ANÁLISE DE DADOS EM GRH

Aula 4:

‘O que posso concluir com base nos dados da minha amostra?’

Introdução à estatística inferencial

Docente: Daniela Craveiro
dcraveiro@iseg.ulisboa.pt

ANÁLISE DE DADOS EM GRH

No final desta aula,

@s alun@s deverão:

- Perceber a diferença entre Estatística Descritiva e Estatística Inferencial
- Perceber o que é uma amostra probabilística e que tipos de técnicas de amostragem existem
- Perceber quais as características de uma Distribuição Normal e o papel da Teoria do Limite Central enquanto fundamento da Estatística Inferencial
- Perceber o que é o Intervalo de Confiança, para que serve, e como é calculado
- Perceber o que são Testes de hipóteses e a relação com o método científico
- Produzir o Intervalo de Confiança de um Média

ANÁLISE DE DADOS EM GRH

- **Estatística Descritiva**
  - **Dá-nos as ferramentas para descrever dados de uma (ou mais variáveis) numa amostra**
    - Medidas de tendência central (médias, modas, etc.)
    - Distribuição de frequências (proporções, percentagens, etc.)
    - Medidas de dispersão (variância, desvio padrão, etc.)
  - **Dá-nos as ferramentas para descrever a relação entre variáveis dados de uma (ou mais variáveis) numa amostra**
    - Medidas de Associação e Correlação

- **Estatística Inferencial**
  - Dá-nos as ferramentas para avaliarmos se a forma como os dados estão distribuídos, ou se a relação entre variáveis na amostra, podem ser inferidos para a população
    - **Intervalos de Confiança**
    - **Testes de Hipóteses**

- **A possibilidade de inferir de uma amostra para uma população depende de duas condições fundamentais:**

  I. **Que amostra seja probabilística**

  II. **Que haja uma forma de demonstrar que a distribuição da amostra segue uma distribuição normal**

Lembrar algumas definições

População

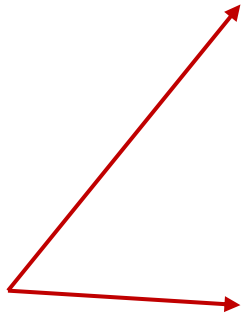
Conjunto de indivíduos, ou outras entidades, que pretendemos estudar.

Base de Amostragem

Lista de todas as unidades da população de interesse a partir da qual a amostra será extraída (ex. lista de números de telefone).

Amostra

Segmento da população de interesse que vai fazer parte do estudo.



Amostra Probabilística. Amostra em que cada elemento da população tem igual probabilidade de ser seleccionado, e é seleccionado independentemente dos outros.

Amostra Não-Probabilística. Amostra que não é escolhida segundo métodos probabilisticos.


TÉCNICAS DE AMOSTRAGEM
NÃO-PROBABILÍSTICAS
Conveniência
Por Quotas
Bola de Neve
PROBABILÍSTICAS
Aleatória Simples
Sistemática
Estratificada
Por clusters

Técnicas de Amostragem

NÃO-PROBABILITICAS	
Por Conveniência	Os membros da amostra são selecionados em função dos interesses do investigador e da facilidade de acesso aos entrevistados.
Por Quotas	Os membros da amostra são selecionados (a partir da base amostral) de modo a que a amostra possa possa reflectir a composição da população de interesse por referência a um conjunto de categorias (género, idade, etc.).
Bola de Neve	Selecciona-se um conjunto de inquiridos de forma aleatória, a quem é depois pedido que indique alguém na população de interesse que possa responder. (O processo de selecção de entrevistados pára quando a adição de novos entrevistas não adiciona mais dados de relevo.)

Técnicas de Amostragem

PROBABILITICAS	
Aleatória Simples	Os membros da amostra são selecionados de forma aleatória (ex. sorteio, Tabela de Números Aleatórios, data de nascimento, etc.) da **base amostral.**
Sistemática	Os membros da base amostral são ordenados de acordo com uma tabela de números aleatórios. É seleccionado, de forma aleatória, um membro da base amostral. Os restantes membros da amostra são escolhidos em função do seu número de identificação usando o seguinte critério (fracção da amostragem): Nº + i Em que i=N/n i, fração de amostragem N, total da população n, tamanho da amostra

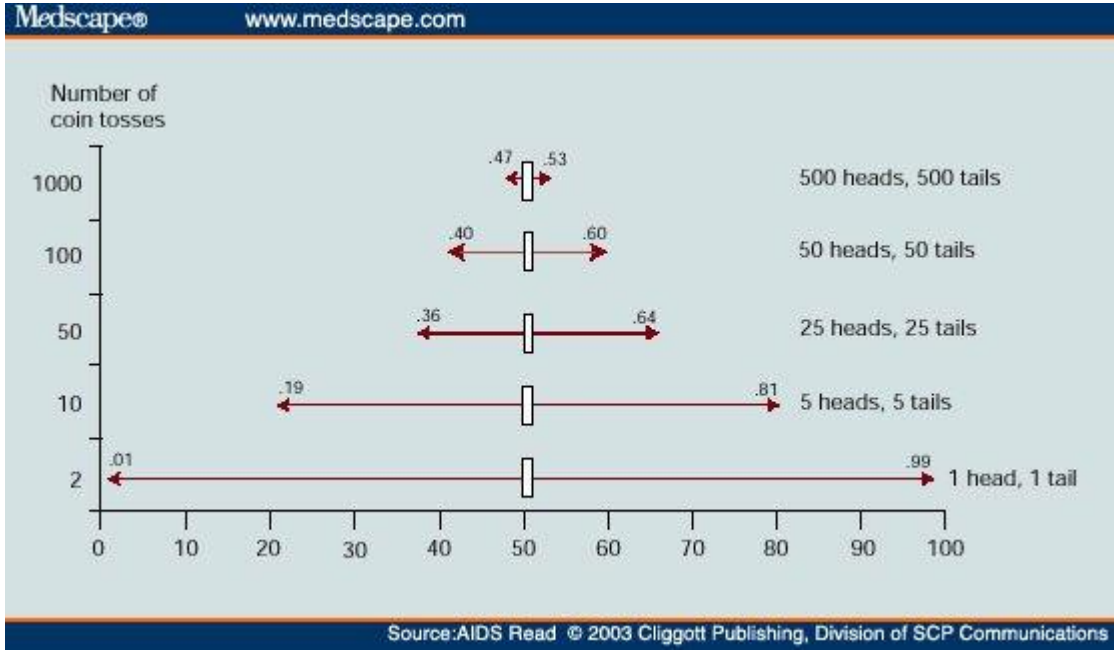
Técnicas de Amostragem

PROBABILITICAS	
Estratificada	Primeiro divide-se a **base amostral** num conjunto de sub-grupos (estratos), mutuamente exclusivos (um membro da população só pode pertencer a um estrato) e exaustivos (nenhum membro da população é omitido). Exemplos de categorias de estratificação: características demográficas, tipo de empresa, tipo de sector económico, etc. Os membros de cada estrato são depois seleccionados de forma aleatória.
Por Clusters	A **base de amostragem** é divida em clusters (Unidades Primárias de Amostragem), formados em função dos interesses do investigador. O investigador pode optar por incluir todos os clusters (Amostragem por Clusters em Um Passo), ou apenas uma fracção, que é seleccionada de forma aleatória (Amostragem por Clusters em Dois Passos). Dentro de cada cluster, seleciona-se de forma aleatória os membros (Unidades Secundárias de Amostragem) a incluir na amostra.

Por que é que o tamanho da amostra é importante?

- **Quanto maior for o tamanho da amostra, menor é a amplitude do intervalo de confiança – o que significa, maior precisão das nossas estimativas**

- **Quanto maior for o tamanho da amostra, maior será a 'potência estatística' do estudo, que mede a probabilidade de encontrar um efeito estatístico que existe na realidade (evitando Erros de Tipo II)**



Fonte: http://gosu.talentrank.co/confidence-interval-and-sample-size/

ANÁLISE DE DADOS EM GRH

Como se calcula o tamanho da amostra adequado que precisamos, se SABEMOS o tamanho da população?

$$Tamanho\ da\ Amostra: \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N}\right)}$$

Em que:

- N: **População (Total)**
- p: **Proporção da amostra (se desconhecida, assume-se 0.5)**
- z **: z-score (se Intervalo de Confiança a 95% = 1.96; se a 99% = 2.57)**
- e **: Margem de erro (se Intervalo de Confiança a 95% = 0.05; se a 99% = 0.01)**

Fonte: https://www.surveymonkey.com/mp/sample-size-calculator/

ANÁLISE DE DADOS EM GRH

- **Como se calcula o tamanho da amostra adequado que precisamos, se NÃO SABEMOS o tamanho da população?**

$$Tamanho\ da\ Amostra = \frac{z^2 \times p(1-p)}{e^2}$$

Em que:

P: **Desvio Padrão (um desvio-padrão de 50%, i.e. 0.5, é considerado um valor aceitável)**

Z **: z-score (se Intervalo de Confiança a 95% = 1.96; se a 99% = 2.57)**

e **: Margem de erro (se Intervalo de Confiança a 95% = 0.05; se a 99% = 0.01)**

Fonte: https://www.qualtrics.com/blog/calculating-sample-size/

- **A possibilidade de inferir de uma amostra para uma população depende de duas condições fundamentais:**

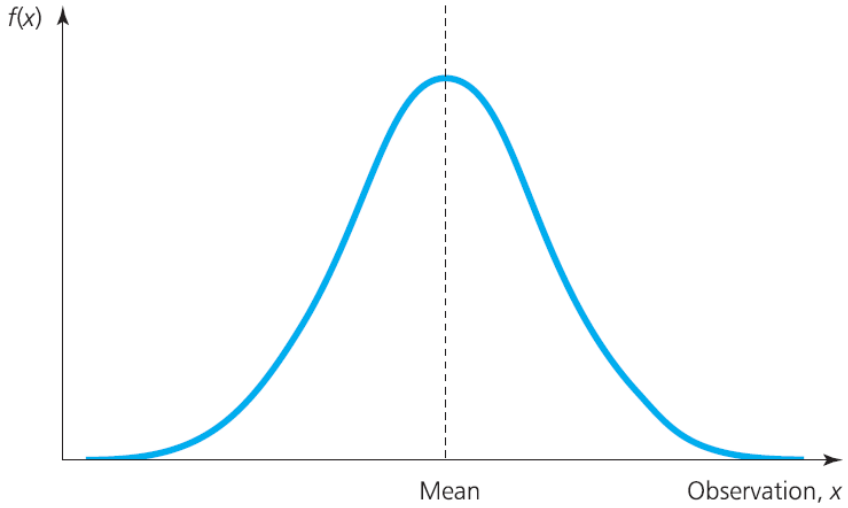
  I. **Que amostra seja probabilística**

  II. **Que haja uma forma de demonstrar que a distribuição da amostra segue uma distribuição normal**

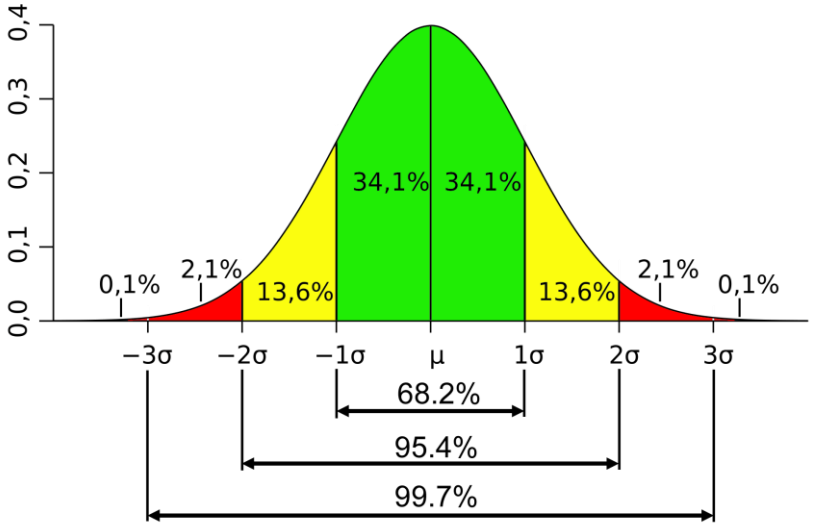
O QUE É UMA DISTRIBUIÇÃO NORMAL?

ANÁLISE DE DADOS EM GRH

- **O que é uma Distribuição Normal (ou Curva de Gauss)?**
  - **Média = Mediana = Moda**
  - **Simétrica**
  - **Distribuição segue a regra dos 3 Sigmas**
    - 34,1% das observações da variável estão dentro de um desvio-padrão da média
    - 68,2% das observações da variável estão dentro de (+ / -) um desvio-padrão da média
    - 95,4% das observações da variável estão dentro de (+ / -) 2 desvio-padrão da média
    - 99,7% das observações da variável estão dentro de (+ / -) 3 desvio-padrão da média



Fonte: Waters, 2011: 354



Fonte:http://en.wikipedia.org/wiki/File: Standard_deviation_diagram.svg

TODAS AS DISTRIBUIÇÕES SÃO NORMAIS?

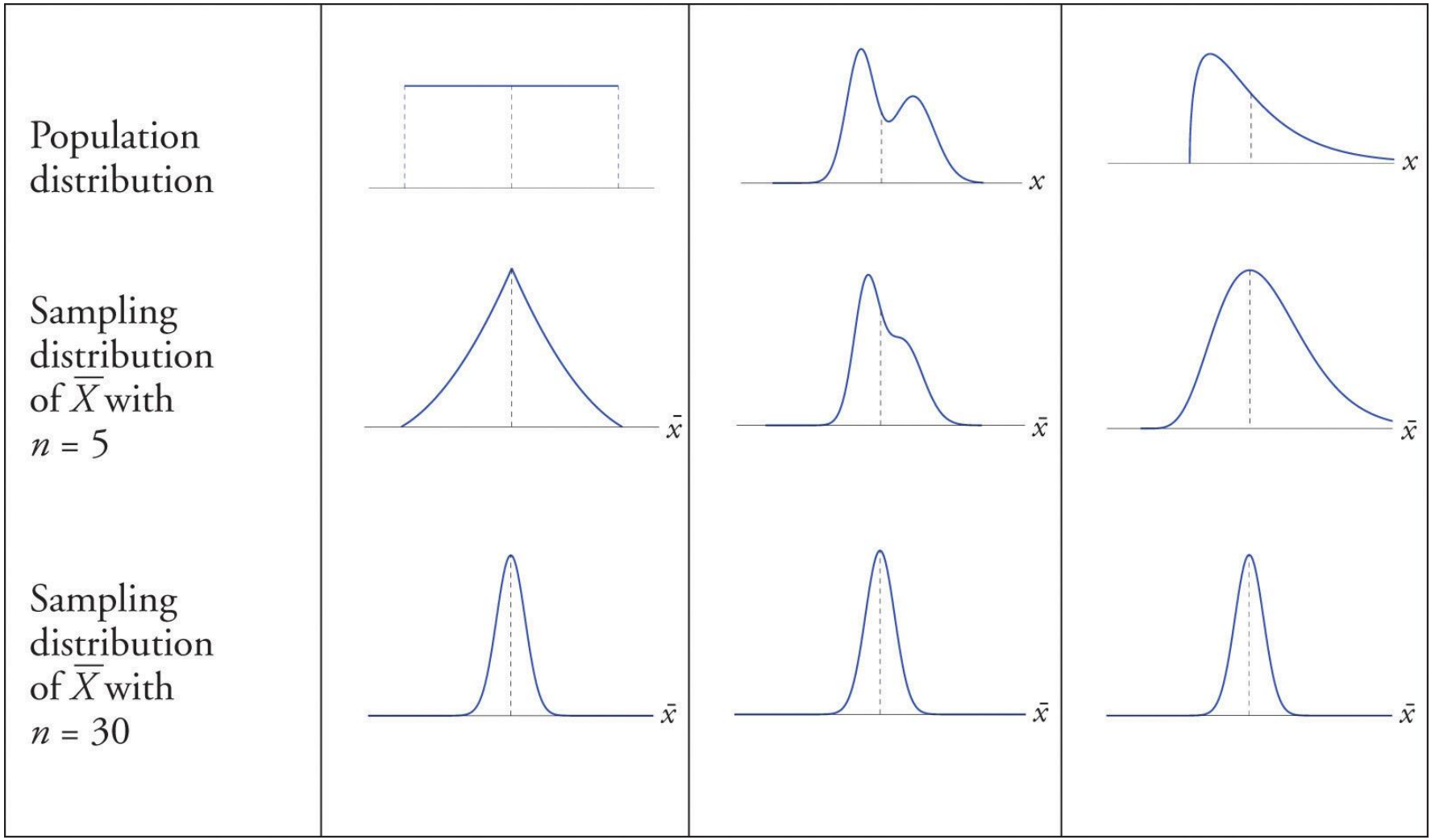
ANÁLISE DE DADOS EM GRH

NÃO.

(MUITAS SIM.)

Acontece que a distribuições das **médias das amostras** tendem para a normal, quando maiores que 30, mesmo que não tenham a forma normal.

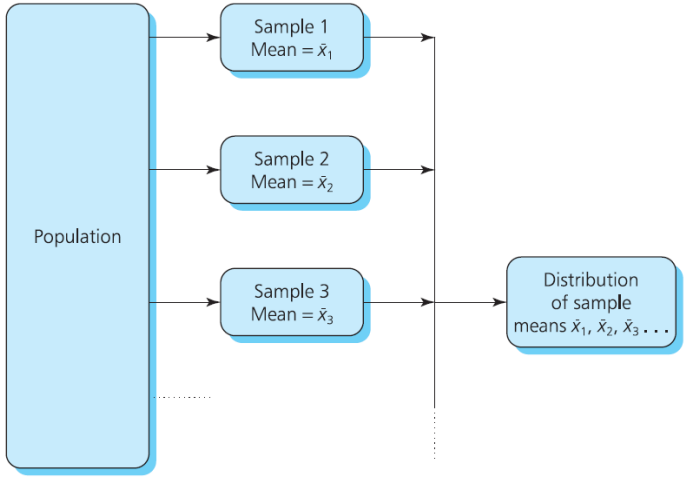
E assim**, porque sabemos coisas (matematicamente) sobre esta distribuição,** conseguimos calcular intervalos de confiança e gerar hipóteses estatísticas testáveis.

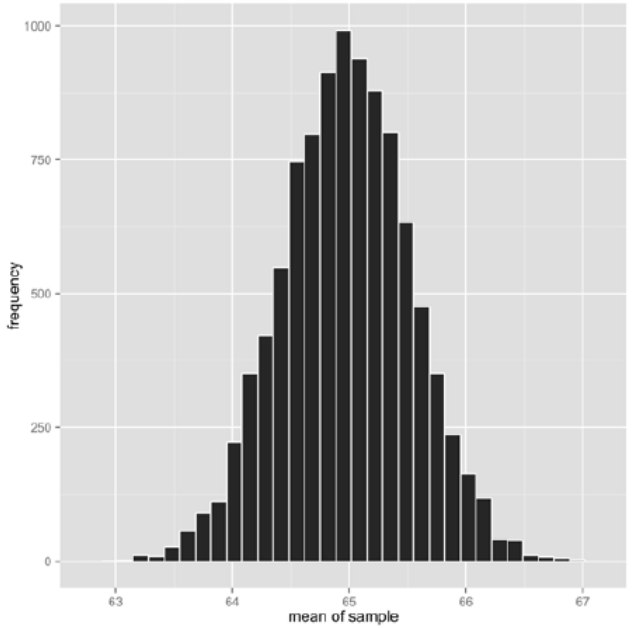


ANÁLISE DE DADOS EM GRH

Distribuição Amostral da Média

- Uma dada população pode dar origem a um número de amostras
- Cada amostra terá uma dada média (a chamada média amostral)
- À forma como se distribuem as médias destas amostras chamamos 'Distribuição Amostral Da Média'





O Teorema do Limite Central parte do um conjunto de propriedades da Distribuição Amostral da Média para fazer a inferência estatística de uma amostra para uma população.

- **O que diz o Teorema do Limite Central**

  - **Quando uma amostra é ≥ 30**

    - **a Distribuição Amostral da Média tende para a uma distribuição normal**
    - **O Desvio-Padrão da Distribuição Amostral da Média é o produto do seguinte rácio:**

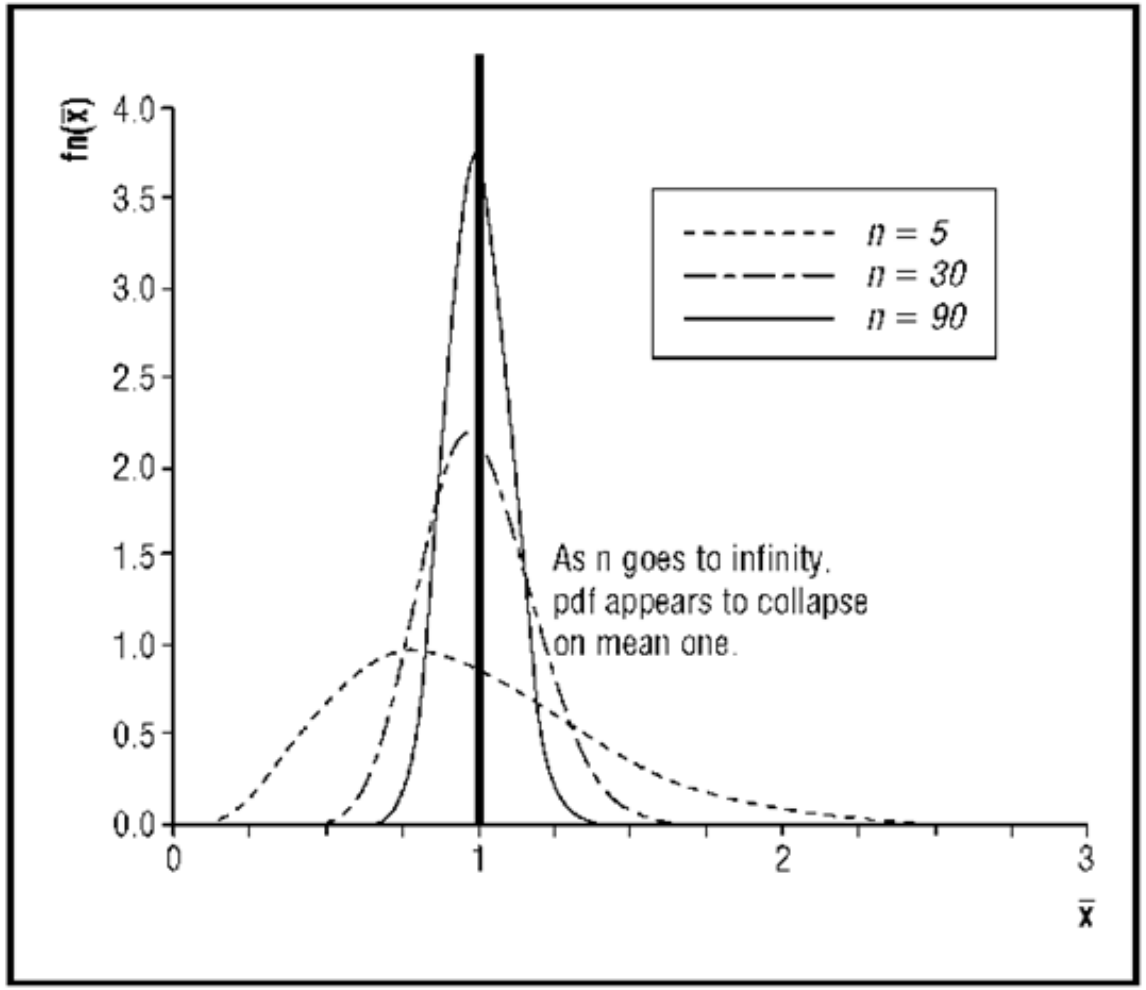
$$\frac{\sigma}{\sqrt{N}}$$

    - σ: **Desvio-Padrão da População**
    - $\sqrt{N}$: **Raiz quadrada do número de observações da amostra**

ANÁLISE DE DADOS EM GRH

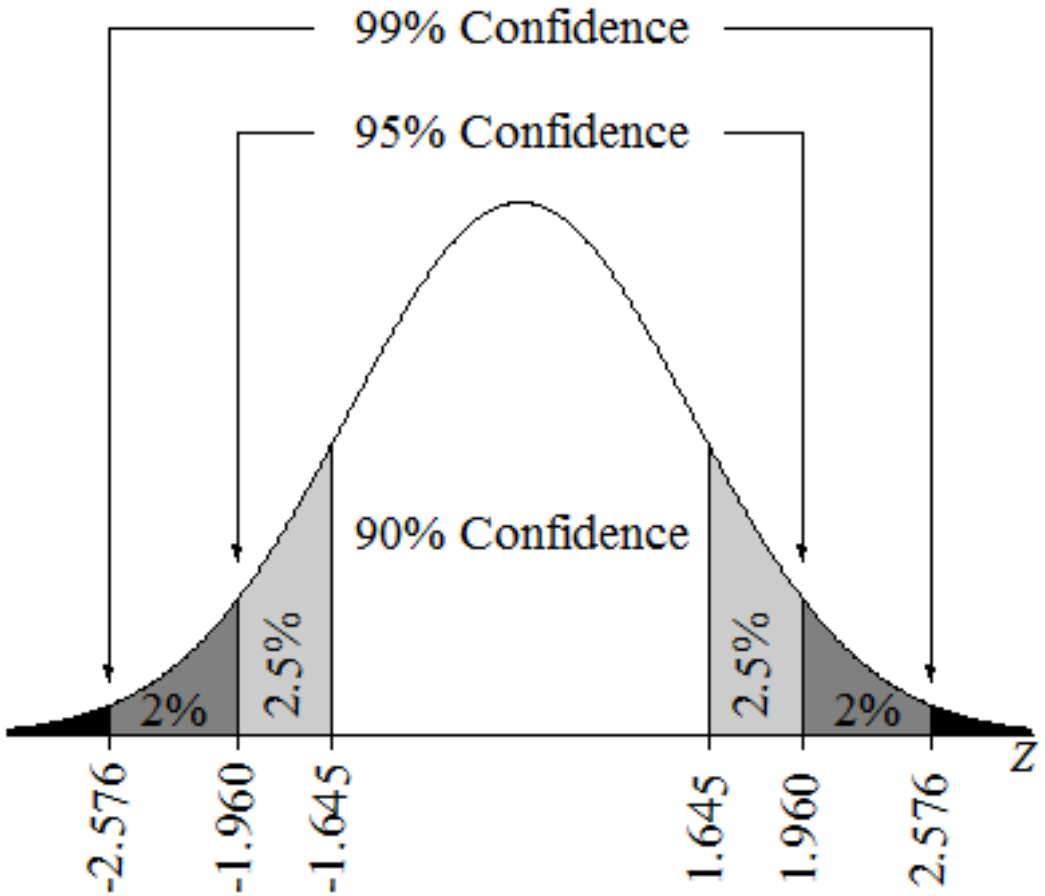
- **O que diz o Teorema do Limite Central**

  - **Quando uma amostra é ≥ 30, a média das médias amostrais tende para média populacional**
  - **A média da nossa amostra pode ser considerada uma aproximação da média da população**
  - **Quanto mais aumenta o tamanho da amostra, menor é o desvio-padrão da distribuição amostral da média, i.e. menor é a probabilidade de erro na amostra**
  - **Podemos com alguns elementos calcular intervalo de confiança das nossas inferências –** i.e: o intervalo de Valores dentro do qual se estima que a média se situe na população, com determinando **grau de confiança**



Fonte: http://what-when-how.com/social-sciences/law-of-large-numbers-social-science/

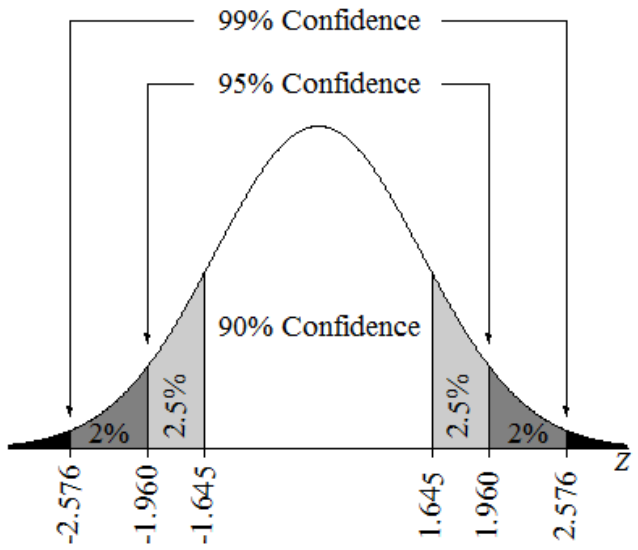
- **Grau de Confiança**
  - **Probabilidade de o intervalo de confiança capturar o parâmetro (neste caso a média) da população**
  - **Por norma** **adota-se um Grau de Confiança de 95%**
    - Se quisermos, podemos adotar um Grau de Confiança maior (99%)…
    - ou menor (90%)
  - **Interpretação:**
    - *Ex*: Intervalo de Confiança com um Grau de Confiança a 95%
    - Se fizéssemos 100 inquéritos, em 95% dos casos o intervalo de confiança iria conter a média da população



Fonte: https://courses.lumenlearning.com/wmopen-concepts-statistics/chapter/estimate-the-difference-between-population-proportions-2-of-3/

- **Grau de Confiança?**

De notar que, associado a um determinado Grau de Confiança, temos sempre um determinado valor crítico (*z*), baseado no Erro-Padrão



Confidence level	Z value
90%	1.65
95%	1.96
99%	2.58
99,9%	3.291

Fonte: http://www.biochemia-medica.com/en/journal/18/2/10.11613/BM.2008.015

Estes valores são usados para calcular a amplitude do Intervalo de Confiança

- **Como se calcula o Intervalo de Confiança?**

MARGEM DE ERRO

$$CI_x^{95} = \bar{x} \pm \{z^{95} \times SE\}$$

CI_x^{95} **Intervalo, com um Grau de Confiança a 95%**

$\bar{x}$ **Média da variável x**

z^{95} **Valor crítico para um Grau de Confiança a 95%**

SE **Erro-Padrão**

- **Como se calcula o Intervalo de Confiança de uma média?**

Isto é equivalente a

MARGEM DE ERRO

$$CI_x^{95} = \bar{x} \pm \left\{ z^{95} \times \frac{SD}{\sqrt{n}} \right\}$$

SD **Desvio-Padrão para um Grau de Confiança a 95%**

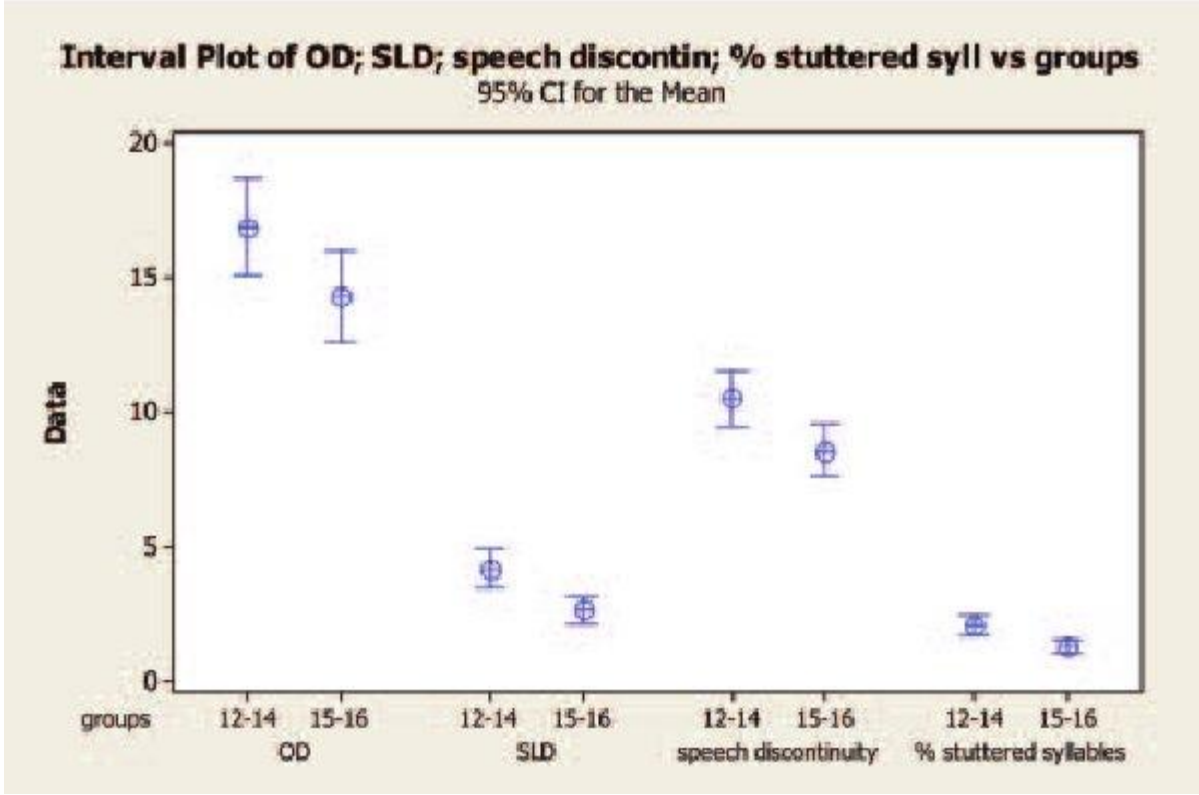
CI_x^{95} **Intervalo, com um Grau de Confiança a 95%**

$\bar{x}$ **Média da amostra**

z^{95} **Valor crítico para um Grau de Confiança a 95%**

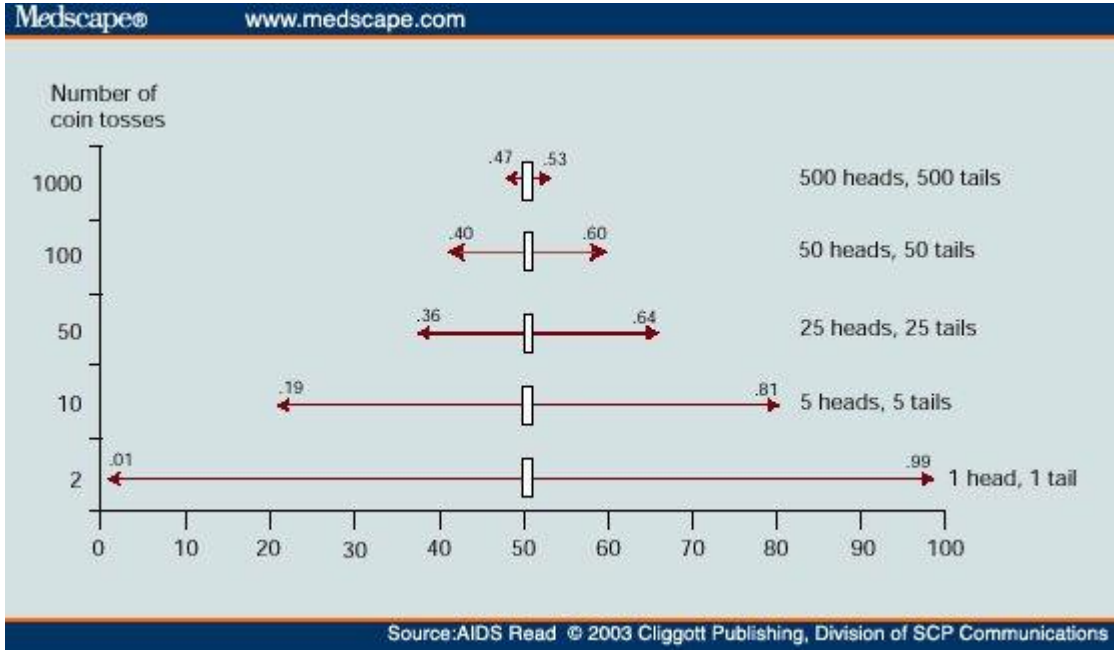
$\sqrt{n}$ **Raiz-quadrada do número de observações da amostra**

- **O que nos diz o Intervalo de Confiança?**
  - **Grau de precisão da média**
    - **Quanto maior a amplitude do Intervalo de Confiança, menor o grau de precisão**



Fonte:
https://www.researchgate.net/publication/5988752_Fluency_variation_in_adolescents/figures?lo=1

- **O que nos diz o Intervalo de Confiança?**
  - **Grau de precisão da média**
    - **Quanto maior a amplitude do Intervalo de Confiança, menor o grau de precisão**
    - **O que afeta a amplitude?**
      - **Quanto maior a amostra, menor é a amplitude do IC**
      - **Quanto maior é o Erro-Padrão, maior é a amplitude do IC**

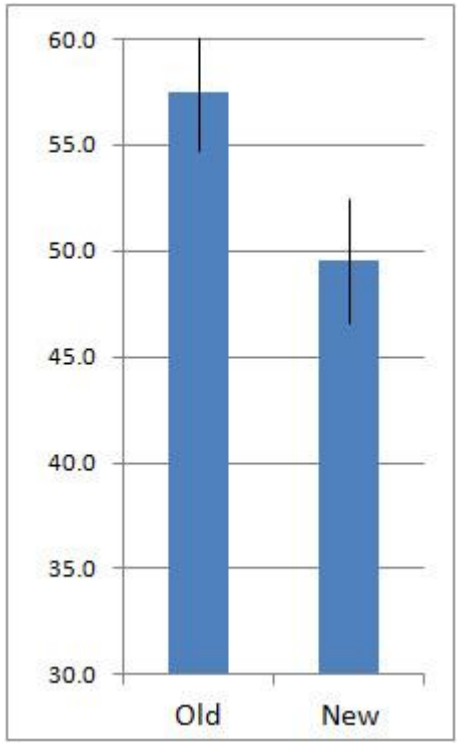


Fonte: http://gosu.talentrank.co/confidence-interval-and-sample-size/

ANÁLISE DE DADOS EM GRH

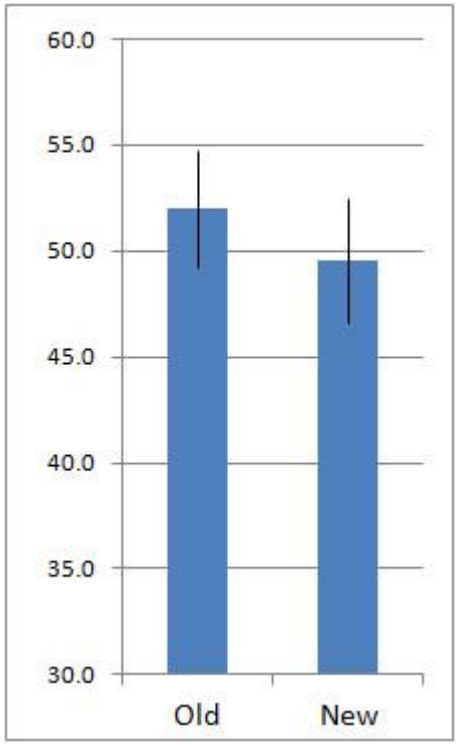
- **O que nos diz o Intervalo de Confiança?**
  - **Significância estatística**

Os intervalos de confiança não se sobrepõem



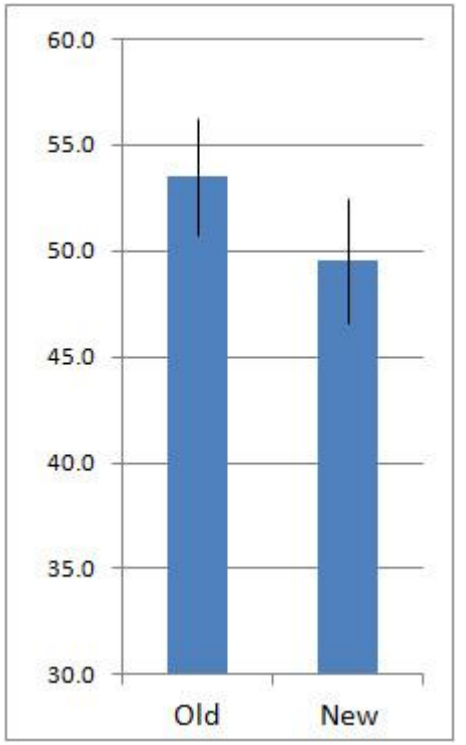
A diferença é estatisticamente significativa

Há uma grande sobreposição entre os intervalos de confiança



A diferença não é estatisticamente significativa

Há alguma sobreposição entre os intervalos de confiança



Não é claro. Para apurar devemos recorrer a um teste estatístico

Fonte: https://m[illegible]

ANÁLISE DE DADOS EM GRH

Temos analisado o caso da média

Mas podemos calcular o Intervalo de Confiança para uma série de estatísticas (proporções, medianas, coeficientes de beta, etc.)

ANÁLISE DE DADOS EM GRH

- **Ex: Intervalo de Confiança de uma proporção**

MARGEM DE ERRO

$$CI_p^{99} = \hat{p} \pm \left(z^{99} \times \sqrt{\frac{\hat{p} * (1 - \hat{p})}{n}} \right)$$

CI_x^{99}

Intervalo, com um Grau de Confiança a 99%

$\hat{p}$

Proporção da categoria, na amostra

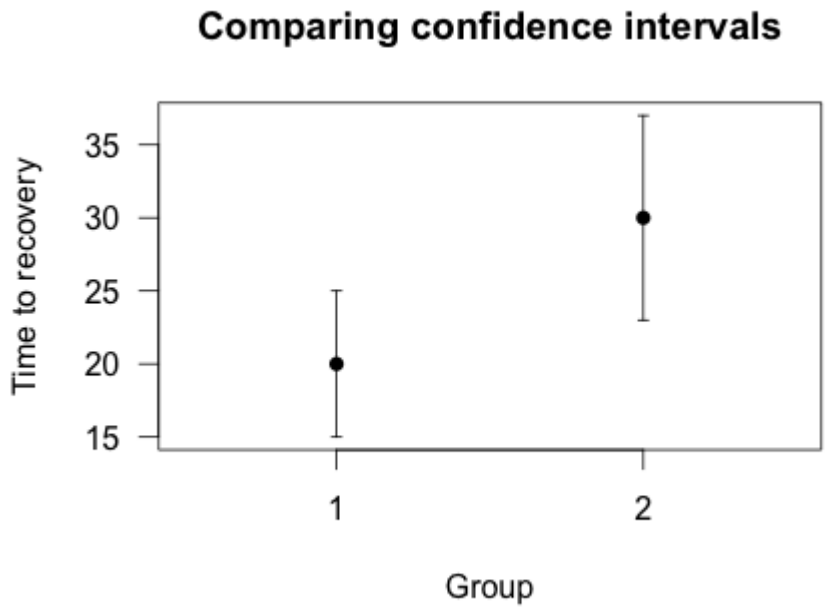
z^{95}

Valor crítico para um Grau de Confiança a 95%

n

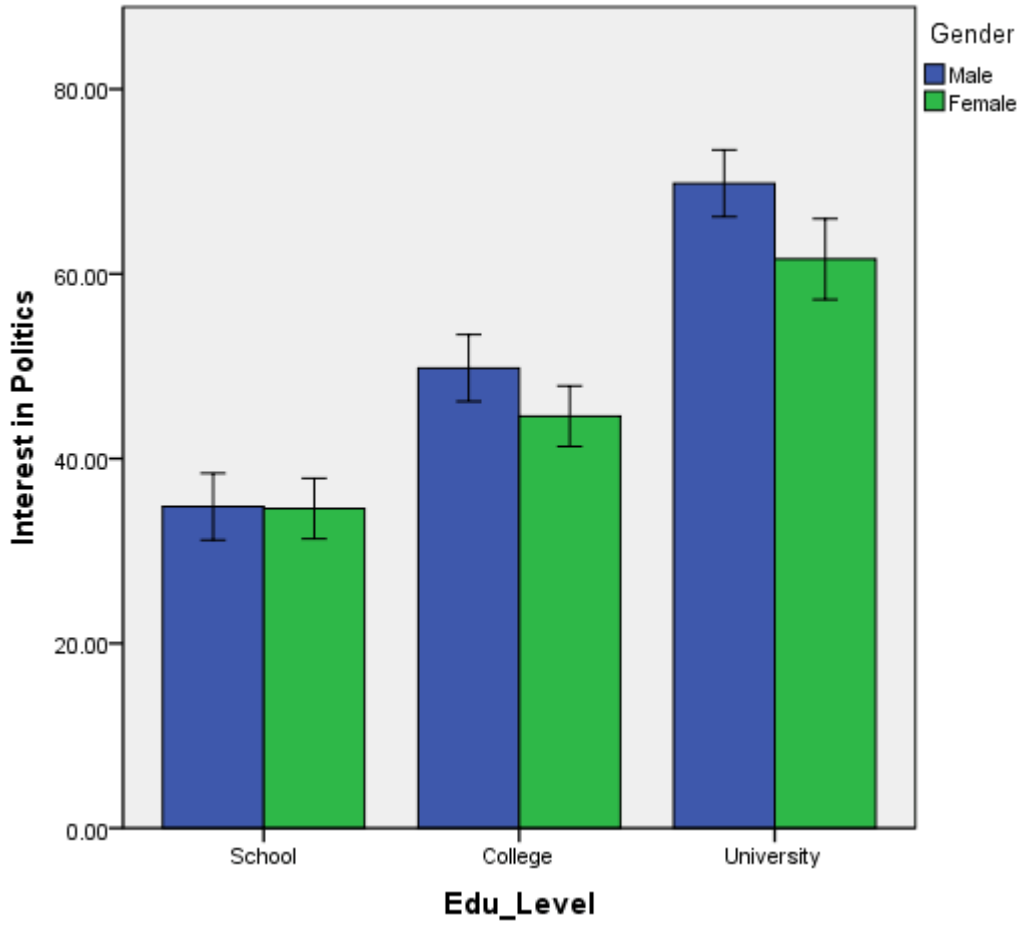
Número de observações da amostra

- **Como devemos representar graficamente um Intervalo de Confiança?**
  - **Se se tratar de uma média**
    - **Gráfico 'Alto-Baixo'**



Fonte: https://www.statisticsdonewrong.com/significant-differences.html

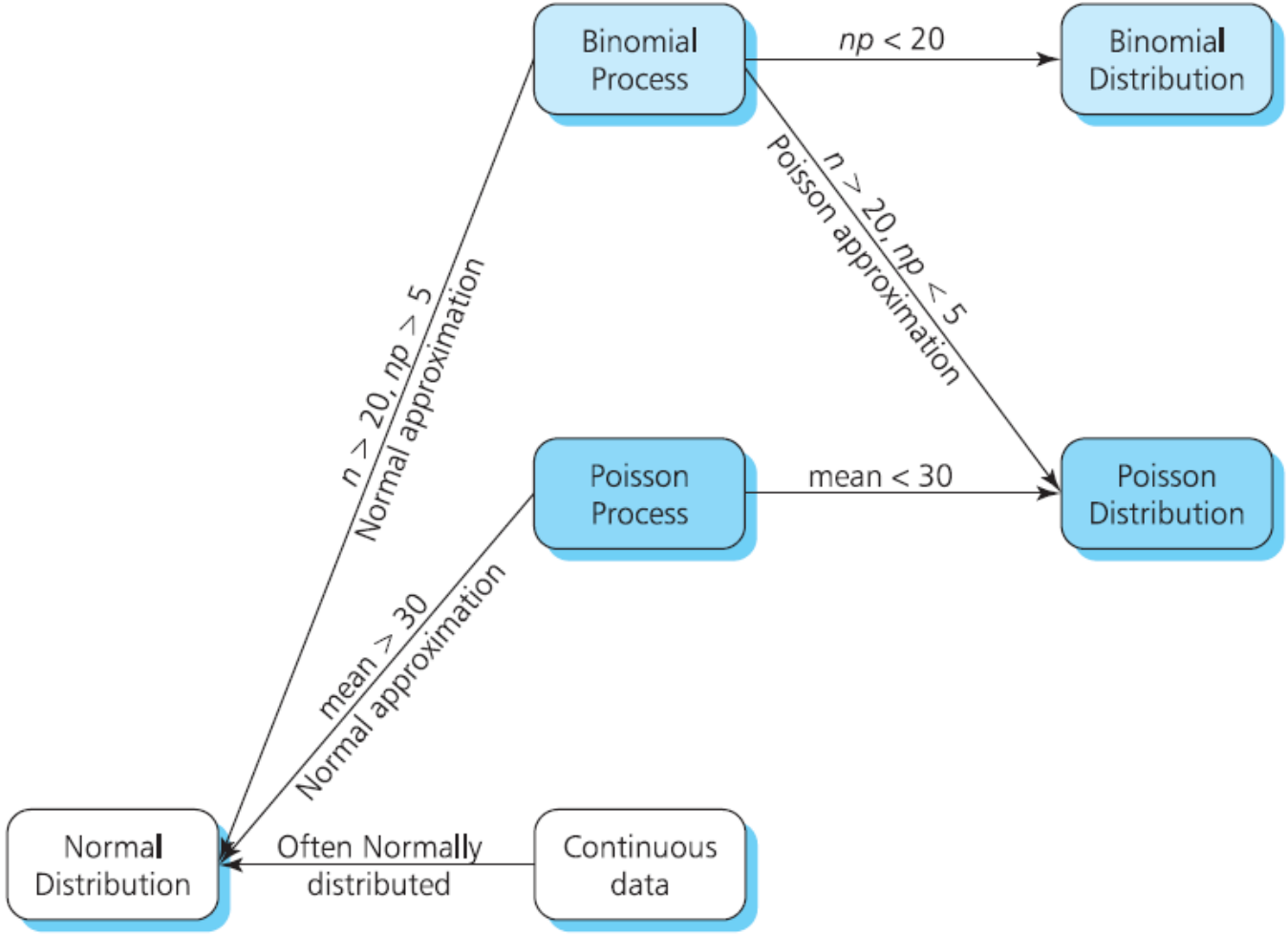
- **Como devemos representar graficamente um Intervalo de Confiança?**
  - **Se se tratar de uma média**
    - **Gráfico 'Alto-Baixo'**
  - **Se se tratar de uma proporção**
    - **Gráfico de Barras com Intervalo de Confiança**



Fonte: https://statistics.laerd.com/spss-tutorials/clustered-bar-chart-using-spss-statistics-2.php

ANÁLISE DE DADOS EM GRH

- **E se a população não segue uma distribuição normal?**
- **Em alguns casos, podemos fazer aproximações… isto é tratar certas distribuições como se tratassem de distribuições normais.**
- **Mas isso não é uma preocupação por agora!**



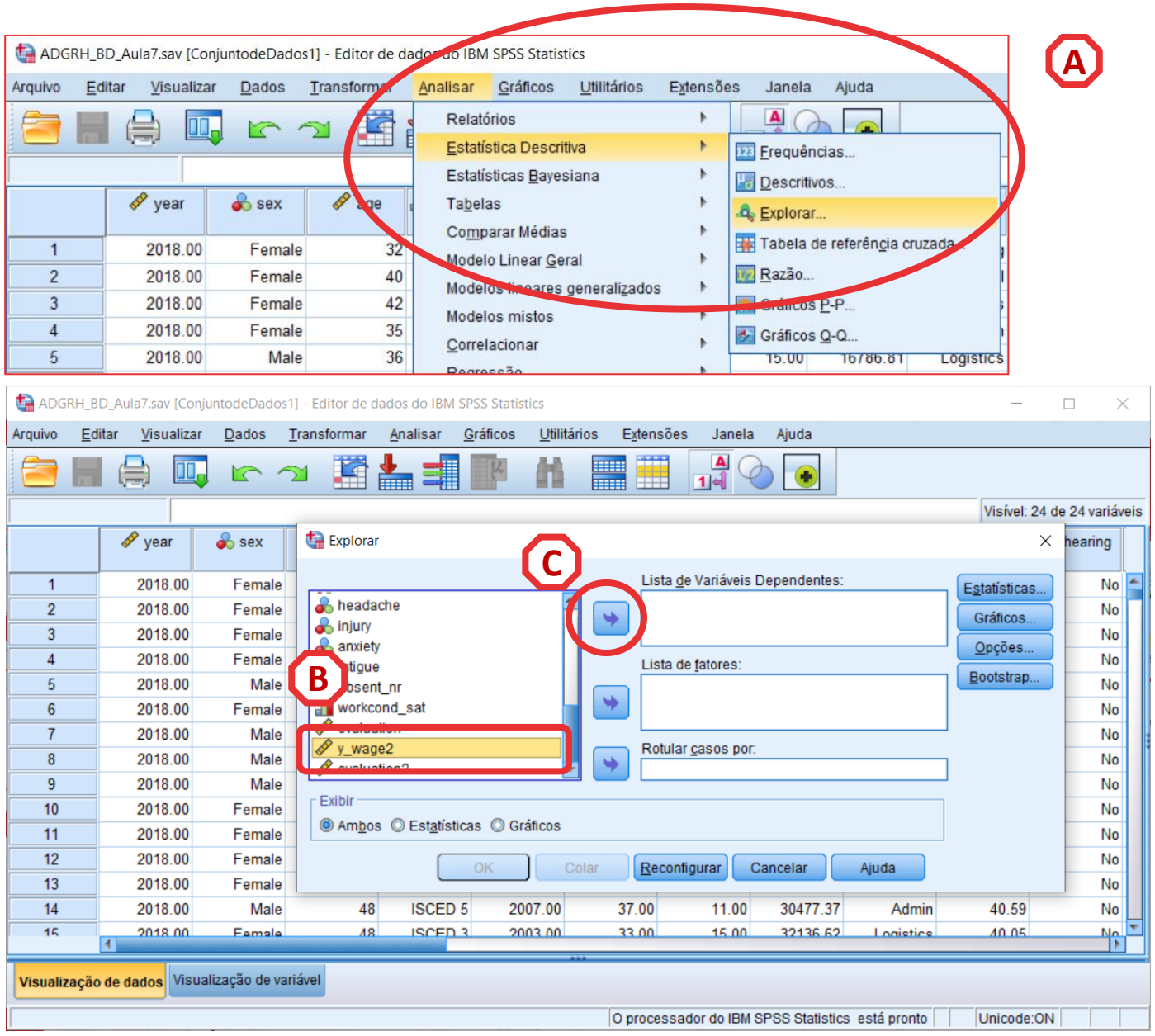
Estatística inferencial

Calcular o Intervalo de Confiança de uma média

Objetivo: Determinar o intervalo de confiança da média da variável que mede os salários na empresa (y_wage2)

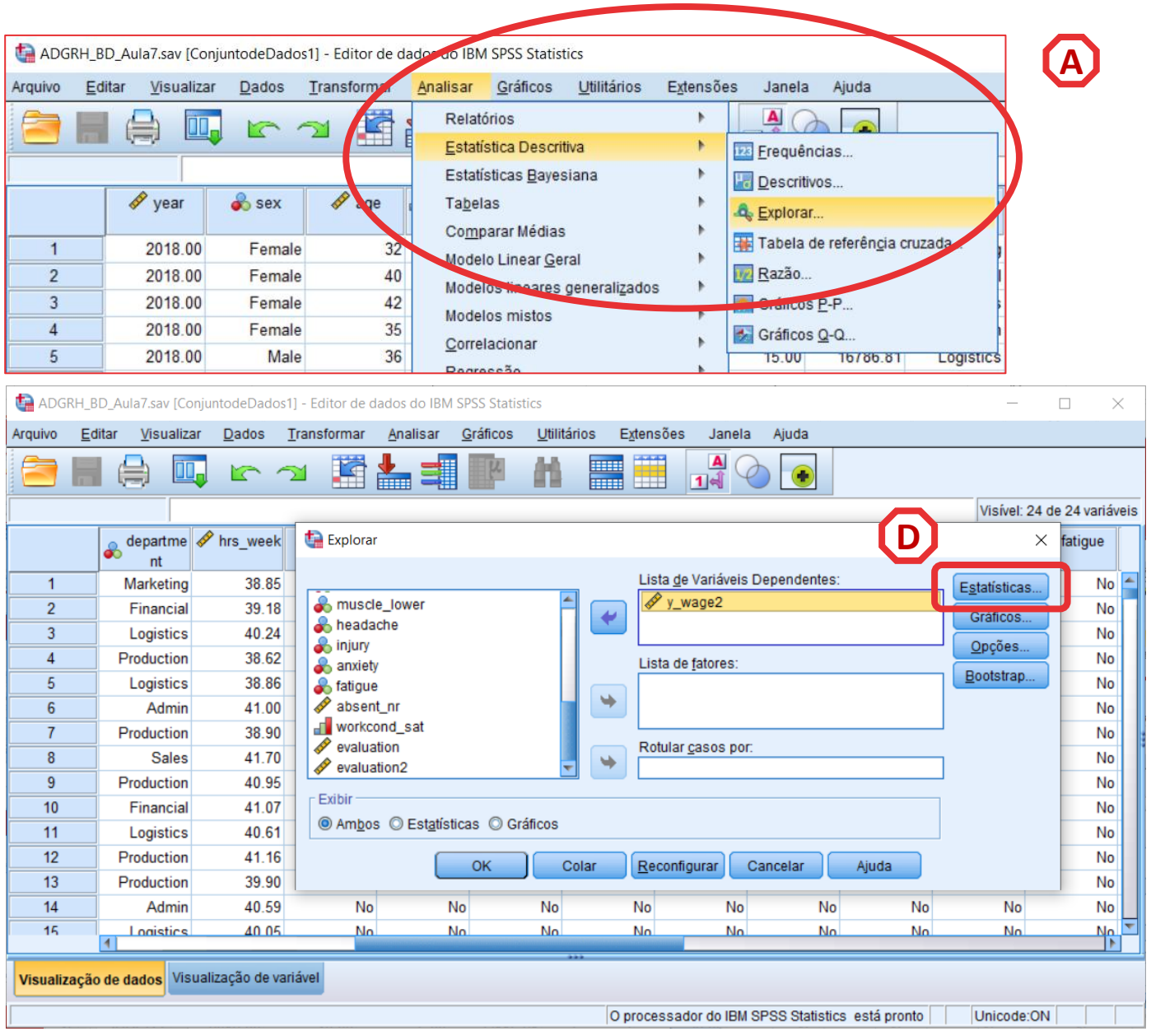
Intervalo de Confiança: Média

- Selecionar ‘Analisar’ / ‘Estatísticas Descritivas’ / ‘Explorar’
- Selecionar a variável ‘y_wage2’
- Colocar na caixa ‘Lista de Variáveis Dependentes’



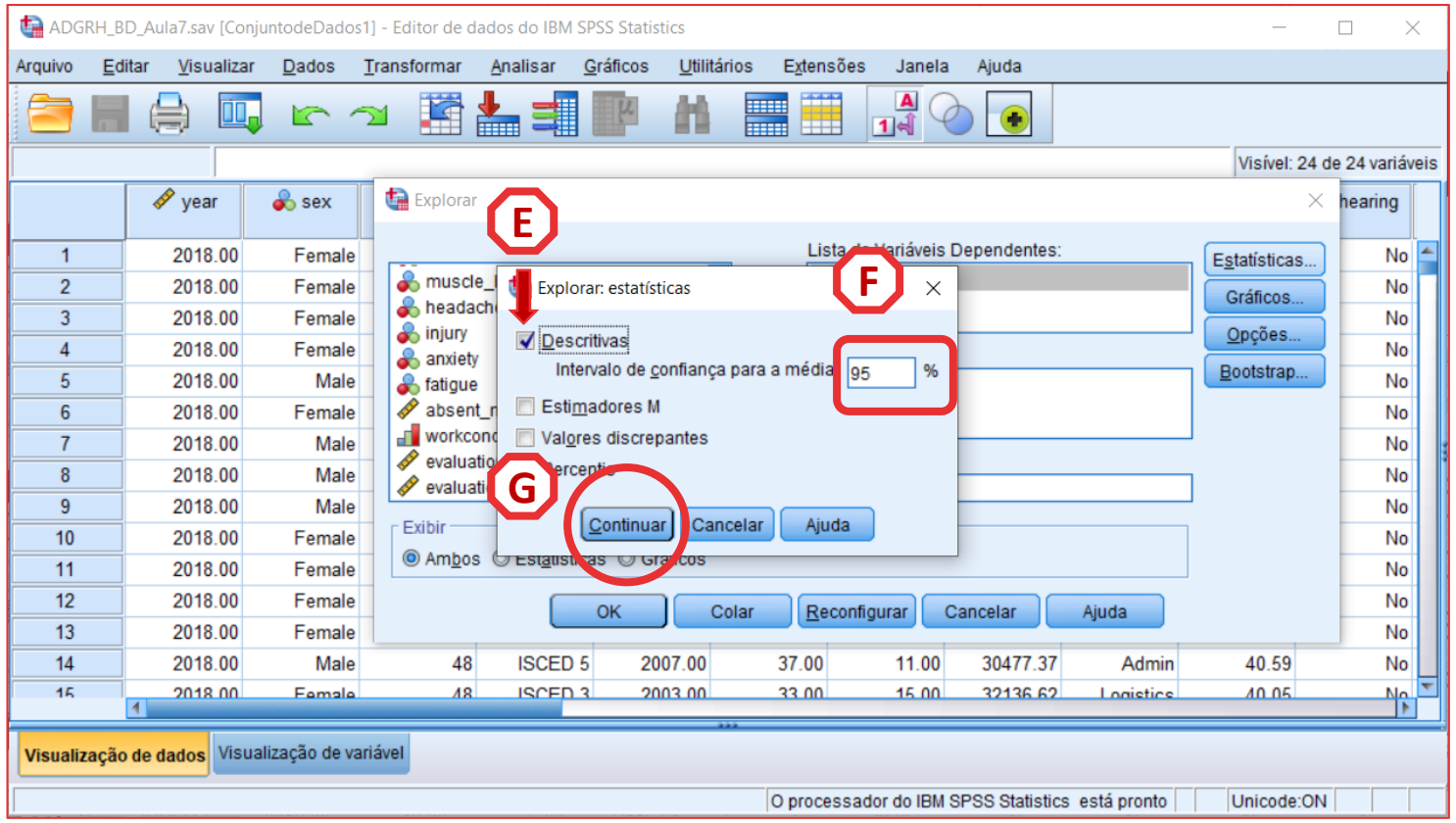
Intervalo de Confiança: Média

- Selecionar ‘Analisar’ / ‘Estatísticas Descritivas’ / ‘Explorar’
- Selecionar a variável ‘y_wage2’
- Colocar na caixa ‘Lista de Variáveis Dependentes’
- Selecionar ‘Estatísticas’



Intervalo de Confiança: Média

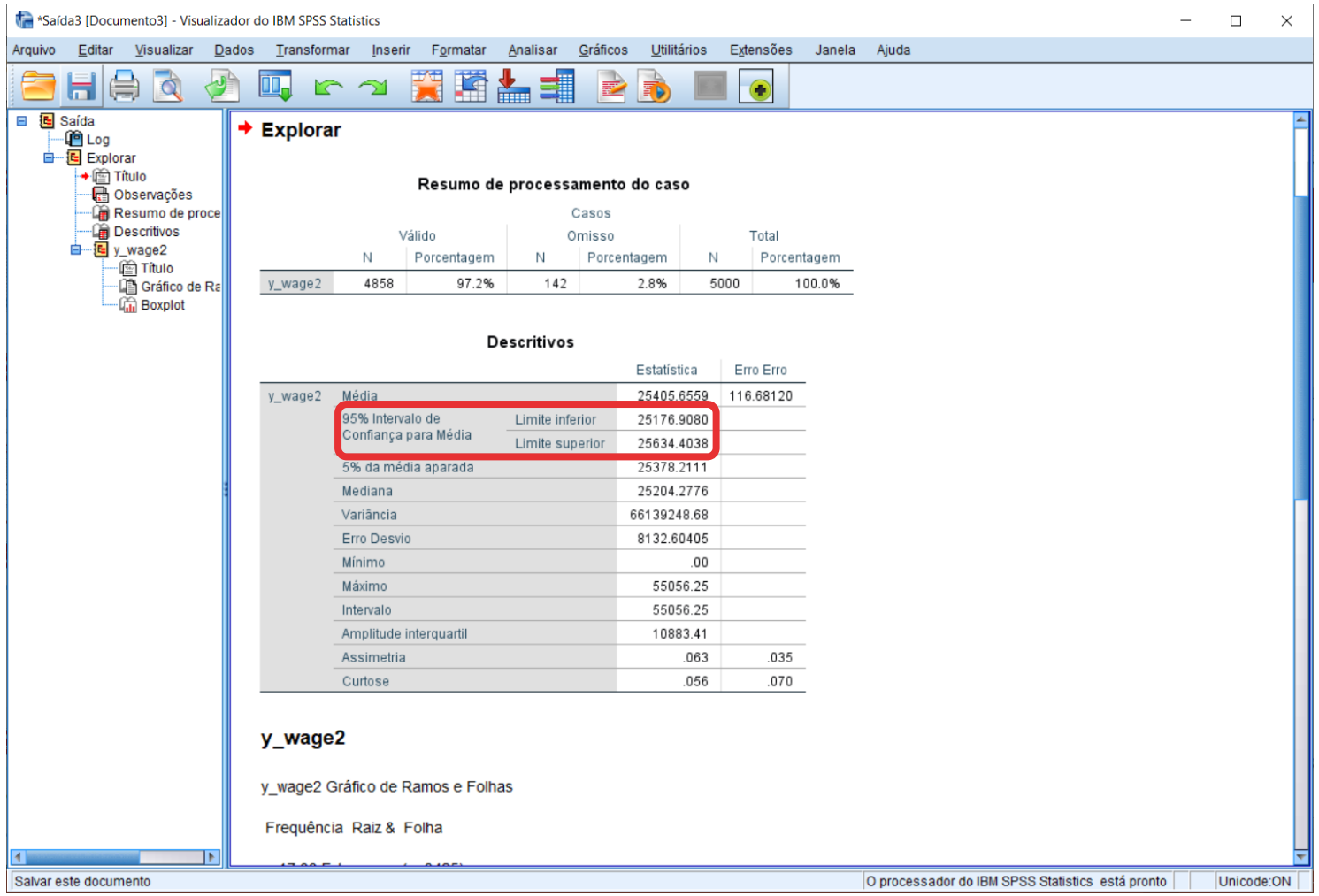
- Selecionar ‘Analisar’ / ‘Estatísticas Descritivas’ / ‘Explorar’ — A
- Selecionar a variável ‘y_wage2’ — B
- Colocar na caixa ‘Lista de Variáveis Dependentes’ — C
- Selecionar ‘Estatísticas’ — D
- Selecionar ‘Descritivos’ — E
- Definir um Grau de Confiança de ’95%’ — F
- Selecionar ‘Continuar’/OK — G



Intervalo de Confiança: Média

O resultado é publicado no 'Visualizador de Resultados'

PODEMOS DIZER, COM 95% DE CONFIANÇA, QUE O VALOR DO SALÁRIO MÉDIO ANUAL NA POPULAÇÃO ESTÁ ENTRE €25.177 E €25.634.



ANÁLISE DE DADOS EM GRH

- **O que precisam mesmo mesmo de saber**
  - Usamos a análise estatística porque queremos descrever e compreender os nossos dados **e tirar conclusões sobre eles.**
  - A estatística inferencial permite tirar conclusões a partir dos meus dados, inferindo sobre a população ou universo que quero estudar.
  - Para isso, tenho de ter uma boa amostra, com tamanho suficiente e, idealmente, probabilística.
  - **As conclusões inferidas têm sempre incerteza (ou erro amostral) e por isso fala-se em margem de erro, intervalos de confiança, e define-se graus de confiança.**

- **O que precisam mesmo mesmo de saber**

  - **Ao definir um nível ou grau de confiança**, consigo estimar a **margem de erro** das minhas conclusões sobre a população (assumindo que a minha amostra é aleatória).
  - Os **níveis de confiança** convencionados para o reporte estatístico são 90%, 95% e 99%. Nas ciências sociais, é habitual usar o nível de 95%.
  - A **margem de erro** das minhas conclusões define **o intervalo de confiança**: o intervalo de confiança é o intervalo estimado para o parâmetro populacional, com base na estimativa da amostra ± margem de erro
  - Não precisamos de saber as fórmulas, mas convém olhara para elas e perceber:
    - Para a mesma amostra, se quiser ter um nível de confiança maior sobre o que se passa na população, terei de aumentar a margem de erro (= intervalo de confiança maior, maior amplitude, menor precisão).
    - Para a mesma amostra, se optar por um nível de confiança mais baixo, vou diminuir a margem de erro(= intervalo de confiança menor, menor amplitude, maior precisão).
    - Quanto maior for a amostra, menor será a margem de erro(= intervalo de confiança menor, menor amplitude do erro, maior precisão).

ANÁLISE DE DADOS EM GRH

- **Intervalo de Confiança**
  - **Fornece um conjunto de valores plausíveis da estimativa (ex. média) na população.**

- **Teste de Hipóteses**
  - **Implica a formulação de hipóteses formais**
  - **Força uma tomada de decisão relativa à significância estatística**

ANÁLISE DE DADOS EM GRH

- **Intervalo de Confiança**
  - **Fornece um conjunto de valores plausíveis da estimativa (ex. média) na população.**
  - **A partir da minha amostra, posso estimar o comprimento médio dos bigodes na população.**
  - **Com base em pressupostos estatísticos e em decisões associadas ao grau de confiança,** consigo apresentar um intervalo de valores (intervalo de confiança) que expressa a incerteza da estimativa.

- **Teste de Hipóteses**
  - **Implica a formulação de hipóteses formais**
  - **Força uma tomada de decisão relativa à significância estatística**
  - **A partir da minha amostra, posso testar se o comprimento médio dos bigodes varia entre gatinhos e gatinhas.**
  - **Com base em pressupostos estatísticos e num grau de confiança definido,** tiro conclusões sobre a probabilidade de existirem diferenças observadas na amostra

MAS PARA O TESTE DE HIPÓTESES, PRECISAMOS DE:

- **Definir as hipóteses estatísticas**
  - **Nula**
  - **Alternativa**
- **Assumir a hipótese nula como verdadeira** (para calcular a sua probabilidade de ocorrer)

- **Teste de Hipóteses**
  - **Envolve a formulação de duas hipóteses alternativas**
    - **Hipótese Nula (H_0)**
      - **Determina o valor do parâmetro da população que se pretende testar (ex. média, proporção, etc.)**
      - **Exprime-se sobre a forma de uma igualdade (=)**
    - **Hipótese Alternativa (H_1)**
      - **Determina que o valor do parâmetro é diferente do que o definido pela Hipótese Nula**
      - **Consequentemente pode exprimir-se de uma destas formas**
        - **$\neq H_0$ - Parâmetro é diferente do que é definido pela Hipótese Nula**
        - **$> H_0$ - Parâmetro é maior do que é definido pela Hipótese Nula**
        - **$< H_0$ - Parâmetro é menor do que é definido pela Hipótese Nula**

PORQUÊ?

ANÁLISE DE DADOS EM GRH

- **Método científico**

  - **Transformação de conceitos em medidas mensuráveis**

  A operacionalização envolve a definição clara de como os conceitos abstratos nas hipóteses serão medidos e observados na prática.

  - **Recolha de dados**

  A operacionalização torna possível realizar observações, experiências, inquéritos, entrevistas, questionários e outras atividades de pesquisa para reunir evidências que apoiam ou refutam as hipóteses.

  - **Operacionalização de hipóteses**

  Após a coleta de dados, os cientistas usam técnicas de análise estatística para avaliar as evidências e determinar se as hipóteses são suportadas ou refutadas. As hipóteses cientificas têm de ser **testáveis, falsificáveis, claras e precisas.**

ANÁLISE DE DADOS EM GRH

- **Mas porquê a hipótese nula?**
  - **Base de referência clara**:

A hipótese nula fornece uma afirmação inicial clara que pode ser testada. A hipótese alternativa, por outro lado, é mais flexível e pode assumir muitas formas diferentes, tornando-a menos útil como ponto de partida para a análise.

  - **Falsificabilidade:**

A hipótese nula é formulada de forma a ser falsificável por meio de observações empíricas, o que significa que estamos dispostos a aceitar a possibilidade de que ela seja rejeitada se houver evidência suficiente para isso. A hipótese alternativa, é geralmente formulada para expressar o efeito ou relação que os pesquisadores esperam encontrar, e não é necessariamente tão facilmente falsificável.

NULA: Não há cisnes negros ALT: Há alguns cisnes negros

  - **Objetividade:**

A hipótese nula ajuda a manter a objetividade na pesquisa, pois não é influenciada pelas expectativas dos pesquisadores. Os cientistas devem testar a hipótese nula mesmo que acreditem que a hipótese alternativa seja verdadeira.

  - **Quantificação da Incerteza:**

A hipótese nula desempenha um papel crucial na quantificação da incerteza associada às observações. Ao comparar os resultados observados com a hipótese nula, os cientistas podem calcular a probabilidade de que os resultados sejam simplesmente devido ao acaso. Isso ajuda a diferenciar entre resultados que podem ser explicados por flutuações aleatórias e aqueles que são realmente significativos.

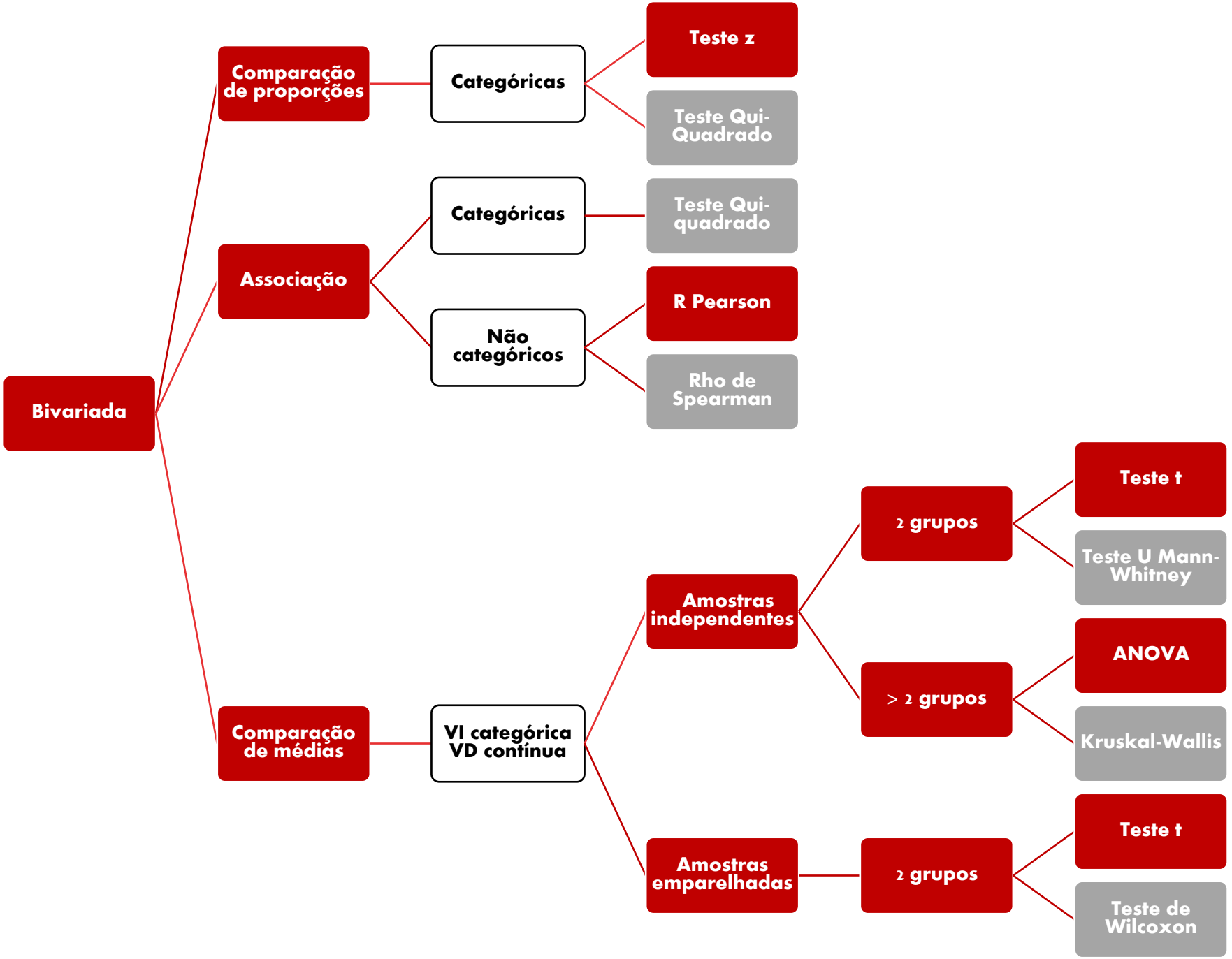
Não há efeito. Vs. Há algum efeito.

ANÁLISE DE DADOS EM GRH

- **A escolha dos testes de hipóteses a usar, vai depender dos objetivos e das características das variáveis**

- **Do nível de medida da variável**
  - Nominal, ordinal, intervalar ou de razão

- **Das características da distribuição das variáveis**
  - Testes paramétricos (Assumem que os dados seguem uma distribuição normal)
  - Testes não paramétricos


Testes Paramétricos
Teste não paramétrcos
Bivariada
Comparação de proporções
Categóricas
Teste z
Teste Qui-Quadrado
Associação
Categóricas
Teste Qui-quadrado
Não categóricos
R Pearson
Rho de Spearman
Comparação de médias
VI categórica VD contínua
Amostras independentes
2 grupos
Teste t
Teste U Mann-Whitney
> 2 grupos
ANOVA
Kruskal-Wallis
Amostras emparelhadas
2 grupos
Teste t
Teste de Wilcoxon

ANÁLISE DE DADOS EM GRH

Para que serve no contexto de um relatório de dados?

Nesta aula são identificados conceitos que suportam a capacidade de inferência da análise de dados de uma amostra.

Saber identificar o tipo de amostragem e tamanho da amostra face à população alvo, é fundamental, bem como compreender as implicações que trás às conclusões da análise.

Reportar a incerteza, referindo intervalos de confiança ou os erros associados ao teste de hipóteses são fundamentais para o rigor no reporte dos dados.

Materiais suplementares

- **Podemos calcular o Intervalo de Confiança de uma proporção/percentagem?**
- **Sim, mas o SPSS é particularmente limitado**
  - **Regra geral: não é possível calcular (diretamente) o Intervalo de Confiança de uma proporção/percentagem**
  - **Exceções**
    - **Quando fazermos testes de significância estatística**
    - **Para variáveis binomiais (com valores 0 e 1)**
      - **Pressuposto: a média dessa variável representa o proporção de observações com valor 1**

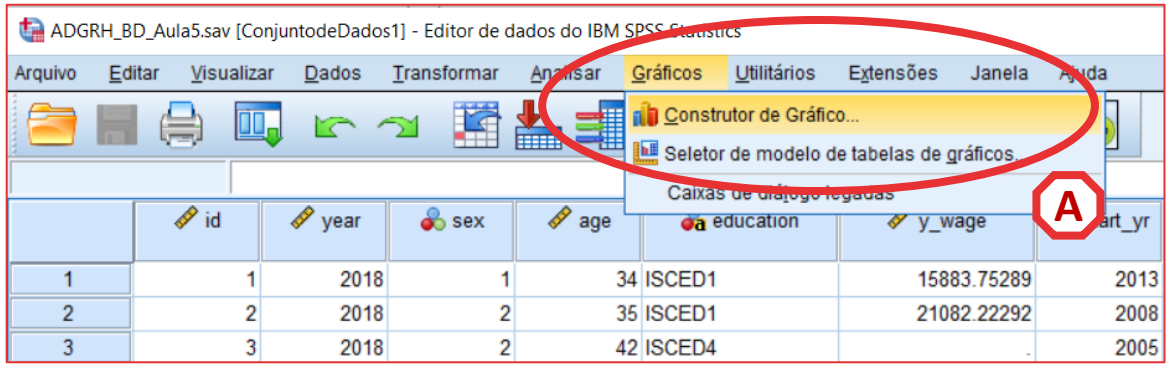
Alternativas?

- **Fazer o cálculo à mão**
- **Representar o Intervalo de confiança graficamente**
- **Procurar calculadoras online:**

https://select-statistics.co.uk/calculators/confidence-interval-calculator-population-proportion/

Gráfico de Barras com Erros

- Selecionar ‘Gráficos’ / ‘Construtor de Gráfico’ (A)
- Selecionar ‘Não mostrar este diálogo novamente’ (B)
- Selecionar ‘OK’ (C)



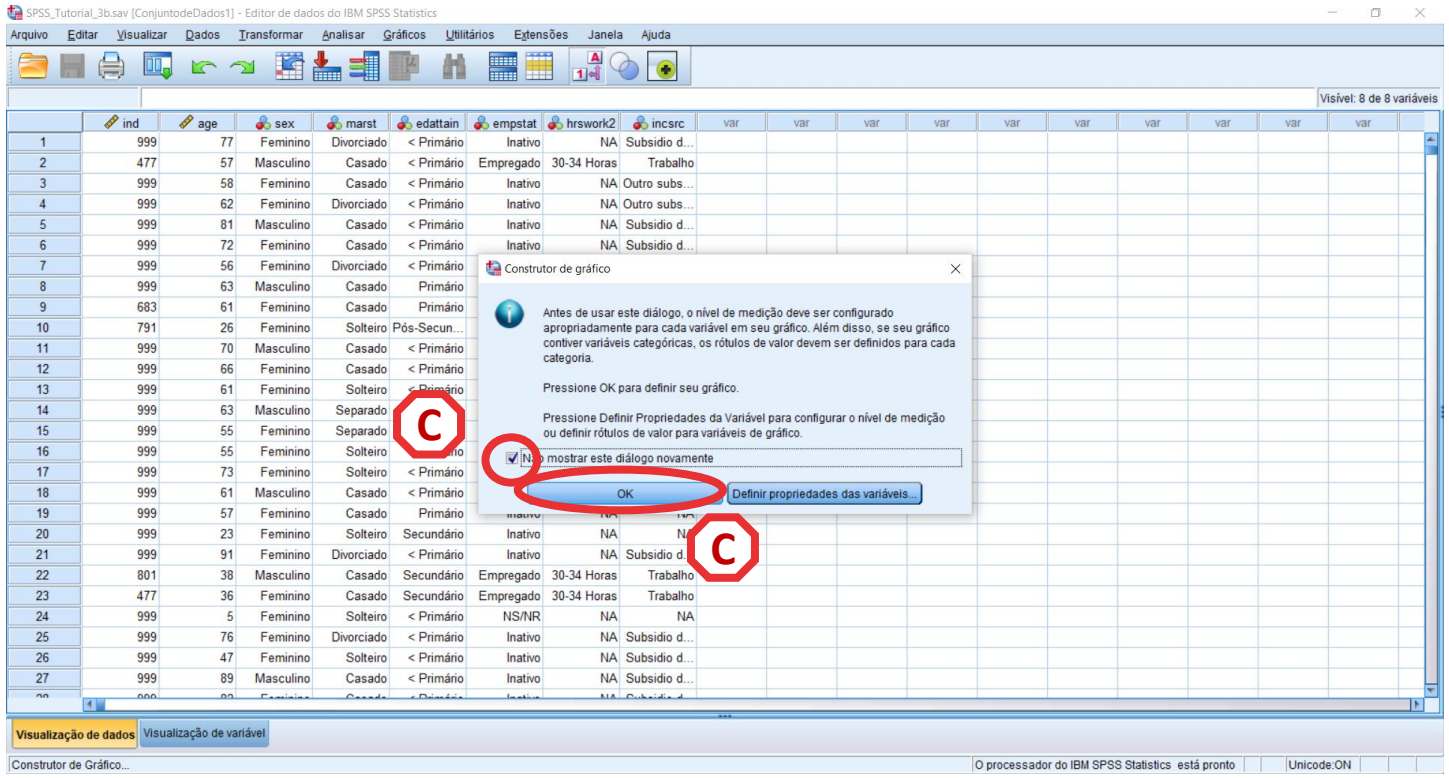


Gráfico de Barras com Erros

- **Selecione ‘Barras’** (D)
- **Selecionar (com duplo-clique) o Gráfico de Barras (simples)** (E)

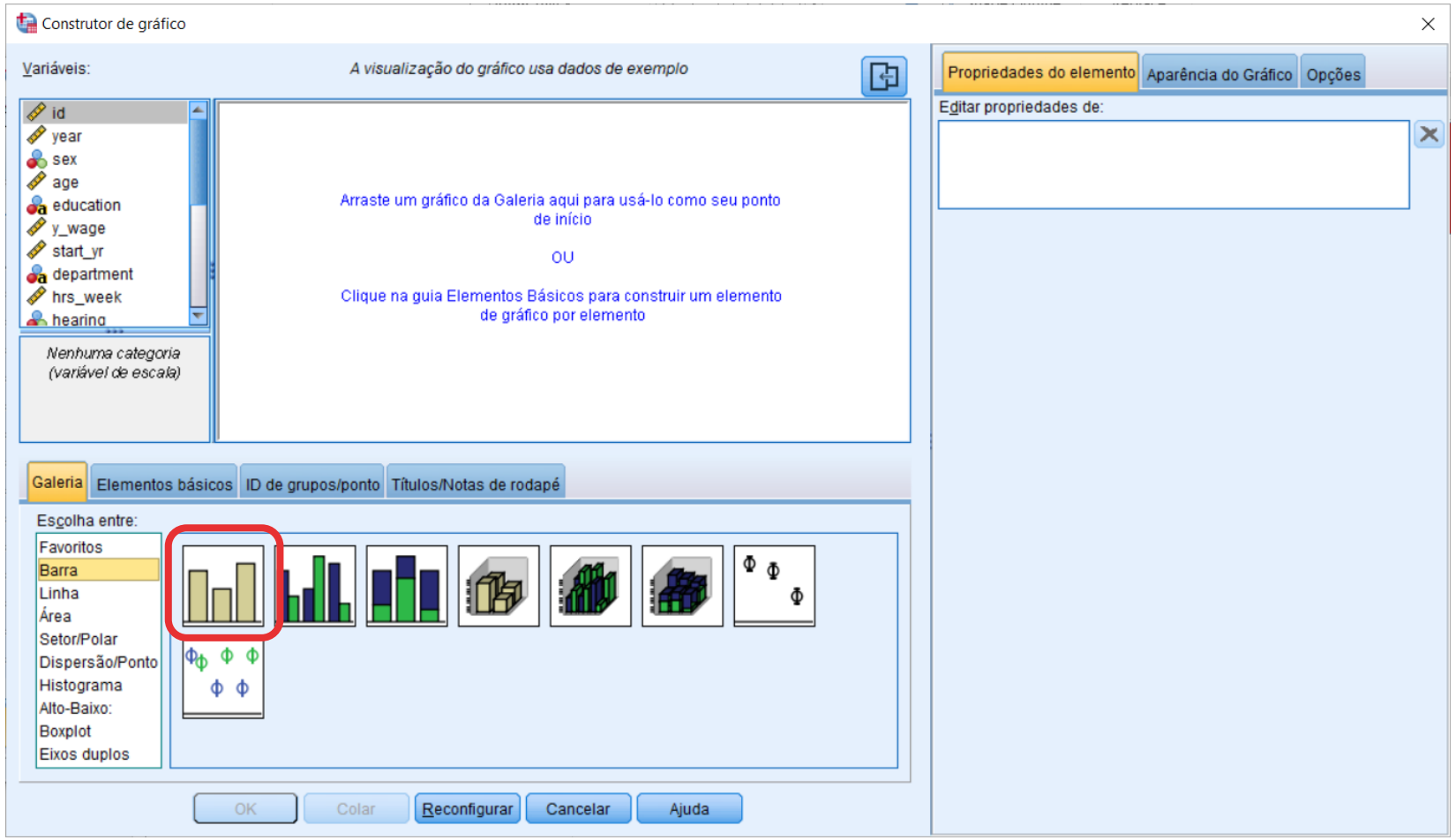


Gráfico de Barras com Erros

- Selecione ‘Barras’ (D)
- Selecionar (com duplo-clique) o Gráfico de Barras (simples) (E)
- Selecionar a variável ‘department’ (F)
- Colocar a variável ‘department’ no ‘Eixo X’ (G)

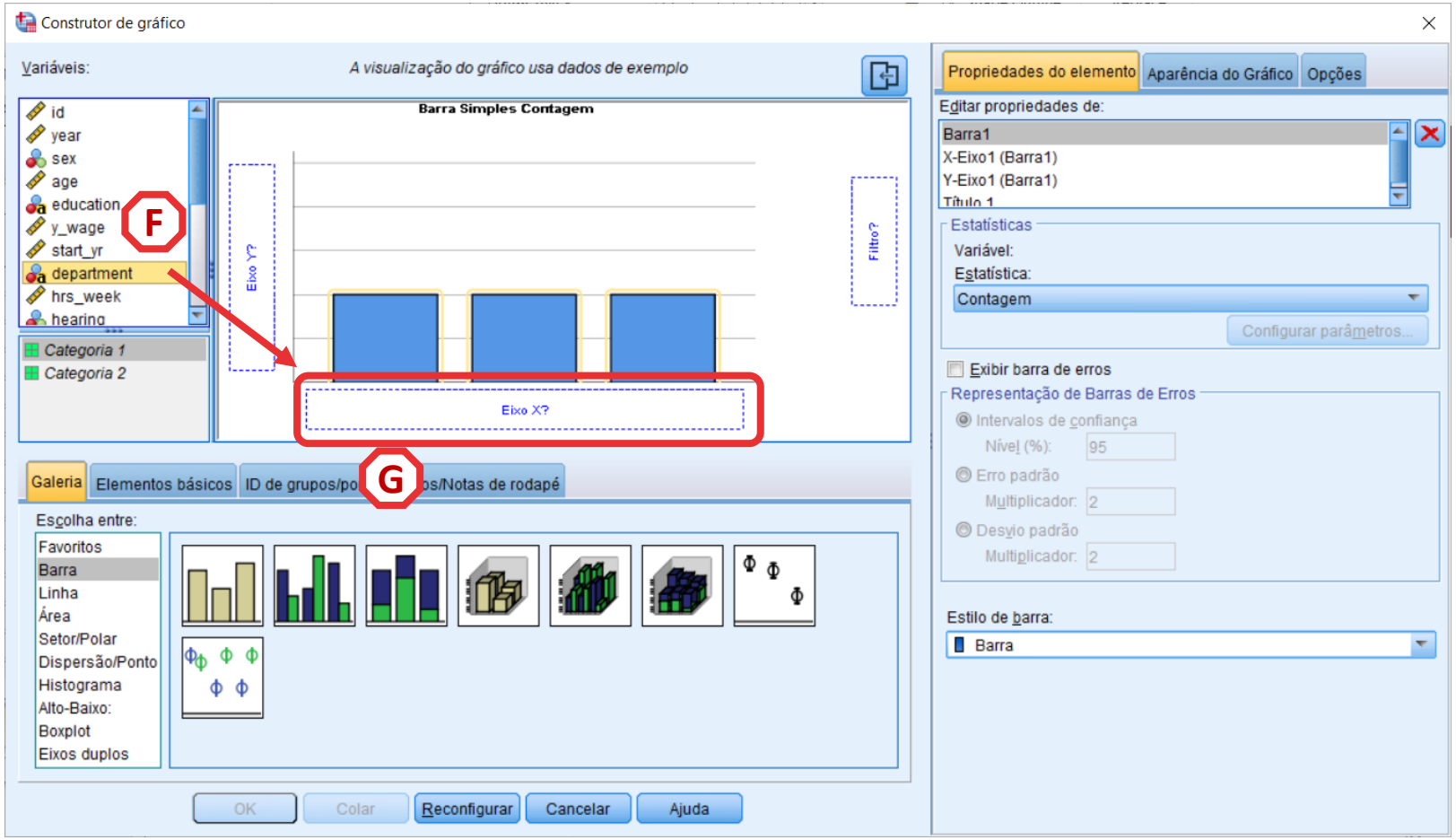


Gráfico de Barras com Erros

- Selecione ‘Barras’
- Selecionar (com duplo-clique) o Gráfico de Barras (simples)
- Selecionar a variável ‘department’
- Colocar a variável ‘department’ no ‘Eixo X’
- Selecionar ‘Exibir Barra de Erros’
- Selecionar ‘OK’

D

E

F

G

H

I

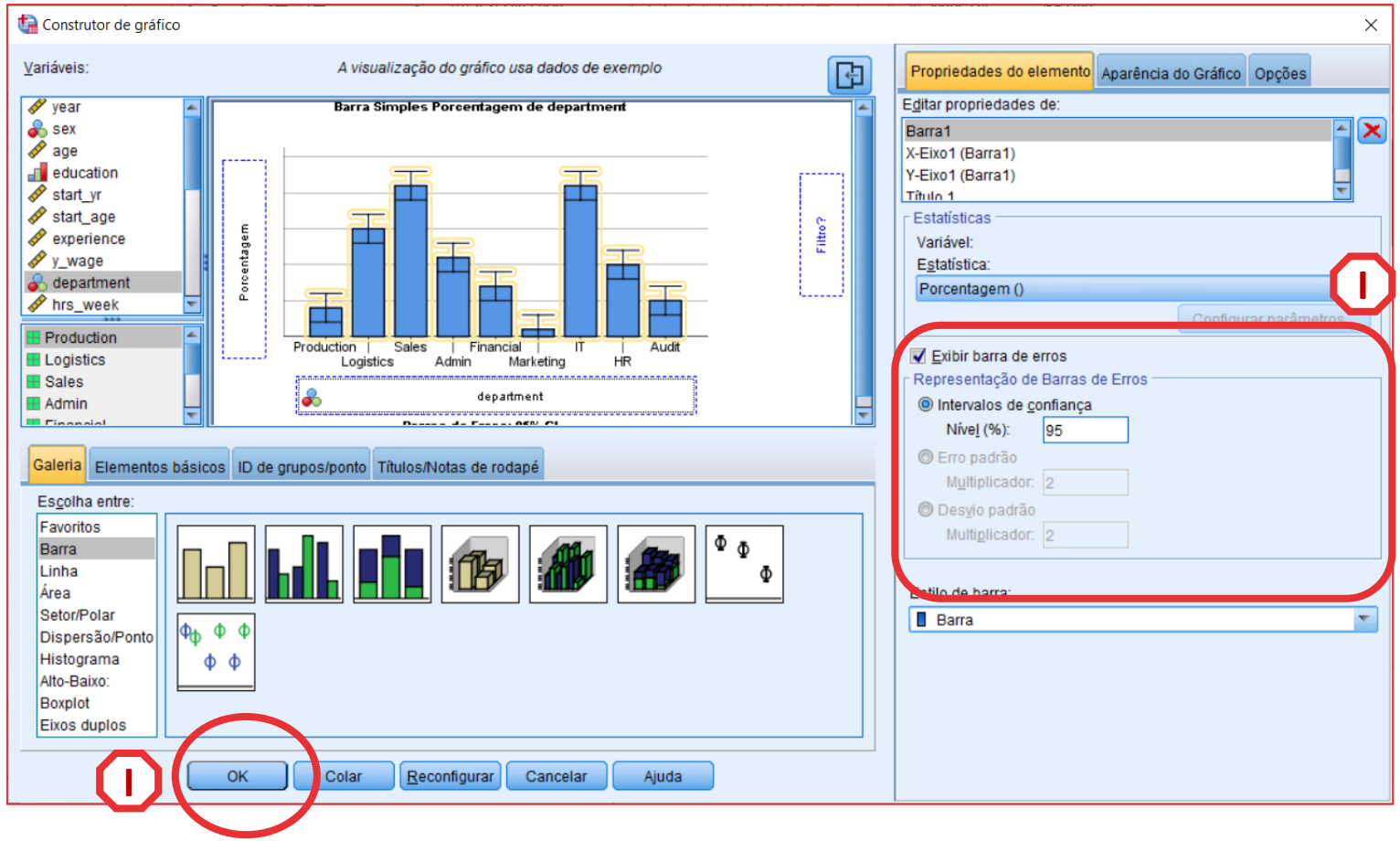
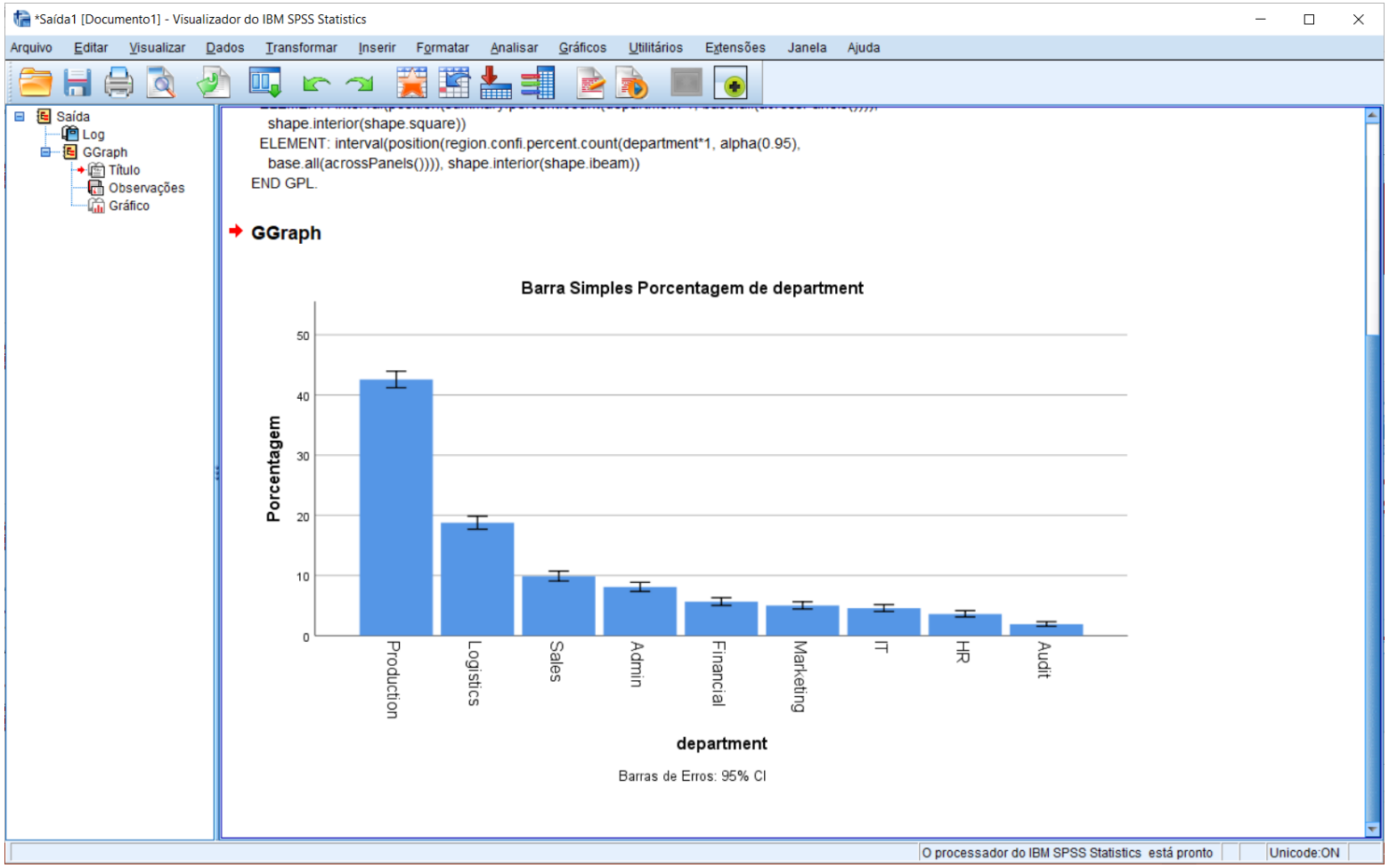


Gráfico de Barras com Erros

- **O gráfico é publicado no ‘Visualizador de Resultados’**



ANÁLISE DE DADOS EM GRH

Na calculadora

Temos de incluir alguma informação e as calculadoras aplicam a fórmula que conhecemos

PODEMOS DIZER, COM 95% DE CONFIANÇA, QUE A PERCENTAGEM DE COLABORADORES NO DEPARTAMENTO DE PRODUÇÃO, NESTA EMPRESA DE 10000 PESSOAS ESTÁ ENTRE

41,6% E 43,6%

Calculator

What is your sample proportion?	42.6	%
What confidence level do you need? Typical choices are 90%, 95%, or 99%	95	%
How big is your sample?	5000	
How big is the population?	10000	
Your confidence interval is	(41.63 , 43.57)	

Alternative Scenarios

With a sample proportion of	1 %	10 %	50 %
Your confidence interval would be	(0.8 , 1.2)	(9.41 , 10.59)	(49.02 , 50.98)
With a confidence level of	90 %	95 %	99 %
Your confidence interval would be	(41.79 , 43.41)	(41.63 , 43.57)	(41.33 , 43.87)
With a sample size of	100	500	10000
Your confidence interval would be	(32.96 , 52.24)	(38.38 , 46.82)	(42.6 , 42.6)
With a population size of	1500	5000	10000
Your confidence interval would be	(NaN , NaN)	(42.6 , 42.6)	(41.63 , 43.57)

https://select-statistics.co.uk/calculators/confidence-interval-calculator-population-proportion/

Exercícios em autonomia

- **Assumimos que esta amostra é probabilística. Quais são os intervalos confiança da média para a totalidade da amostra do número de dias em baixa médica (absent_nr) para o nível de 90% e 95%?**
- **Represente os intervalos de confiança a 95% para as categorias de nível de satisfação com as condições de trabalho (workcond_sat)**
- **Qual a percentagem de pessoas com sintomas de ansiedade nesta empresa (se tamanho da população for N= 10.000)?**