

## Intervalos de confiança

Fornece um intervalo de valores, centrado na estatística amostral, no qual julgamos, com um risco conhecido de erro, estar o parâmetro da população.

Utilizar os arquivos **AMP\_modulo.sprint.csv** e **AMP\_esforco.sprint.csv**. Os dois arquivos se fazem necessários porque a empresa pretende fazer uma análise das sprints por módulo.

- O arquivo AMP\_esforco\_sprint.csv contém as seguintes variáveis (colunas): ID\_SPINT (identificador da sprint), Estimativa\_UCP (estimativa do esforço do projeto medido em Use Case Points); Estimativa\_Horas (esforço previsto em horas, estimado a partir dos UCPs); Realizado\_Horas (esforço realizado em horas); Entrega\_no\_Prazo (flag identificando se a sprint terminou no prazo '1' ou não '0').
- O arquivo AMP\_modulo\_sprint.csv contém o relacionamento dos identificadores de módulos com os identificadores de sprints. Para obter os dados combinados é necessário fazer um join.

```
% Formatação para legibilidade
format short g

% Ler o arquivo AMP modulo sprint
% Ler o arquivo AMP esforco sprint
MS = readtable('AMP_modulo_sprint.csv');
ES = readtable('AMP_esforco_sprint.csv');

% Outer Join
Sprints = outerjoin(MS,ES,'Keys',{'ID_SPINT'}, 'MergeKeys',true);

% Transformar ID_MODULO para o tipo categorical para permitir filtragem por linha
Sprints.ID_MODULO = categorical(Sprints.ID_MODULO)
```

Sprints = 80x6 table

...

	ID_MODULO	ID_SPINT	Estimativa_UCP	Estimativa_Horas
1	M01	1	16.744	91.89
2	M01	2	41.333	225.74
3	M02	3	26.025	111.65
4	M02	4	11.483	49.69
5	M01	5	44.328	261.53
6	M01	6	49.295	283.86
7	M04	7	23.503	101.03
8	M03	8	43.74	186.99
9	M02	9	39.851	171.31
10	M01	10	17.165	130.8
11	M04	11	15.947	68.55
12	M05	12	66.551	284.08

	ID_MODULO	ID_SPRINT	Estimativa_UCP	Estimativa_Horas
13	M02	13	22.135	95.15
14	M04	14	10.395	44.69
15	M05	15	61.567	264.66
16	M01	16	3.7973	16.32
17	M02	17	23.381	100.51
18	M04	18	2.4077	10.35
19	M01	19	25.664	110.32
20	M02	20	39.604	170.25
21	M01	21	8.3113	135.73
22	M05	22	25.121	107.99
23	M01	23	57.063	288.29
24	M02	24	9.804	42.15
25	M03	25	20.241	87.01
26	M05	26	26.631	114.48
27	M01	27	15.045	84.68
28	M04	28	9.4157	40.48
29	M03	29	42.27	173.66
30	M04	30	15.382	66.13
31	M05	31	43.976	189.04
32	M02	32	4.7349	20.35
33	M04	33	35.471	152.48
34	M03	34	37.33	160.47
35	M05	35	6.8153	29.3
36	M01	36	12.016	51.65
37	M05	37	21.062	90.54
38	M01	38	8.5145	43.6
39	M04	39	47.069	217.28
40	M05	40	26.984	116
41	M01	41	26.808	115.24
42	M03	42	5.6851	94.44
43	M02	43	36.388	156.43
44	M02	44	4.8065	20.66
45	M03	45	6.3179	117.16

	ID_MODULO	ID_SPRINT	Estimativa_UCP	Estimativa_Horas
46	M04	46	14.757	63.44
47	M01	47	50.394	173.64
48	M03	48	6.9571	49.91
49	M01	49	45.294	214.71
50	M04	50	44.634	191.87
51	M05	51	64.621	277.79
52	M02	52	53.913	226.7
53	M03	53	36.304	156.06
54	M05	54	17.448	75.01
55	M04	55	16.226	69.75
56	M05	56	26.027	111.89
57	M02	57	8.1363	34.98
58	M03	58	30.196	129.81
59	M04	59	9.0114	38.74
60	M03	60	5.5682	63.94
61	M02	61	12.839	55.19
62	M01	62	45.99	154.71
63	M02	63	11.261	48.41
64	M03	64	4.4002	18.92
65	M04	65	38.199	164.21
66	M05	66	56.46	242.71
67	M03	67	5.4351	53.36
68	M04	68	51.396	220.25
69	M02	69	13.636	58.62
70	M05	70	41.154	176.91
71	M04	71	48.494	208.47
72	M01	72	7.2325	31.09
73	M03	73	5.9796	65.7
74	M02	74	7.0998	30.52
75	M01	75	15.126	65.02
76	M03	76	26.398	113.48
77	M04	77	11.882	59.92
78	M04	78	22.987	98.82

	ID_MODULO	ID_SPRINT	Estimativa_UCP	Estimativa_Horas
79	M05	79	30.437	130.84
80	M05	80	11.289	48.53

## Intervalo de confiança para a média

### Questão 1

Responder as seguintes perguntas sobre as sprints do módulo M02 que foram realizadas em menos do que 300 horas.

a) Quantos dessas sprints existem?

Dicas: (i) use uma expressão lógica para criar um vetor com o índice das sprints do módulo M02 que foram realizadas com menos do que 300 horas de esforço. (ii) utilize o vetor de índice para construir um vetor contendo apenas as sprints com essas características. (iii) o tamanho do vetor é igual à quantidade de projetos.

b) Qual a média de horas realizadas dessas sprints?

Dica: Utilize a função do MatLab para calcular a média.

c) Calcular o intervalo de confiança de 95% da média das horas realizadas para as sprints com essas características.

Dica: A função ICMP calcula o intervalo de confiança para a média da população. Recebe dois argumentos: um vetor com os valores da amostra e o nível de confiança. Retorna dois valores correspondentes ao início e fim do intervalo de confiança.

d) Calcular a amplitude do intervalo de confiança de 95%. (DICA: fim - início)

e) Calcular o intervalo de confiança de 99% da média das horas realizadas para as sprints com essas características.

Se o seu código estiver correto o resultado deve ser [ 39.21 , 144.92 ]

f) Calcular a amplitude do intervalo de confiança de 99%.

Coloque seu código abaixo.

```
% SUBSTITUIR OS COMANDOS ABAIXO CONFORME INDICADO
indice = (Sprints.Realizado_Horas < 300 & Sprints.ID_MODULO == 'M02');
sprintm02menor300 = length(indice);
% Quantidade de sprints do módulo M02 realizadas em menos do que 300 horas
% Substituir o comando a seguir
fprintf("Quantidade de sprints do módulo M02 realizadas em menos do que 300 horas = %d\n", sprintm02menor300);
```

Quantidade de sprints do módulo M02 realizadas em menos do que 300 horas = 80

```
% Média das horas realizadas para as sprints do módulo M02
% realizadas em menos do que 300 horas
% Substituir o comando a seguir
```

```
sprintm02menor300Media = mean(indice);
fprintf("Média das horas realizadas = %.2f\n", sprintm02menor300Media);
```

Média das horas realizadas = 0.20

```
% Intervalo de confiança de 95% da média as horas realizadas para sprints
% do módulo M02 realizadas em menos do que 300 horas
% Substituir o comando a seguir
%icmp95M02Menor300 = [0, 0];
fprintf("ICMP_95 da média as horas realizadas = [%.2f, %.2f]\n", ICMP(indice, .95));
```

ICMP\_95 da média as horas realizadas = [0.11, 0.29]

```
% Amplitude do Intervalo de confiança da média de 95% das sprints
% do módulo M02 realizadas em menos do que 300 horas
fprintf("Amplitude do ICMP_95 das horas realizadas = %.2f\n", (ICMP(indice, .95) - ICMP(indice, .05)));
```

Amplitude do ICMP\_95 das horas realizadas = -0.06  
Amplitude do ICMP\_95 das horas realizadas = 0.06

```
% Intervalo de confiança de 99% da média as horas realizadas para sprints
% do módulo M02 realizadas em menos do que 300 horas
% Substituir o comando a seguir
icmp99M02M300 = [0, 0];
fprintf("ICMP_99 da média das horas realizadas = [%.2f, %.2f]\n", ICMP(indice, .99));
```

ICMP\_99 da média das horas realizadas = [0.08, 0.32]

```
% Amplitude do Intervalo de confiança da média de 99% das sprints do módulo M02
% realizadas em menos do que 300 horas
% Substituir o comando a seguir
amplitude99 = 0;
fprintf("Amplitude do ICMP_99 da média das horas realizadas = %.2f\n", ICMP(indice, .1) - ICMP(indice, .99));
```

Amplitude do ICMP\_99 da média das horas realizadas = 0.11  
Amplitude do ICMP\_99 da média das horas realizadas = -0.11

## Intervalo de confiança para proporção

### Questão 2

Responder as seguintes perguntas sobre a proporção das sprints que foram entregues no prazo.

Dica: A variável Entrega\_no\_Prazo tem 1s (sprints entregues no prazo) e 0s (sprints não entregues no prazo).

a) Quantos projetos foram entregues no prazo?

Dica: A quantidade de sprints entregues no prazo é igual a quantidade de 1s na variável Entrega\_no\_Prazo.

b) Qual a proporção de projetos entregues no prazo?

Dica: É possível calcular a proporção de duas maneiras: (i) Dividir a quantidade de sprints entregues no prazo pela quantidade total de sprints. (ii) Calcular a média da variável Entrega\_no\_Prazo. Esse cálculo fornece a proporção porque a variável Entrega\_no\_Prazo é uma variável de Bernoulli (apenas 0s e 1s), e a média da variável de Bernoulli é igual à proporção.

c) Qual o intervalo de confiança de 99% da proporção das sprints que foram entregues no prazo?

Se o seu código estiver correto o resultado deve ser: [ 0.69, 0.94]

Dica: A função ICPP calcula o intervalo de confiança para a proporção da população. Recebe dois argumentos: a proporção dos dados da amostra (proporção dos projetos entregues no prazo) e o nível de confiança. Retorna dois valores correspondentes ao início e fim do intervalo de confiança.

```
% SUBSTITUIR OS COMANDOS ABAIXO CONFORME INDICADO
% Calcular quantidade de sprints entregues no prazo
n = sum(Sprints.Entrega_no_Prazo);
fprintf("Quantidade de sprints entregues no prazo = %d\n", n);
```

```
Quantidade de sprints entregues no prazo = 65
```

```
% Calcular a proporção de sprints entregues no prazo
% Substituir o comando a seguir
```

```
fprintf("Proporção de sprints entregues no prazo = %.2f\n", mean(Sprints.Entrega_no_Prazo));
```

```
Proporção de sprints entregues no prazo = 0.81
```

```
% Calcular o intervalo de confiança de 99% das sprints entregues no prazo
% Substituir o comando a seguir
```

```
fprintf("ICPP_99 para as sprints entregues no prazo = [%.2f, %.2f]\n", ICPP(Sprints.Entrega_no_Prazo, 0.99));
```

```
ICPP_99 para as sprints entregues no prazo = [0.70, 0.93]
```

## Intervalo de confiança para variância

### Questão 3

Responder as seguintes perguntas sobre as sprints do módulo M02 que foram realizados em menos do que 300 horas.

a) Calcular a variância das horas realizadas para essas sprints.

Dica: Utilizar a função var do MatLab.

b) Calcular o intervalo de confiança de 95% da variância das horas realizadas para essas sprints.

Dica: A função ICVP calcula o intervalo de confiança para a variância da população. Recebe dois argumentos: um vetor com os valores da amostra e o nível de confiança. Retorna dois valores correspondentes ao início e fim do intervalo de confiança.

c) Calcular o desvio padrão do esforço realizado em horas para esses projetos.

Dica: lembrar que o desvio padrão é igual à raiz quadrada da variância.

d) Calcular o intervalo de confiança de 95% do desvio padrão das horas realizadas para essas sprints.

Se seu código estiver correto o resultado deve ser: [53.00 , 111.04]

```
% SUBSTITUIR OS COMANDOS ABAIXO CONFORME INDICADO
indice = (Sprints.Realizado_Horas < 300 & Sprints.ID_MODULO == 'M02');
sprintsM02Menor300 = Sprints.Realizado_Horas(indice);

% Variância das horas realizadas para as sprints do módulo M02 realizadas
% em menos do que 300 horas
% Substituir o comando a seguir
varianciaM02Menor300 = var(indice);
fprintf("Variância das horas realizadas = %.2f\n", varianciaM02Menor300);
```

Variância das horas realizadas = 0.16

```
% Intervalo de confiança da variância de 95% das horas realizadas das sprints dos módulos
% realizadas em menos do que 300 horas
% Substituir o comando a seguir

fprintf("ICVP_95 da variância = [%.2f, %.2f]\n", ICVP(indice, .95));
```

ICVP\_95 da variância = [0.12, 0.23]

```
% Desvio padrão da variância das horas realizadas nas sprints do módulo M02
% realizadas em menos do que 300 horas
% Substituir o comando a seguir

fprintf("Desvio padrão da variância = %.2f\n", std(indice));
```

Desvio padrão da variância = 0.40

```
% Intervalo de confiança de 95% das horas realizadas nas sprints do módulo M02
% realizadas em menos do que 300 horas
% Substituir o comando a seguir

fprintf("ICDPP_95 da variância = [%.2f, %.2f]\n", sqrt(ICVP(indice, .95)));
```

ICDPP\_95 da variância = [0.35, 0.48]

## FORMATIVA

Realizar as alterações necessárias.

Colocar a opção Output Inline.

Salvar em pdf.

Fazer o upload no AVA.

Não envie o LiveScript.

## Funções

```
function icmp = ICMP(X, c)
alfa = 1 - c;
n = size(X, 1);
Xbarra = mean(X);
S = sqrt(var(X));
x1 = (Xbarra - tinv(1-alfa/2, n-1)*S/sqrt(n));
x2 = (Xbarra + tinv(1-alfa/2, n-1)*S/sqrt(n));
icmp = [x1, x2];
end
```

```
function icpp = ICPP(p, n, c)
alfa = 1 - c;
S = sqrt((p*(1-p)));
p1 = (p - norminv(1-alfa/2, 0, 1)*S/sqrt(n));
p2 = (p + norminv(1-alfa/2, 0, 1)*S/sqrt(n));
icpp = [p1, p2];
end
```

```
function icvp = ICVP(X, c)
alfa = 1 - c;
n = size(X, 1);
X2Sup = chi2inv(alfa/2, (n-1));
X2Inf = chi2inv((1-alfa/2), (n-1));
S2 = var(X);
x1 = ((n-1)*S2/X2Inf);
x2 = ((n-1)*S2/X2Sup);
icvp = [x1, x2];
end
```