ICMC – USP – SME122
Introdução à Inferência Estatística
2ª PROVA – 2º/2007 – 13/12/2007

1. Dois fornecedores (A e B) fabricam um certo componente utilizado em impressoras. A resistência ao impacto é uma característica importante. Segundo afirmação do fornecedor B, as peças por ele produzidas têm maior resistência média ao impacto. Uma amostra aleatória de 10 componentes produzidos pelo fornecedor A resultou em uma média de 290 unidades e desvio padrão de 12 unidades, ao passo que de uma amostra de 16 peças produzidas pelo fornecedor B obteve-se média igual a 321 unidades e desvio padrão igual a 22 unidades. Os dados corroboram a afirmação do fornecedor B?

   Solução. Supomos que a resistência é uma variável aleatória com distribuição normal. Pelo enunciado, as duas amostras são independentes (fornecedores diferentes). Representamos as médias e as variâncias (desconhecidas e diferentes) da resistência por μ_A, σ_A^2, μ_B e σ_B^2, conforme o fornecedor. Ainda segundo o enunciado, $n_A = 10$, $\overline{X}_A = 290$, $s_A = 12$, $n_B = 16$, $\overline{X}_B = 321$, $s_B = 22$. A afirmação do fornecedor B pode ser verificada testando $H_0 : \mu_A = \mu_B$ contra $H_1 : \mu_A < \mu_B$. Calculamos a estatística de teste

   $$t_0 = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{s_A^2/n_A + s_B^2/n_B}} = -4,639$$

   com graus de liberdade estimados (após arredondamento) por

   $$g = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{\frac{s_A^4}{n_A^2(n_A-1)} + \frac{s_B^4}{n_B^2(n_B-1)}} = 24.$$

   Adotando um nível de significância de 5%, obtemos o valor crítico $t_c = -1,711$. Rejeitamos H_0, pois $t_0 < t_c$. Logo, a um nível de significância de 5%, os dados corroboram a afirmação do fornecedor B.

2. Os resultados abaixo foram obtidos em um experimento em que um mesmo grupo de pessoas realizou tarefas típicas em duas interfaces diferentes de um mesmo programa computacional. Os valores na tabela são os tempos de realização das tarefas (metade dos indivíduos usou primeiro a interface 1 e depois a interface 2).

Indivíduo	Tempo de realização (s)	
	Interface 1	Interface 2
1	37,0	17,8
2	25,8	20,2
3	16,2	16,8
4	24,2	41,4
5	22,0	21,4
6	33,4	38,4
7	23,8	16,8
8	58,2	32,2
9	33,6	27,8
10	24,4	23,2
11	21,2	20,6
12	36,2	32,2
13	29,8	53,8
14	23,4	29,6

Baseando-se nos dados acima, podemos concluir que as duas interfaces apresentam o mesmo grau de dificuldade para a realização das tarefas?

Solução. Supomos que o tempo de realização é uma variável aleatória com distribuição normal. Pelo enunciado, as duas amostras são dependentes, pois os mesmos indivíduos foram utilizados nas duas etapas da coleta de dados ($n = 14$ pares de observações). Representamos por D a diferença entre os tempos de realização com as duas interfaces. A pergunta formulada pode ser respondida testando $\mathrm{H}_0 : \mu_D = 0$ contra $\mathrm{H}_1 : \mu_D \neq 0$. Calculamos $\overline{D} = \sum_{i=1}^{n} D_i/n = 1,21\,s$, $s_D = \{\sum_{i=1}^{n}(D_i - \overline{D})^2/(n-1)\}^{1/2} = 12,68\,s$, $t_0 = \sqrt{n}\overline{D}/s_D = 0,358$. Com $n-1 = 13$ graus de liberdade e com um nível de significância de 5%, obtemos $t_c = 2,160$. E como $|\ t_0\ | < t_c$, não rejeitamos H_0. De acordo com os dados e adotando um nível de significância de 5%, concluímos que as duas interfaces apresentam o mesmo grau de dificuldade para a realização das tarefas.

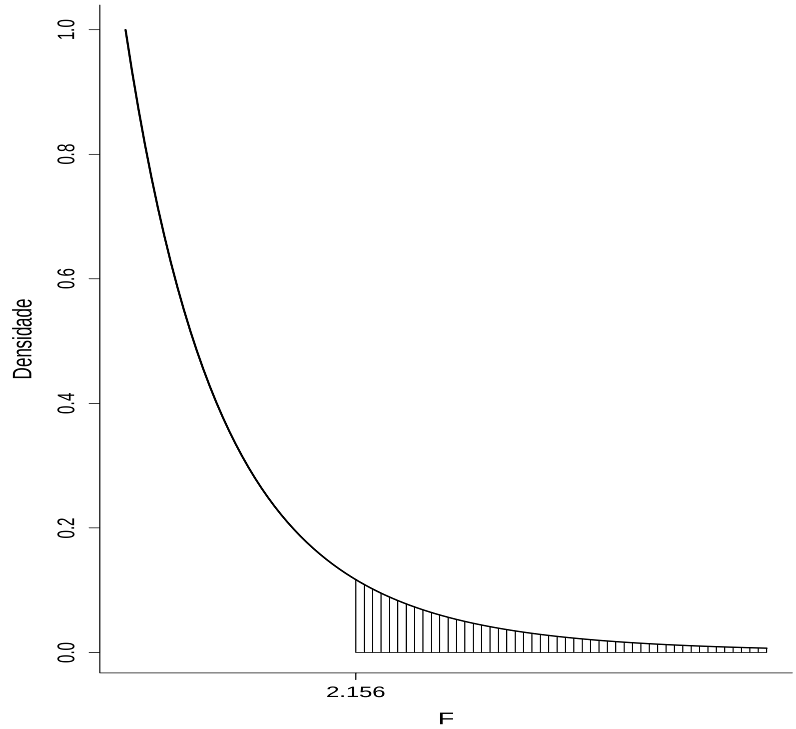
3. Analistas de uma loja de departamentos estão interessados em verificar se existem diferenças entre as quantias médias faturadas entre três diferentes formas de pagamento: dinheiro (D), cheque (C) e cartão de crédito (CC). Um levantamento de vendas (em milhares de R$) em um certo período forneceu os dados abaixo:

	Forma de pagamento		
	D	C	CC
	56,00	80,90	73,25
	20,50	51,29	56,65
Vendas	37,37	40,95	123,21
(1.000 R$)	28,64	72,65	56,50
		132,47	37,29
		60,32	44,65
		60,00	40,64

(a) Complete a tabela abaixo e responda à questão de interesse.

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	F
Entre formas	$k-1=\mathbf{2}$	$14677,6 - 11400,7 = \mathbf{3276,9}$	$\frac{3276,9}{2} = \mathbf{1638,45}$	$\frac{1638,45}{760,05} = \mathbf{2,156}$
Intraformas	$n-k=\mathbf{15}$	11400,7	$\frac{11400,7}{15} = \mathbf{760,05}$	
Total	$n-1=\mathbf{17}$	14677,6		

Solução. Supomos que o faturamento é uma variável aleatória com distribuição normal. Denotando por μ_{D}, μ_{C} e μ_{CC} os faturamentos médios nas três formas de pagamento, a questão formulada pode ser respondida testando a hipótese $\mathrm{H}_0 : \mu_{\mathrm{D}} = \mu_{\mathrm{C}} = \mu_{\mathrm{CC}}$. De acordo com os dados fornecidos, temos $k = 3$ formas de pagamento, $n_{\mathrm{D}} = 4$, $n_{\mathrm{C}} = 7$, $n_{\mathrm{CC}} = 7$ e $n = n_{\mathrm{D}} + n_{\mathrm{C}} + n_{\mathrm{CC}} = 18$. Na tabela acima obtemos $F_0 = 2,156$. A um nível de significância de 5%, o valor crítico $f_c = 3,68$ é encontrado na tabela da distribuição F com 2 e 15 graus de liberdade. Como $F_0 < f_c$, não rejeitamos H_0. Portanto, com 5% de significância, os dados permitem concluir que não há diferença entre as quantias médias faturadas nas três formas de pagamento.

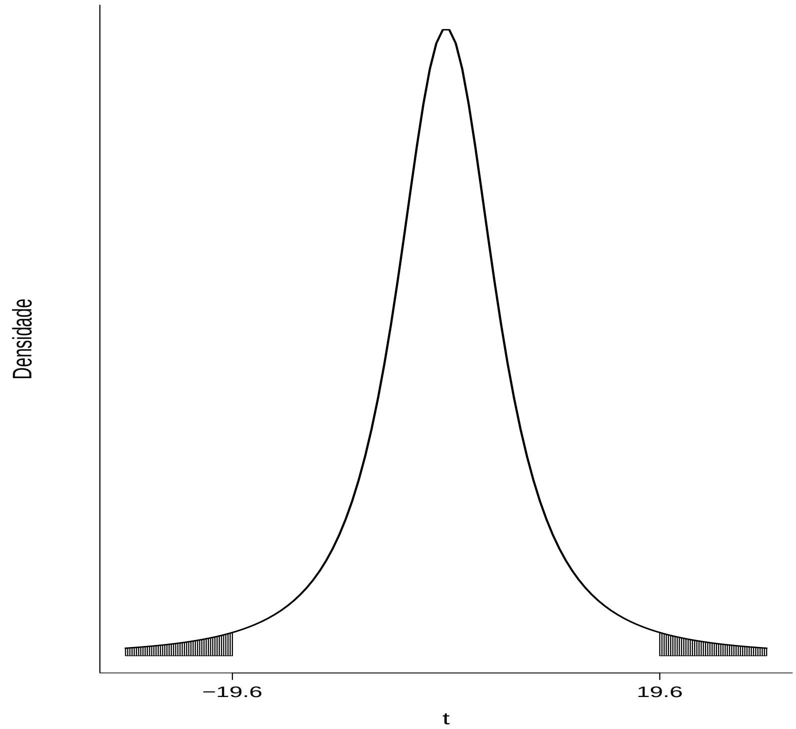


(b) Indique como obter graficamente o nível descritivo do valor de F calculado no item (a). Solução. O nível descritivo está indicado pela área hachurada no gráfico acima.

4. Resultados anteriores indicam que um método comumente utilizado para aliviar um certo tipo de dor é efetivo em 60% dos pacientes tratados. Resultados experimentais com um novo tratamento aplicado a 120 pacientes mostraram que 85 pacientes sentiram alívio da dor. É possível concluir que o novo método é superior ao método já em uso?

   Solução. O parâmetro a ser testado é a proporção de pacientes que sentiram alívio da dor ao utilizarem o novo tratamento, denotada por θ. A pergunta formulada menciona a superioridade do novo método em relação a um tratamento comum (efetivo em 60% dos pacientes tratados; ou seja, $\theta_0 = 0,6$). Esta pergunta pode ser respondida testando $\mathrm{H}_0 : \theta = 0,6$ contra $\mathrm{H}_1 : \theta > 0,6$. De acordo com os dados, a proporção amostral pode ser estimada por $\widehat{\theta} = 85/120 = 0,708$. Aproximamos a distribuição da proporção amostral pela distribuição normal. Adotamos um nível de signficância de 5%, de modo que $z_c = 1,645$ (tabela da distribuição t de Student). Calculamos $z_0 = \sqrt{n}(\widehat{\theta} - \theta_0)/\sqrt{\theta_0(1-\theta_0)} = 2,42$. Rejeitamos H_0, pois $z_0 > z_c$. Assim, com 5% de signficância, concluímos a partir dos dados que o novo método é superior ao método já em uso.

5. Uma máquina em uma fábrica de bebidas é ajustada para liberar uma certa quantidade de líquido. Uma amostra de 25 garrafas envasadas resultou em um volume médio de $32,4$ cm^3 e um desvio padrão de $0,74$ cm^3. (a) Os dados confirmam o argumento de que a quantidade média de líquido envasado é $29,5$ cm^3? (b) Indique como obter graficamente o nível descritivo correspondente a sua resposta no



item (a).

SOLUÇÃO. Supomos que a quantidade de líquido envasado é uma variável aleatória com distribuição normal (representada por X). De uma amostra com $n = 25$ garrafas, obtivemos média amostral $\overline{X} = 32,4$ cm^3 e desvio padrão amostral $s = 0,74$ cm^3. A pergunta do item (a) pode ser respondida testando $\mathrm{H}_0 : \mu = 29,5$ contra $\mathrm{H}_1 : \mu \neq 29,5$ ($\mu_0 = 29,5$). Calculamos $t_0 = \sqrt{n}(\overline{X} - \mu_0)/s = 19,59$, valor excessivamente alto. Com $n - 1 = 24$ graus de liberdade e um nível de signficância usual (5% ou 1%), temos $\mid t_0 \mid > t_c$, de forma que H_0 é rejeitada. Logo, os dados coletados refutam o argumento de que a quantidade média de líquido envasado é $29,5$ cm^3. O nível descritivo corresponde á area da região hachurada no gráfico acima (notando que a figura está sem escala).