Ex.3.30

6 de dezembro de 2020 15:31

3.30. Numa Estação Florestal estudam-se problemas de intercepção de precipitação. Nesse sentido em 12 dias de chuva (suponha que se trata de uma amostra aleatória de dias com precipitação) são colocados dois udómetros para medir a quantidade de precipitação: um numa zona desarborizada e outro sob as copas das árvores. As leituras da quantidade de água em cada udómetro (medidas em cm de altura) deram os seguintes valores (cada coluna corresponde a um dia):

zona descoberta	5.87	1.30	2.34	2.82	5.89	9.09	1.93	9.27	4.65	4.35	5.00	8.43
sob coberto	4.96	1.14	2.24	2.26	4.75	7.83	1.86	8.85	4.17	3.65	4.08	7.99

a) Estime a precipitação média na zona desarborizada e na zona sob coberto.

b) Será admíssivel supor que a 95% de confiança as precipitações médias nos dois casos são iguais? Explicite as hipóteses necessárias à resolução desta questão.

c) Tendo em conta que, pela própria natureza do problema, o nível de precipitação sob coberto não deverá ser superior ao nível da correspondente precipitação a descoberto, indique um procedimento estatístico mais adequado para avaliar se existem diferenças significativas nos dois casos.

a) $\bar{x} = (5.87 + 1.30 + \cdots + 8.43) / 12 = 5.07$ cm

$\bar{y} = (4.96 + 1.14 + \cdots + 7.99) / 12 = 4.48$ cm

b)

NESTE EXEMPLO, É FUNDAMENTAL COMPREENDER QUE AS DUAS AMOSTRAS (X-ZONA DESCOBERTA, E Y-SOB COBERTO) SÃO EMPARELHADAS, POIS OS 12 DIAS SÃO OS MESMOS (NOTE-SE QUE EXISTE UMA ELEVADA CORRELAÇÃO ENTRE X E Y).

NESSE CASO, EM VEZ DE CONSIDERAR X E Y SEPARADAMENTE, CONSIDERA-SE UMA NOVA V.A. QUE É A DIFERENÇA $D = X - Y$.

NO FUNDO, TRABALHA-SE ENTÃO COM UMA ÚNICA V.A. AS OBSERVAÇÕES PARA A V.A. D SÃO OBTIDAS A PARTIR DOS $\{x_1, \ldots, x_n\}$ E DOS $\{y_1, \ldots, y_n\}$ FAZENDO $\{d_1 = x_1 - y_1, d_2 = x_2 - y_2, \ldots, d_n = x_n - y_n\}$.

zona descoberta	5.87	1.30	2.34	2.82	5.89	9.09	1.93	9.27	4.65	4.35	5.00	8.43
sob coberto	4.96	1.14	2.24	2.26	4.75	7.83	1.86	8.85	4.17	3.65	4.08	7.99
DIFERENÇA:	0.91	0.26	0.1	0.56	0.14	1.26	0.07	0.42	0.48	0.7	0.92	0.44

com $\bar{d} = 0.5967$ cm e $s_d = 0.397$ cm

D_i É A V.A. DIFERENÇA ENTRE PP NA ZONA DESCOBERTA E SOB COBERTO NO DIA i

SUPÕE-SE QUE $D \cap N(\mu_D, \sigma_D)$

É POSSÍVEL RESPONDER À ALÍNEA b) DE 2 FORMAS ALTERNATIVAS:

A - CONSTRUINDO UM I.C. PARA μ_D

OU

B - TESTANDO A HIPÓTESE $H_0: \mu_D = 0$

A:

μ_D diferença de médias	amostras emparelhadas; σ_D desconhec. e pop. das diferenças normal	$\bar{d} - t_{\alpha/2;\,(n-1)}\, s_D/\sqrt{n} < \mu_D < \bar{d} + t_{\alpha/2;\,(n-1)}\, s_D/\sqrt{n}$

O I.C. A 95% É, DADO QUE $t_{\alpha/2;(n-1)} = t_{0.025;(11)} = 2.2$

TABELA

	α							
n	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.32492	1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	318.30884
2	0.28868	0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	22.32712
3	0.27667	0.76489	1.63774	2.35336	3.18245	4.54070	5.84091	10.21453
4	0.27072	0.74070	1.53321	2.13185	2.77645	3.74695	4.60409	7.17318
5	0.26718	0.72669	1.47588	2.01505	2.57058	3.36493	4.03214	5.89343
6	0.26483	0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.20763
7	0.26317	0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	4.78529
8	0.26192	0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	4.50079
9	0.26096	0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.29681
10	0.26018	0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.14370
				ALÍNEA c)	ALÍNEA b)			
11	0.25956	0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.02470
12	0.25903	0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	3.92963

ASSIM, O I.C. A 95% PARA μ_D É

$$\left]\, 0.5967 - 2.2 \times \frac{0.397}{\sqrt{12}} \,,\, 0.5967 + 2.2 \times \frac{0.397}{\sqrt{12}} \,\right[$$

0.5967-2.2*0.397/sqrt(12)=0.34457113744489

0.5967+2.2*0.397/sqrt(12)=0.84882886255511

OU SEJA, COM 95% DE CONFIANÇA PODE AFIRMAR-SE QUE

$$0.344 < \mu_D < 0.848$$

E, POR ISSO, A MÉDIA DAS DIFERENÇAS ENTRE DESCOBERTO E SOB COBERTO NÃO É NULA.

B:

μ_D diferença de duas médias	amostras emparelhadas; pop.das diferenças normal e σ_D desconhec.	$\mu_D = 0$ $\mu_D \le 0$ $\mu_D \ge 0$	$\mu_D \neq 0$ $\mu_D > 0$ $\mu_D < 0$	$T = \dfrac{\overline{D}}{S_D/\sqrt{n}}$ c/ $(n-1)$ g.l.	$T < -t_{\alpha/2}$ ou $T > t_{\alpha/2}$ $T > t_\alpha$ $T < -t_\alpha$

NESTE CASO SEGUEM-SE OS PASSOS DO TESTE À HIPÓTESE DA MÉDIA DA

Passos a seguir na construção de um **teste estatístico**

1. Identificar o(s) parâmetro(s); especificar H_0 e H_1 e o nível de

NESTE CASO SEGUEM-SE OS PASSOS DO TESTE À HIPÓTESE DA MÉDIA DA DIFERENÇA SER IGUAL OU DIFERENTE DE ZERO

Passos a seguir na construção de um **teste estatístico**

1. Identificar o(s) parâmetro(s); especificar $\mathbf{H}_0$ **e** $\mathbf{H}_1$ e o nível de significância α.
2. Escolher uma variável aleatória – **Estatística de Teste**, que sob $\mathbf{H}_0$ terá distribuição conhecida (pelo menos aproximadamente).
3. Definir a **região de rejeição** ou **região crítica** – **RC** (conjunto de valores da estatística que são menos "plausíveis" caso $\mathbf{H}_0$ seja verdadeira, portanto levam a rejeitar $\mathbf{H}_0$).
4. Calcular **o valor** da estatística de teste, para a amostra observada.
5. Se o valor calculado $\in$ RC $\longrightarrow$ rejeita-se H_0
o valor calculado $\notin$ RC $\longrightarrow$ não se rejeita H_0

1. O PARÂMETRO É μ_D.

$$\underbrace{H_0 : \mu_D = 0}_{\text{TEM O BENEFÍCIO DA DÚVIDA}} \quad vs \quad H_1 : \underbrace{\mu_D \neq 0}_{\text{TEM O ÓNUS DA PROVA}}$$

2. A ESTATÍSTICA DE TESTE É $T = \dfrac{\overline{D}}{S_D/\sqrt{n}}$

SOB H_0, $T = \dfrac{\overline{D}}{S_D/\sqrt{n}} \cap t_{(n-1)}$

3. A REGIÃO CRÍTICA É DADA POR $T < -t_{\alpha/2}$ **ou** $T > t_{\alpha/2}$
OU SEJA $]-\infty, -2.2[\cup]2.2, +\infty[$, POIS $t_{\alpha/2(n-1)} = 2.2$ (VER ACIMA).

4. DADO QUE $\bar{d} = 0.5967$ cm e $s_d = 0.397$ cm,
O VALOR CALCULADO DE T É $T_{CALC} = \dfrac{0.5967}{0.397/\sqrt{12}} = 5.2$

0.5967/(0.397/sqrt(12))=5.206623258822916

5. DADO QUE $T_{CALC} = 5.2$ PERTENCE À REGIÃO CRÍTICA, REJEITA-SE H_0 A UM NÍVEL DE SIGNIFICÂNCIA $\alpha = 0.05$ E CONCLUI-SE QUE EXISTE UMA DIFERENÇA SIGNIFICATIVA ENTRE "DESCOBERTO" E "SOB COBERTO".

c) Tendo em conta que, pela própria natureza do problema, o nível de precipitação sob coberto não deverá ser superior ao nível da correspondente precipitação a descoberto, indique um procedimento estatístico mais adequado para avaliar se existem diferenças significativas nos dois casos.

NESTE CASO, APENAS É POSSÍVEL RESPONDER À QUESTÃO COM UM TESTE DE HIPÓTESE, POIS TRATA-SE DE UM TESTE UNILATERAL:

μ_D diferença de duas médias	amostras emparelhadas; pop.das diferenças normal e σ_D desconhec.	$\mu_D = 0$	$\mu_D \neq 0$	$T = \dfrac{\overline{D}}{S_D/\sqrt{n}}$	$T < -t_{\alpha/2}$ ou $T > t_{\alpha/2}$
		$\mu_D \leq 0$	$\mu_D > 0$	c/ $(n-1)$ g.l.	$T > t_\alpha$
		$\mu_D \geq 0$	$\mu_D < 0$		$T < -t_\alpha$

É NECESSÁRIO DECIDIR SE $H_0: \mu_D \leq 0$ OU $H_0: \mu_D \geq 0$

NOTE-SE QUE H_0 É A HIPÓTESE QUE TEM O BENEFÍCIO DA DÚVIDA. REJEITAR H_0 REQUER MAIOR EVIDÊNCIA EXPERIMENTAL DO QUE SIMPLESMENTE ADMITIR H_0.

COMO A MÉDIA OBSERVADA É $\bar{d} = 0.5967$, ENTÃO ADMITE-SE TRIVIALMENTE A HIPÓTESE $\mu_D \geq 0$.

A HIPÓTESE INTERESSANTE A TESTAR É ENTÃO $H_0: \mu_D \leq 0$. SE SE REJEITAR $H_0: \mu_D \leq 0$, CONCLUI-SE COM EVIDÊNCIA SIGNIFICATIVA QUE $\mu_D > 0$.

O TESTE É ENTÃO O SEGUINTE:

1. TESTAR PARÂMETRO μ_D. $H_0: \mu_D \leq 0$ vs $H_1: \mu_D > 0$

TEM O BENEFÍCIO DA DÚVIDA — TEM O ÓNUS DA PROVA

Passos a seguir na construção de um **teste estatístico**

1. Identificar o(s) parâmetro(s); especificar $\mathbf{H}_0$ **e** $\mathbf{H}_1$ e o nível de significância α.
2. Escolher uma variável aleatória – **Estatística de Teste**, que sob $\mathbf{H}_0$ terá distribuição conhecida (pelo menos aproximadamente).
3. Definir a **região de rejeição** ou **região crítica** – **RC** (conjunto de valores da estatística que são menos "plausíveis" caso $\mathbf{H}_0$ seja verdadeira, portanto levam a rejeitar $\mathbf{H}_0$).
4. Calcular **o valor** da estatística de teste, para a amostra observada.
5. Se o valor calculado $\in$ RC $\longrightarrow$ rejeita-se H_0
   o valor calculado $\notin$ RC $\longrightarrow$ não se rejeita H_0

2. A ESTATÍSTICA DE TESTE É $T = \dfrac{\overline{D}}{S_D/\sqrt{n}}$

SOB H_0, $T = \dfrac{\overline{D}}{S_D/\sqrt{n}} \cap t_{(n-1)}$

3. DADO QUE $\alpha = 0.05$, $t_{\alpha;(n-1)} = t_{0.05;(11)} = 1.7958$

A REGIÃO CRÍTICA É ENTÃO $]1.7958, +\infty[$

NOTA: SE O VALOR OBSERVADO PARA T FOR MUITO ELEVADO É NATURAL QUE REJEITE $H_0: \mu_D \leq 0$.

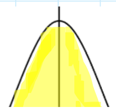
4. DADO QUE $\bar{d} = 0.5967$ cm e $s_d = 0.397$ cm,

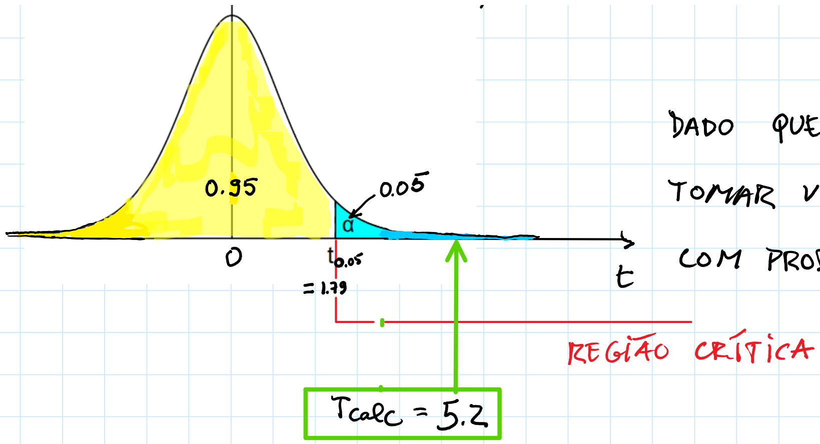
O VALOR CALCULADO DE T É $T_{CALC} = \dfrac{0.5967}{0.397/\sqrt{12}} = 5.2$

0.5967/(0.397/sqrt(12))=5.206623258822916

5. DADO QUE $T_{CALC} = 5.2$ PERTENCE À REGIÃO CRÍTICA, REJEITA-SE H_0 A UM NÍVEL DE SIGNIFICÂNCIA $\alpha = 0.05$ E CONCLUI-SE QUE $\mu_D > 0$, OU SEJA QUE A PRECIPITAÇÃO A "DESCOBERTO" É SIGNIFICATIVAMENTE SUPERIOR À PRECIPITAÇÃO "SOB COBERTO".

FUNÇÃO DENSIDADE DA V.A T SOB $H_0: \mu_D \leq 0$.





DADO QUE SOB H_0, T DEVERIA TOMAR VALORES ABAIXO DE 1.79 COM PROB. 0.95, ENTÃO É POUCO PROVÁVEL, SOB H_0, QUE T_{calc} ESTEJA NA R. CRÍTICA.

POR ISSO REJEITA-SE H_0 E CONCLUI-SE H_1.