DEST
Departamento de Estatística

Estatística Inferencial - Semana 3
Distribuição amostral

1. Um experimento genético, envolve uma população de moscas de frutas que consiste em 1 macho (Mike) e 3 fêmeas, chamadas Ana, Bárbara e Cristina. Suponha que duas moscas de frutas sejam selecionadas aleatoriamente *com reposição*.
   a) Depois de listar as 16 diferentes amostras possíveis, ache a proporção de fêmeas em cada amostra e, então, use uma tabela para descrever a distribuição amostral da proporção de fêmeas.
   b) Ache a média da distribuição amostral.
   c) A média da distribuição amostral (item b) é igual à proporção populacional de fêmeas?

2. As idades (anos) dos quatro presidentes dos Estados Unidos quando foram assassinados no exercício do cargo são 56 (Lincoln), 49 (Garfield), 58 (McKinley) e 46 (Kennedy).
   a) Supondo que duas das idades sejam selecionadas com reposição, liste as 16 diferentes amostras possíveis.
   b) Ache a média de cada uma das 16 amostras e, então, resuma a distribuição amostral das médias no formato de uma tabela que represente uma distribuição probabilidade.
   c) Compare a média populacional com a média das médias amostrais.

3. Repita o Exe 2 usando a mediana em lugar de médias.

4. Considere o seguinte problema (adaptado de Magalhães & Lima, 2006):
   Um fabricante afirma que sua vacina contra gripe imuniza em 80% dos casos. Uma amostra de 25 indivíduos entre os que tomaram a vacina foi sorteada e testes foram feitos para verificar a imunização ou não desses indivíduos.
   a) No contexto do problema identifique:
      - a população,
      - o parâmetro de interesse,
      - o estimador,
      - a estimativa,
      - a distribuição amostral.
   b) Se o fabricante estiver correto, qual é a probabilidade da proporção de imunizados na amostra ser inferior a 0,75? E superior a 0,85?

5. Uma variável aleatória Y tem distribuição normal, com média 100 e desvio padrão 10.
   a) Qual a $P(90 < Y < 110)$?
   b) Se $\bar{Y}$ for a média de uma amostra de 16 elementos retirados dessa população, calcule $P(90 < \bar{Y} < 110)$.

6. A máquina de empacotar um determinado produto o faz segundo uma distribuição normal, com média μ e desvio padrão 10 g.
   a) Em quanto deve ser regulado o peso médio μ para que apenas 10% dos pacotes tenham menos do que 500 g?

b) Com a máquina assim regulada, qual a probabilidade de que o peso total de 4 pacotes escolhidos ao acaso seja inferior a 2 kg?

7. Utilizando algum recurso computacional ou tabela, calcule as probabilidades a seguir, conforme a distribuição da v.a. Y:

$Y \sim t_{20}$	$Y \sim \chi_{16}$	$Y \sim F_{(10,7)}$
$P(-2,85 \leq Y \leq 2,85)$	$P(8,91 < Y < 32,85)$	$P(Y > 3,18)$
$P(Y < -2,85)$	$P(Y > 8,91)$	$P(Y > 0,15)$
$P(Y > 2,85)$	$P(Y > 32,85)$	$P(Y > 5,35)$
$P(Y > 2,12)$	$P(Y > 22,80)$	$P(Y < 7,41)$
$P(Y < -3,01)$	$P(Y < 10,12)$	$P(Y < 1)$

8. Para cada uma das 3 distribuições propostas no exercício anterior, encontre o valor de y tal que:
   a) $P(Y < y) = 0.90$
   b) $P(Y < y) = 0.025$
   c) $P(Y < y) = 0.01$
   d) $P(Y > y) = 0.975$
9. Um estudo que investiga a relação entre idade e despesas médicas anuais amostra aleatoriamente 100 indivíduos em uma cidade da Califórnia. Espera-se que a amostra tenha uma média de idade semelhante à de toda a população.
   a) Se o desvio padrão das idades de todos os indivíduos em Davis for $\sigma = 15$, encontre a probabilidade de que a idade média dos indivíduos da amostra esteja dentro de dois anos da idade média de todos os indivíduos na cidade. (Dica: encontre a distribuição amostral da idade média da amostra e use o teorema do limite central. Você não precisa saber a média da população para responder, mas se isso facilitar, use um valor como $\mu = 30$.)
   b) A probabilidade seria maior ou menor se $\sigma = 10$? Por quê?
10. O teste de conhecimentos gerais chamado *Graduate Record Examination* (GRE) tem componentes que medem o raciocínio verbal e o raciocínio quantitativo. O exame verbal e o exame quantitativo têm cada um uma pontuação mínima de 200 e máxima de 800. Nos últimos anos, a pontuação total nos dois exames teve aproximadamente uma distribuição normal com uma média de cerca de 1050 e desvio padrão de cerca de 200.
    a) Qual a probabilidade de obter pontuação total
       (i) abaixo de 1200 e
       (ii) acima de 1200?
    b) Dos participantes do teste GRE que pontuaram acima de 1.200, qual proporção deles teve pontuação acima de 1.400?
    c) Um grupo de 25 alunos formou um grupo de estudos para se preparar para o GRE. Para eles, a média de suas 25 pontuações totais é 1200. Se eles fossem uma amostra aleatória dos alunos que estão fazendo o exame, explique por que isso teria sido um resultado muito incomum.
11. Para uma população normal com variância conhecida σ^2, responda as seguintes questões: (para resposta considere o arredondamento na terceira casa decimal e sempre que a resposta for porcentagem apresente o valor decimal entre 0 e 1)

a) Qual é o nível de confiança para o intervalo $\bar{y} - 2,14\,\sigma/\sqrt{n} \leq \mu \leq \bar{y} + 2,14\,\sigma/\sqrt{n}$
b) Qual é o nível de confiança para o intervalo $\bar{y} - 2,49\,\sigma/\sqrt{n} \leq \mu \leq \bar{y} + 2,49\,\sigma/\sqrt{n}$
c) Qual é o nível de confiança para o intervalo $\bar{y} - 1,85\,\sigma/\sqrt{n} \leq \mu \leq \bar{y} + 1,85\,\sigma/\sqrt{n}$

12. Considere a seguinte equação aplicada para obter um intervalo de confiança bilateral de $100(1-\alpha)\%$ para o parâmetro μ de uma distribuição normal com variância conhecida σ^2 a partir de uma amostra aleatória de n observações:

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}$$

em que $z_{\alpha/2}$ é o ponto superior da distribuição normal padrão que delimita à sua direita $\alpha/2$ de área.

Para as respostas considere 3 casas decimais.

a) Qual é o valor de $z_{\alpha/2}$ nessa equação que fornece 99% de confiança?
b) Qual é o valor de $z_{\alpha/2}$ nessa equação que fornece 95% de confiança?
c) Qual é o valor de $z_{\alpha/2}$ nessa equação que fornece 90% de confiança?

13. Considere a seguinte equação aplicada para obter um intervalo de confiança bilateral de $100(1-\alpha)\%$ para o parâmetro σ^2 de uma distribuiçãoo normal a partir de uma amostra aleatória de n observações:

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]$$

em que $\chi^2_{\alpha/2}$ e $\chi^2_{1-\alpha/2}$ são pontos da distribuição χ^2 com $n-1$ graus de liberdade, que delimita à sua direita $\alpha/2$ de área. Considerando uma amostra aleatória de 15 elementos:

(Para as respostas considere 3 casas decimais).

a) Qual é o valor de $\chi_{\alpha/2}$ nessa equação que fornece 99% de confiança?
b) Qual é o valor de $\chi_{\alpha/2}$ nessa equação que fornece 95% de confiança?
c) Qual é o valor de $\chi_{\alpha/2}$ nessa equação que fornece 90% de confiança?

14. Considere um estudo no qual deseja-se estimar a proporção de solicitações atendidas e resolvidas de uma central do usuário através de uma amostra aleatória simples.

a) Se a amostra for de 4000 solicitações, qual será a margem de erro (com confiança de 95%) para a estimação da proporção de resolvidas?
b) Para este mesmo tamanho de amostra, qual seria a confiança associada a uma margem de erro de $\pm 0,01$ (1 %)?
c) Qual deveria ser o tamanho da amostra para se obter a estimativa com uma margem de erro de 2,5% com 95% de confiança?
d) E para uma margem de erro de 3% com 99% de confiança?

15. Antes de uma eleição, um determinado partido está interessado em estimar a probabilidade p de eleitores favoráveis ao seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores eram favoráveis ao candidato.

a) Utilizando a informação da amostra piloto, determine o tamanho da amostra para que, com 0,8 de probabilidade, o erro cometido na estimação seja de no máximo 0,05.
b) Se na amostra final, com o tamanho obtido no item anterior, observou-se que 51% dos eleitores eram favoráveis ao candidato, construa um intervalo de confiança para p, com confiança de 95%.
c) Decidiu-se coletar uma amostra de tamanho 150. Qual o erro máximo (margem de erro) que cometemos com probabilidade de 0,95 e p igual ao dado no item anterior?
d) Qual a influência do tamanho da amostra na amplitude do intervalo de confiança, considerando fixados $p = 0,51$ e o nível de confiança em $0,95$?

16. Num grupo de pacientes, o nível de colestrol é uma variável aleatória X com distribuição Normal de média desconhecida e variância 64 (mg/ml)^2.

a) Para uma amostra de 46 indivíduos, que forneceu um nível médio de colesterol de 120 mg/ml, calcule o intevalo com 95% de confiança para a média populacional.
b) Para uma amostra de 100 indivíduos, que forneceu um nível médio de colesterol de 120 mg/ml, calcule o intevalo com 95% de confiança para a média populacional.
c) Para uma amostra de 150 indivíduos, que forneceu um nível médio de colesterol de 120 mg/ml, calcule o intevalo com 95% de confiança para a média populacional.
d) Qual a influência do tamanho da amostra na amplitude do intervalo de confiança, dado que fixamos $\sigma = 8$ e o nível de confiança em 0,95?

17. Um pesquisador está investigando o tempo de reação de um novo medicamento. Em sua pesquisa 20 pacientes foram sorteados ao acaso, receberam o medicamento e tiveram o seu tempo de reação anotado. Os dados coletados foram os seguintes (em minutos): 2,9; 3,4; 3,5; 4,1; 4,6; 4,7; 4,5; 3,8; 5,3; 4,9; 4,8; 5,7; 5,8; 5,0; 3,4; 5,9; 6,3; 4,6; 5,5; 6,2.

a) Obtenha um intervalo de confiança de 90% para a verdadeira média populacional.

b) Obtenha um intervalo de confiança (90% de confiança) para a variância dos tempos de reação.

18. Entre milhares de casos de pneumonia não tratada com sulfa, a porcentagem que desenvolveu complicações foi de 13%. Com o intuito de saber se o emprego da sulfa diminuiria essa porcentagem, 113 casos de pneumonia foram tratados com sulfapiridina e destes, 6 apresentaram complicações. Com base nesse resultado, o que você pode dizer sobre o emprego de sulfa na porcentagem de complicações em casos de pneumonia? Considere um nível de confiança de 90%. Justifique sua resposta.

19. A Leishmaniose Visceral é uma doença importante e que, se não for tratada corretamente, pode levar a óbito. Todo caso diagnosticado de Leishmaniose Visceral deve ser notificado às autoridades de saúde. Num estudo sobre o número de dias entre o início dos sintomas da Leishmaniose Visceral e a notificação do caso às autoridades, uma pesquisadora deseja estimar o número médio de dias entre os sintomas e a notificação usando um intervalo de 95% de confiança. Sabendo que ela gostaria que o erro de estimação fosse a metade do desvio-padrão do número de dias e supondo que o número de dias entre o início dos sintomas e a notificação tenha distribuição Gaussiana, responda: quantos casos de Leishmaniose Visceral, no mínimo, ela deve estudar?

20. A vida média de baterias automotivas de uma certa marca está sendo estudada. Baseado em estudos similares, com outras marcas, é possível admitir que a vida dessas baterias segue a distribuição Normal com desvio padrão de 4,5 meses. De qual tamanho deverá ter a amostra, para que a amplitude do intervalo de 90% de confiança para a vida média seja de 3 meses?

Respostas

---

1. a) Serão 16 pares possíveis, todos equiprováveis com probabilidade 1/16 e a distribuição da proporção amostral de fêmeas é:

$\hat{p}$ =prop(fêmeas)	p
0	1/16
1/2	6/16
1	9/16

   b) A média da proporção amostral é $E(\hat{p}) = 0 \cdot 1/16 + 1/2 \cdot 6/16 + 1 \cdot 9/16 = 0.75$

   c) A proporção populacional de fêmeas é 3/4=0.75 que é igual à média da proporção amostral. Este resultado indica que em média a proporção amostral é igual à proporção populacional.

---

2. a) As 16 amostras possíveis são todos os pares dois a dois das idades dos quatro presidentes.

   b) A distribuição amostral das médias ficará:

Média	P(Média)
46	1/16
47,5	2/16
49	1/16
51	2/16
52	2/16
52,5	2/16
53,5	2/16
56	1/16
57	2/16
58	1/16

   c) A média populacional é 52,25 e a média das médias amostrais obtida usando a distribuição obtida no item b é:

$$E(\bar{X}) = 46 \cdot 1/16 + \cdots + 58 \cdot 1/16 = 52.25$$

   notamos que ambas são iguais.

---

3. b) As medianas de amostras de tamanho dois são exatamente iguais as médias amostrais, então a distribuição das medianas será a mesma do exercício anterior.

Mediana	P(Mediana)
46	1/16
47,5	2/16
49	1/16
51	2/16
52	2/16
52,5	2/16
53,5	2/16
56	1/16
57	2/16
58	1/16

   c) A mediana populacional é 52.5 e a média das medianas obtida usando a distribuição construída no item (b) é $E(\text{Mediana}) = 52.25$, observamos que elas são diferentes.

---

4. a)

$$Y: \text{ imunizado (0: não, 1: sim)}$$
$$Y \in \{0, 1\}$$
$$Y \sim \text{Ber}(p)$$

- Os indivíduos que receberam a vacina.
- A proporção (p) de indivíduos imunizados entre todos os que receberam a vacina (*na população*).
- O cálculo da proporção de indivíduos imunizados na amostra $\hat{p} = \sum_{i=1}^{n} Y_i/n$.
- A proporção observada em uma determinada amostra (no caso na amostra de $n = 25$ indivíduos).
- A distribuição amostral do estimador, ou seja, a distribuição que seria obtida caso fossem obtidas estimativas de *diversas* amostras.

Distribuição amostral aproximada:

$$\hat{p} = \frac{\sum_{i=1}^{n} y_i}{n} \sim \text{N}(p, \frac{p(1-p)}{n})$$

b)

$$\hat{p} \sim \text{N}\left(p = 0,80 \ , \ \frac{p(1-p)}{n} = \frac{0,80(1-80)}{25}\right)$$

$$P(\hat{p} < 0,75 | p = 0,80) = 0.266$$
$$P(\hat{p} > 0,85 | p = 0,80) = 0.266$$

---

5. a) $P(90 < Y < 110) = 0,68$
   b) $P(90 < \bar{Y} < 110) = 1,00$

---

6. a) $\mu = 512.8$
   b) 0.0052

7.

$Y \sim t_{20}$	$Y \sim \chi_{16}$	$Y \sim F_{(10,7)}$
$P(-2,85 \leq Y \leq 2,85) = 0.9901$	$P(8,91 < Y < 32,85) = 0.9093$	$P(Y > 3,18) = 0.0691$
$P(Y < -2,85) = 0.0049$	$P(Y > 8,91) = 0.9171$	$P(Y > 0,15) = 0.9959$
$P(Y > 2,85) = 0.0049$	$P(Y > 32,85) = 0.0077$	$P(Y > 5,35) = 0.0182$
$P(Y > 2,12) = 0.0234$	$P(Y > 22,80) = 0.1192$	$P(Y < 7,41) = 0.9928$
$P(Y < -3,01) = 0.0035$	$P(Y < 10,12) = 0.1397$	$P(Y < 1) = 0.4834$

---

8. a) y=1.3253, y=23.5418 e y=2.7025
   b) y=-2.086, y=6.9077 e y=0.2532
   c) y=-2.528, y=5.8122 e y=0.1923
   d) y=2.086, y=28.8454 e y=4.7611

---

9. a) 0.8175
   b) Seria 0.9545, portanto maior porque o desvio padrão da média amostral (também conhecido como erro padrão da média amostral) diminui, ou seja, a distribuição é menos dispersa em torno da média do que no item (a).

10. a) i) 0.773
ii) 0.227

b) 0.177

c) A probabilidade de uma amostra aleatória de 25 alunos obter uma média de 1200 ou mais é 0.

11.

a) $z_{\alpha/2} = 2,14$ então $1 - \alpha = 0.968$
b) $z_{\alpha/2} = 2,49$ então $1 - \alpha = 0.987$
c) $z_{\alpha/2} = 1,85$ então $1 - \alpha = 0.936$

12.

a) $z_{\alpha/2} = 2,576$
b) $z_{\alpha/2} = 1,960$
c) $z_{\alpha/2} = 1,645$

13.

a) Para 14 graus de liberdade: $\chi^2_{\alpha/2} = 31.319$
b) Para 14 graus de liberdade: $\chi^2_{\alpha/2} = 26.119$
c) Para 14 graus de liberdade: $\chi^2_{\alpha/2} = 23.685$

14.

a) $M.E. = 1.96\sqrt{\frac{0,5(1-0,5)}{4000}} = 0.0155$
b) $1 - \alpha = 79.4$ %
c) $n = 1537$
d) $n = 1844$

15.

a) $n \approx 158$
b) [0,432; 0,590]
c) 0.080
d) Conforme aumentamos o tamanho da amostra, a amplitude do intervalo de confiança diminui. Há mais informação disponível nos dados.

16.

a) [117,69; 122,31]
b) [118,43; 121,57]
c) [118,72; 121,28]
d) Conforme aumentamos o tamanho da amostra, a amplitude do intervalo de confiança diminui. Há mais informação disponível nos dados.

17.

a) $[4.359; 5.131]$
b) $[0.625; 1.86]$

18.

O IC de 90% para proporção de complicações entre os pacientes na população que fazem uso de sulfa é: (0.019; 0.088). Agora interprete este resultado e responda a pergunta.

19. $n = 16$

20. $n = 25$