Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Probabilidade e Estatística 2
1^o semestre de 2020
Gustavo L. Gilardoni

Lista de Exercícios 1

1. Para avaliar a diferença $\mu_A - \mu_B$ entre as médias de duas populações uma pesquisadora colheu uma amostra de 400 indivíduos da população com média μ_A e outra de tamanho 100 da população com média μ_B. As amostras forneceram médias $\bar{x}_A = 18.93$ e $\bar{x}_B = 17.00$ e desvios padrões $s_A = 12.00$ e $s_B = 8.00$ respectivamente.

(i) Construa um IC com nível 95% para $\mu_A - \mu_B$; (ii) Teste ao nível de significância 5% se $\mu_A = \mu_B$ e calcule o p-valor.

2. Uma pesquisadora deseja comparar duas dietas para emegrecer, digamos A e B. Para isso, um grupo de $n = 100$ pacientes foi alocado ao acaso a cada uma das duas dietas. Entre os 50 pacientes que seguiram a dieta A por um período de 30 dias, constatou-se uma redução média de peso corporal de 4.2kg com um desvio padrão de 1.2kg, enquanto para os 50 pacientes que seguiram a dieta B, a perda média e o desvio padrão foram 2.8 e 0.7kg respectivamente.

(i) Existe evidência que os pacientes da dieta A perdem em média pelo menos um kg a mais do que os da dieta B? Use $\alpha = 0.05$; (ii) Calcule o p-valor do teste; (iii) Construa um IC de nível 95% para a diferença média de perda de peso entre os pacientes da dieta A e os da B e (iv) Se a pesquisadora está planejando realizar outro experimento semelhante a este também com 100 pacientes e com as mesmas dietas, você recomendaria que ela alocasse 50 pacientes a cada dieta? Explique.

3. Para estudar o efeito de um aditivo para gasolina uma montadora escolheu ao acaso 14 carros da linha de montagem que circularam com e sem o aditivo numa pista de testes por 300km (a ordem na qual usaram e não usaram o aditivo foi também determinada ao acaso). O consumo em litros é apresentado na seguinte tabela.

Carro	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Sem	22.3	22.8	22.9	22.1	25.9	20.9	21.0	22.8	19.0	22.1	20.4	22.6	20.5	22.7
Com	20.5	21.9	21.7	20.2	22.2	20.4	19.9	20.1	19.4	19.6	19.4	20.4	20.5	19.9

A montadora deve recomendar o uso do aditivo se ele poupar em média pelo menos um litro de gasolina para o percurso de 300km.

(i) Ao nível de significância de 5%, a montadora deve recomendar o uso do aditivo? (ii) Calcule o p-valor do teste e (iii) construa um intervalo com 95% de confiança para a redução no consumo devido ao uso do aditivo.

4. Um grupo de 100 estudantes foi submetido a um teste de resistência física. Para cada estudante a freqüência cardíaca foi registrada antes e depois do teste. O objetivo do experimento era verificar se o aumento da freqüência cardíaca devido a realização do teste não supera em média 20 batimentos por minuto. Antes do teste, a freqüência cardíaca dos 100 estudantes teve média de 77 e desvio padrão de 10 batimentos por minuto. Após o teste, a freqüência média passou para 96 enquanto o desvio padrão se manteve em 10 batimentos por minuto. O aumento na freqüência cardíaca, calculado para cada estudante como a diferença entre os batimentos após e antes do teste, teve média de 19 e desvio padrão de 4 batimento por minuto.

(i) Teste, ao nível de significância de 1%, se o aumento médio da frequência cardíaca devido ao teste é inferior a 20 batimentos; (ii) Calcule o p-valor e (iii) Construa um IC com nível 99% para o aumento dos batimentos.

5. Num estudo visando estudar o efeito de um analgésico, 200 pacientes com enxaqueca foram alocados aleatoriamente para ser tratados com o analgésico e com um placebo. Dos 100 pacientes tratados com o analgésico, 65 manifestaram ter melhorado após uma hora, enquanto dos outros 100 pacientes tratados com o placebo, somente 50 sentiram melhoras após uma hora.

(i) Construa um intervalo com confiança 95% para a diferença das proporções de pacientes que melhoram com o analgésico e com o placebo; (ii) teste, ao nível de significância 5%, se o analgésico é melhor do que o placebo.

6. Faz quase 150 anos, Charles Darwin suspeitou que plantas obtidas mediante fecundação cruzada deviam ser mais fortes do que aquelas obtidas por autofecundação (Darwin, *The effects of cross and self-fertilisation in the vegetable kingdom*, Londres, 1876).

Para estudar este problema ele usou 15 mudas obtidas por autofecundação e 15 obtidas por fecundação cruzada e plantou elas em 15 vasos de forma que cada vaso recebeu uma muda de cada tipo. Conseqüentemente, ele pode assumir que as duas mudas plantadas no mesmo vaso receberam a mesma quantidade de nutrientes, luz, etc. Após um certo tempo, ele mediu a altura das plantas. Os dados coletados por Darwin, com as alturas medidas em polegadas, estão na seguinte tabela.

Vaso	1	2	3	4	5	6	7	8
Cruzada	23.5	12.0	21.0	22.0	19.1	21.5	22.2	20.2
Auto	17.4	20.4	20.0	20.0	18.2	18.8	18.8	15.1
Dif.	6.1	-8.4	1.0	2.0	0.9	2.7	3.4	5.1

Vaso	9	10	11	12	13	14	15	Média	D.Padrão
Cruzada	18.2	21.8	23.2	21.0	21.1	23.0	12.1	20.12	3.58
Auto	16.5	18.0	16.1	18.0	11.8	15.5	18.1	17.51	2.25
Dif.	1.7	3.8	7.1	3.0	9.3	7.5	-6.0	2.61	4.72

Para ajudar na análise estatística Darwin recorreu ao seu amigo e meio-primo Francis Galton (que, alguns anos depois, inventaria o termo *Regressão*, que vamos estudar na Unidade 3).

(i) O que você acha do experimento delineado por Darwin? (Considere que foi realizado em 1876, muito antes do desenvolvimento da maioria das técnicas estudadas no nosso curso); (ii) Imagine que você é Francis Galton. Como você ajudaria o seu meio-primo?

7. Uma produtora rural quer saber qual de duas variedades de tomates plantar. Para isso preparou 8 mudas de cada variedade que foram plantadas em condições semelhantes. Após a colheita, o rendimento (em kg) de cada planta foi o seguinte:

Var. A	10.6 7.3 10.0 9.2 8.7 7.2 6.3 10.2
Var. B	8.4 9.9 7.3 9.2 5.9 7.6 7.9 8.0

Supondo que o rendimento das duas variedades é normalmente distribuído com variâncias iguais, (i) teste, ao nível de significância $\alpha = 0.10$, se a variedade A oferece maior rendimento em média do que a B e (ii) Construa um IC com nível 90% para a diferença média entre os rendimentos.

8. Se U é uma variável aleatória com distribuição χ^2 (qui-cuadrado) com 15 graus de liberdade, calcule (i) $P(U < 6.262)$ e $P(U > 27.488)$; (ii) Dois números a e b tais que $P(U < a) = P(U > b) = 0.05$.

9. O seguinte resultado foi visto na aula: *Se $Z_1 \dots, Z_n$ é uma amostra aleatória da distribuição Normal Padrão, então $U = \sum_{i=1}^{n} U_i^2$ segue uma distribuição χ^2 com n graus de liberdade.* Use esse resultado para mostrar que, quando $n \to \infty$, a distribuição de $\frac{U-n}{\sqrt{2n}}$ tende à distribuição Normal Padrão.

10. Uma amostra de $n = 20$ tambaquis apresentou concentração média de mercúrio de $\bar{x} = 0.182$mg/kg e desvio padrão de $s = 0.067$mg/kg.

Calcule ICs para a média μ e o desvio padrão σ de essa população. Use $\alpha = 0.10$ nos dois casos.

11. Seja F uma variável aleatória com distribuição F com $m = 8$ gl no numerador e $n = 10$ gl no denominador.

Ache (i) $P(F < 0.299)$ e $P(F > 3.072)$; (ii) Dois números a e b tais que $P(F < a) = P(F > b) = 0.025$.

12. *Sejam U_m e V_n duas variáveis aleatórias independentes tais que $U_m \sim \chi^2_m$ e $V_n \sim \chi^2_n$. Então a variável aleatória $F_{m,n} = \frac{U/m}{V/n}$ segue uma distribuição F de Snedecor com m gl no numerador e n gl no denominador.*

Use o resultado anterior para mostrar que $P(F_{m,n} < a) = P(F_{n,m} > a^{-1})$.

13. No Exercício 7, teste ao nível $\alpha = 0.05$ se as variâncias das duas populações são iguais. Construa um IC de nível 90% para a razão entre essas variâncias.

Soluções

1. (i) $-0.030 < \mu_A - \mu_B < 3.890$; (ii) $z_{obs} = 1.93$, $z_{0.975} \cdot 1.96$, não rejeita H_0, p-valor $= 0.054$.

2. (i) $z.obs = 2.036 > 1.645$, logo rejeita H_0. (ii) 0.021. (iii) $1.015 < \mu_A - \mu_B < 1.785$. (iv) usando o resultado atual como amostra piloto, recomendaríamos alocar aproximadamente 63 à dieta A e 37 à dieta B.

3. (i) Queremos testar H_a : $\mu_{com} - \mu_{sem} < -1$. Logo $t_{obs} = \sqrt{14}[\bar{d} + 1]/s_d \doteq \sqrt{14}(-1.564 + 1)/1.156 \doteq 1.826$. O valor crítico é $-t_{13;0.05} = -1.771$ e como $t_{obs} < -1.771$, rejeitamos a H_0 (i.e. ao nível 5% existe evidência que o uso do aditivo poupa em média pelo menos 1 litro de gasolina para o percurso de 300km). (ii) $P(t_{13} < -1.826) = 0.045$. (iii) O IC para $\mu_d = \mu_{com} - \mu_{sem}$ é $\bar{d} \pm t_{13;0.975}\, s_d/\sqrt{14} \doteq -1.564 \pm (2.160)\, 1.156/\sqrt{14} \doteq (-2.232; -0.897)$ litros (i.e. com 95% de confiança, o aditivo poupa entre 0.897 e 2.232 litros).

4. (i) $z_{obs} = (10)\,(19 - 20)/4 = -2.5$, $-z_{0.99} \doteq -2.326$, logo rejeitamos H_0; (ii) p-valor $= 0.006$ ou 0.6%; (iii) $17.970 < \mu_d < 20.030$.

5. (i) $(0.015; 0.285)$ ou com 95% de confiança entre 1.5 e 28.5% mais pacientes melhoram com o analgésico quando comparado com o placebo; (ii) $\hat{p}_c = 0.575$, $z_{obs} = 2.170$, rejeitamos a hipótese H_0 de igualdade das proporções (i.e. existe evidência que o analgésico é melhor do que o placebo); o p-valor do teste é 0.015 ou 1.5%.

6. (i) O experimento foi delineado exatamente da mesma forma que um pesquisador moderno faria. É fantástico como, baseado na intuição, Darwin fez o experimento certo, sem ter acesso a toda a teoria moderna do delineamento experimental. Uma observação, porém, é que na realização do experimento o desvio padrão das diferenças (4.72) é consideravelmente maior do que o das medições individuais (3.58 e 2.25), que é o contrário do que esperaríamos num experimento pareado. (ii) Em terminologia moderna, Darwin queria testar se $\mu_d = \mu_{cruzada} - \mu_{auto} > 0$. Para esse teste, o estatístico é $t_{obs} = 2.146$ e o valor crítico para $\alpha = 0.05$ é 1.76, de forma que para esse α existe evidência que as mudas com fecundação cruzada crescem mais do que as auto auto fecundadas. O p-valor do teste é 0.025 ou 2.5%.

7. (i) $s_p \doteq 1.415$, $t_{obs} = 0.936$, $t_{14,0.1} = 1.345$, não rejeitamos a H_0 de igualdade de rendimentos médios; (ii) $-0.584 < \mu_A - \mu_B < 1.909$.

13. Temos que $s_A^2 \doteq (1.593)^2$ e $s_B^2 \doteq (1.212)^2$. Assim, o estatístico do teste é $F_{obs} = s_B^2/s_A^2 \doteq 0.578$. Como o teste é bilateral, os valores críticos são $F_{7,7,0.025} \doteq 0.200$ e $F_{7,7,0.9755} \doteq 4.995$. Logo, não rejeitamos a H_0 de igualdade de variâncias. Como $P(F_{7,7} < F_{obs}) \doteq 0.243 < 0.757 \doteq P(F_{7,7} > F_{obs})$, o p-valor do teste é $2(0.243) \doteq 0.486$. (ii) O IC com 90% de confiança para a razão σ_B^2/σ_A^2 é $[(s_B^2/s_A^2)/F_{7,7,0.95}; (s_B^2/s_A^2)/F_{7,7,0.05}] \doteq (0.578/3.787; 0.578/0.264) \doteq (0.153; 2.190)$.