxOKHWFwDBsQ9C96BAGBEBs&biw=1248&bih=769&dpr=2.2#imgrc=802Ad8C87rxcOM

Basic Statistics

Population/Parameter vs. Sample/Statistics

Parameter	Name	Statistics
μ	Mean	\bar{X}
σ^2	Variance	s^2
σ	Standard Deviation	s

* Sample statistics are used to estimate parameters of a population

Example of Population and Sample

The labor-management cooperation team conducted a job satisfaction survey with 2,000 of the entire employees, 500 from each part of the business.

- Population and parameter:
- Sample and statistics:

Source: https://www.google.pt/search?q=population+vs+sample&rlz=1C1CHZL_pt-PTPT734PT734&source=lnms&tbm=isch&sa=X&ved=0ahUKewjm_pqC2-TcAhVFLBoKHQ2WACwQ_AUICigB&biw=1248&bih=769#imgrc=94gYoz-Erv9_UM

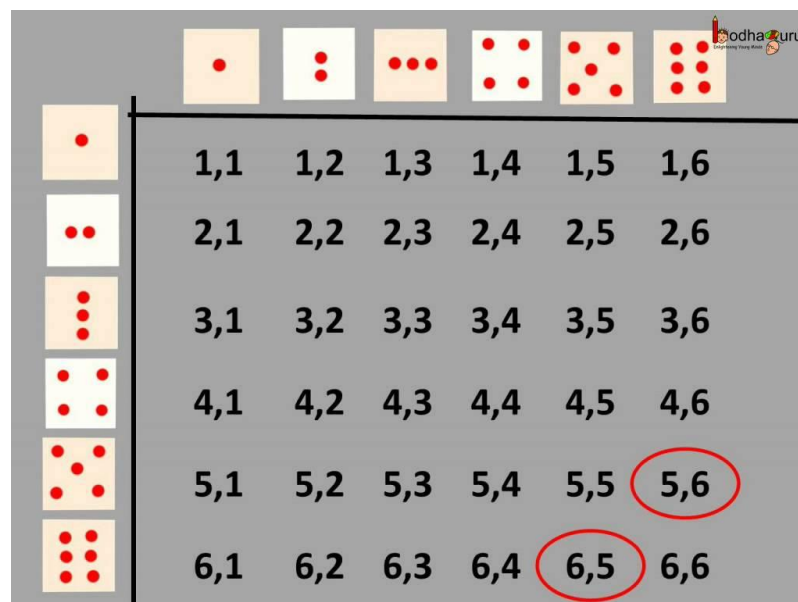
- **Amostragem casual**

Suponha que (x_1, x_2, \dots, x_n) é uma amostra de n observações da população a que corresponde um vector aleatório (X_1, X_2, \dots, X_n) . Quando as n variáveis aleatórias observadas são independentes e identicamente distribuídas (iid), diz-se que se tratar de uma amostra casual.

- **Inferência estatística paramétrica e não paramétrica**

- **Espaço-amostra**

O espaço-amostra é o conjunto de todas as amostras possíveis de serem selecionadas de uma população.



	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

Source: https://www.google.pt/search?rlz=1C1CHZL_pt-PTPT734PT734&tbm=isch&q=%22sample+space%22&chips=q:sample+space,online_chips:events&sa=X&ved=0ahUKewi1zfvgyeXcAhWSxIUkHVFCB0Q4IYINSgO&biw=1248&bih=769&dp_r=2.2#imgrc=1R3o9qyq-Q9pTM:

- **Espaço-parâmetro**

- **Estatística**

Uma estatística é uma variável aleatória ou vector aleatório, $T = T(X_1, X_2, \dots, X_n)$, função da amostra aleatória (X_1, X_2, \dots, X_n) , que não envolve qualquer parâmetro desconhecido.

Parameter name	Population parameter symbol	Sample statistic
Number of cases	N	n
Mean	μ (mu)	\bar{x} (Sample mean)
Proportion	π (Pi)	P (Sample proportion)
Variance	σ^2 (Sigma-square)	s^2 (Sample variance)
Standard deviation	σ (Sigma)	s (sample standard deviation)
Correlation	ρ (rho)	r (Sample correlation)
Regression Coefficient	β (beta)	b (sample regression coefficient)

Source: https://www.google.pt/search?q=sample+statistics&rlz=1C1CHZL_pt-PTPT734PT734&source=lnms&tbn=isch&sa=X&ved=0ahUKEwj8zMKy-XcAhUMaBoKHS8VA4UQ_AUICigB&biw=1248&bih=769#imgrc=lfzV6rHlmpBqYM

- **Distribuições por amostragem**

As distribuições das estatísticas designam-se por **distribuições por amostragem** ou **distribuições amostrais**.

- **Distribuições mais utilizadas**

- **População Normal: Distribuição da média**

$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$
---	---

- **População Normal: Distribuição da variância**

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

- **População Normal: Distribuição da diferença de médias**

$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(\nu)$
$T = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	onde ν é o maior inteiro contido em r , $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{S_2'^2}{n}\right)^2}$

○ **População Normal: Distribuição da relação entre duas variâncias**

$$\frac{S_1'^2}{S_2'^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$$

○ **Grandes amostras: Distribuições por amostragem assintóticas (TLC)**

Média	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \mu}{S'/\sqrt{n}} \stackrel{a}{\sim} N(0,1)$
Diferença de médias	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \stackrel{a}{\sim} N(0,1)$

○ **População de Bernoulli**

Proporção	$\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} N(0,1)$
Diferença de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \stackrel{a}{\sim} N(0,1)$	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{m} + \frac{\bar{X}_2(1-\bar{X}_2)}{n}}} \stackrel{a}{\sim} N(0,1)$
Igualdade de proporções	$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \hat{\theta}(1-\hat{\theta})}} \stackrel{a}{\sim} N(0,1)$ onde $\hat{\theta} = \frac{m\bar{X}_1 + n\bar{X}_2}{m+n}$	