

Trabalho Parte 2

Trabalho Parte 2: Inferência Estatística – intervalo de confiança e testes de hipóteses

Dicionário dos dados:

Variável	Tipo	Nome da variável	Descritivo e unidades
Idade	Objetivo	age	int (days)
Altura	Objetivo	height	int (cm)
Peso	Objetivo	weight	float (kg)
Gênero	Objetivo	gender	1: Feminino 2: Masculino
Pressão sistólica (contração e saída de sangue)	Exame	ap_hi	Int (mmHg)
Pressão diastólica (relaxamento e entrada de sangue)	Exame	ap_lo	Int (mmHg) 1: normal, 2: above normal, 3: well above normal
Colesterol	Exame	cholesterol	1: normal, 2: above normal, 3: well above normal
Glicose	Exame	gluc	Binary 0: Não fumante 1: Fumante
Se é fumante	Subjetivo	smoke	binary
Se ingere bebida alcoólica	Subjetivo	alco	binary
Se pratica atividade física	Subjetivo	active	binary
Presença ou ausência de doença cardiovascular	Target	cardio	binary

Definições:

População: Considere todos os elementos (70k indivíduos) como a sua população.

Amostra: uma amostra aleatória dessa População

Obs: utilize algum programa (R/Python/Matlab/Scilab/Octave/..) ou software (Excel/Calc/...) para realizar o processo de amostragem aleatória simples.

1. Análise exploratória simples – Parte 1

2. Distribuições amostrais – Parte 1

3. Inferência Estatística – Parte 3

3.1 Intervalos de confiança e Testes de Hipóteses

Explique brevemente (com suas palavras) as questões abaixo.

3.1.1 Explique a diferença entre estimação pontual e intervalar.

3.1.2 Descreva o intervalo de confiança de 95% para uma média μ da população. Como o intervalo é interpretado.

3.1.3 Quais os fatores que afetam o comprimento (amplitude) de um intervalo de confiança para uma média?

Intervalos de confiança

3.1.4 Descreva as similaridades e diferenças entre a distribuição t e a distribuição normal padrão. Se você tentasse construir um intervalo de confiança, quando usaria uma no lugar da outra?

3.1.5 Tome uma amostra aleatória de tamanho $n = 20$ e calcule os intervalos de confiança bilaterais para as médias da variável pressão sistólica e diastólica da população de mulheres fumantes maiores de 45 anos. (escolha um α entre 0.05 e 0.01)

- a. Interprete esses intervalos de confiança
- b. O intervalo calculado a partir de sua amostra acertou os verdadeiros parâmetros populacionais?

3.1.6 Assuma que você conhece a σ^2 populacional da altura dos homens maiores que 50 anos (obtenha esse valor de sua população). Qual deveria ser o tamanho da amostra para que o intervalo de confiança bilateral 95% para a média da altura tenha uma amplitude de 5 cm?

3.1.7 Tome uma amostra aleatória de tamanho $n = 300$ e calcule os intervalos de confiança bilaterais para as proporções de indivíduos que ingerem bebida alcoólica e de indivíduos com presença de doença cardiovascular, para homens e mulheres fumantes maiores de 50 anos. (escolha um α entre 0.05 e 0.01)

- a. Interprete esses intervalos de confiança
- b. Os intervalos calculados a partir de sua amostra acertaram os verdadeiros parâmetros populacionais?

3.1.8 Tome uma amostra aleatória de tamanho $n = 45$ e calcule os intervalos de confiança bilaterais para a variância da variável peso, da população com presença de doença cardiovascular, homens, com menos de 50 anos. (escolha um α entre 0.05 e 0.01)

- a. Interprete o intervalo de confiança
- b. O intervalo calculado a partir de sua amostra acertou o verdadeiro parâmetros populacional?

Testes de hipóteses

3.1.9 Teste a hipótese de que a proporção de homens que ingerem bebida alcoólica é maior que a de mulheres. Explique como foi feito o processo amostral. (Escolha um tamanho de amostra > 300 , escolha um α entre 0.05 e 0.01).

1. Descreva as hipóteses, nula e alternativa
2. Escolha o valor de α ,
3. Defina a estatística do teste,
4. Calcule o valor-P,
5. Decida o teste
6. Conclua o teste com base nas evidências da amostra.

3.1.10 Faça outros 3 testes, um para diferença de médias, diferença/razão de variâncias, e outro para diferença de proporções.

1. Descreva as subpopulações escolhidas (população 1 e 2)
2. Descreva as hipóteses, nula e alternativa
3. Escolha o valor de α ,
4. Defina a estatística do teste,
5. Calcule o valor-P,
6. Decida o teste
7. Conclua o teste com base nas evidências da amostra.

Para os testes verifique os pressupostos de normalidade dos dados, no caso do **teste de médias**, e se as variâncias são iguais ou diferentes (homogeneidade das variâncias). Para esse trabalho, caso não sejam atendidos os pressupostos apenas anote essa observação. Se preferir o teste não-paramétrico U de Mann-Whitney pode ser aplicado nesses casos de diferenças de médias que não atendem aos pressupostos.

Teste de Homogeneidade de Variâncias: Teste de Razão de Variâncias ou Teste de Levene`s

Normalidade dos dados: QQ-Plot, Teste de Shapiro-Wilk, ou Kolmogorov-Smirnov ou Anderson-Darling.