Distribuições Amostrais

Prof. Eduardo Bezerra

Inferência Estatística

6 de abril de 2018

Roteiro

Roteiro

Distribuições amostrais

Considere a realização de uma amostragem aleatória simplesde uma população para produzir uma amostra de n elementos.

- Uma afirmação eventualmente feita sobre essa população será baseada em alguma estatística T, que é uma função da amostra $(X_1, X_2, \ldots, X_n)$.
- Colhida essa amostra, teremos observado um valor particular de T.

A validade dessa afirmação seria melhor compreendida se soubéssemos o que acontece quando produzimos **todas** as amostras da população.

- Isto é, qual é a distribuição de T quando $(X_1, X_2, \ldots, X_n)$ assume todos os valores possíveis.

Distribuições amostrais (cont.)

A distribuição obtida considerando todas as possíveis amostras de uma população é denominada distribuição amostral (*sampling distribution*) da estatística T.

Distribuições amostrais (cont.)

O procedimento geral para obtenção dessa distribuição envolve os seguintes componentes:

(a) uma população X, com determinado parâmetro de interesse θ;

(b) todas as amostras retiradas da população, de acordo com certo plano amostral;

Para cada amostra, calculamos o valor t (*estatística pontual*) da estatística T.

Os valores t formam uma nova população, cuja distribuição recebe o nome de distribuição amostral de T.

Distribuições amostrais - exemplo 01

Para estudar a proporção de eleitores que votariam em um candidato, são recolhidas amostras, cada uma de 300 eleitores.

- Na primeira, 175 responderam *sim*, o que gera uma estimativa pontual de 0,58 para a proporção.
- Na segunda, suponha que o valor obtido tenha sido 181, o que corresponde ao valor 0,60 para a estimativa da estatística.

A repetição desse processo n resulta em n valores para a estatística (i.e., a proporção amostral), cada um deles é uma *estimativa pontual* do valor do parâmetro sob estudo (i.e., a proporção populacional).

Distribuições amostrais - exemplo 01

Para estudar a proporção de eleitores que votariam em um candidato, são recolhidas amostras, cada uma de 300 eleitores.

- Na primeira, 175 responderam *sim*, o que gera uma estimativa pontual de 0,58 para a proporção.
- Na segunda, suponha que o valor obtido tenha sido 181, o que corresponde ao valor 0,60 para a estimativa da estatística.

A repetição desse processo n resulta em n valores para a estatística (i.e., a proporção amostral), cada um deles é uma *estimativa pontual* do valor do parâmetro sob estudo (i.e., a proporção populacional).

Distribuições amostrais - exemplo 02

Considere o *lançamento de uma moeda 50 vezes*, em que a estatística de interesse é X = qtd. de caras obtidas.

- A obtenção da distribuição amostral é feita por meio de um modelo binomial $b(50, p)$, onde p é a probabilidade de ocorrência de cara em um lançamento.
- Para julgar a honestidade dessa moeda, devemos verificar se $p = 0{,}5$.

Considere uma amostra em que se obteve 36 caras.

- Nessas condições, $\Pr(X \geq 36 | n = 50, p = 0{,}5) = 0{,}0013 = 0{,}13\%$.
- Portanto, caso a moeda seja honesta, a probabilidade resultarem 36 ou mais caras em 50 lançamentos é muito pequena, o que evidencia ser pouco provável que $p = 0{,}5$.

Distribuições amostrais - exemplo 02

Considere o *lançamento de uma moeda 50 vezes*, em que a estatística de interesse é X = qtd. de caras obtidas.

- A obtenção da distribuição amostral é feita por meio de um modelo binomial $b(50, p)$, onde p é a probabilidade de ocorrência de cara em um lançamento.
- Para julgar a honestidade dessa moeda, devemos verificar se $p = 0{,}5$.

Considere uma amostra em que se obteve 36 caras.

- Nessas condições, $\Pr(X \geq 36 | n = 50, p = 0{,}5) = 0{,}0013 = 0{,}13\%$.
- Portanto, caso a moeda seja honesta, a probabilidade resultarem 36 ou mais caras em 50 lançamentos é muito pequena, o que evidencia ser pouco provável que $p = 0{,}5$.

Roteiro

Distribuição amostral da média

Suponha que uma amostra aleatória de tamanho n seja retirada de uma população normal com média μ e variância σ^2.

- Cada observação X_i da amostra aleatória terá, portanto, a mesma distribuição normal sendo amostrada.

Pela *propriedade reprodutiva* da distribuição normal, a estatística denominada média amostral (*sample mean*)

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

tem uma distribuição normal com média

$$\mu_{\overline{X}} = \frac{1}{n}(\mu + \mu + \ldots + \mu) = \mu$$

e variância

$$\sigma^2_{\overline{X}} = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{\sigma^2}{n}.$$

Distribuição amostral da média

Pelo TLC, o que foi declarado na página anterior é verdadeiro não somente para população de distribuição normal, mas para um população com **qualquer** distribuição.

Teorema de Limite Central (TLC)

Se a amostragem for feita a partir de uma população de distribuição desconhecida **qualquer**, finita ou infinita, com média μ e variância σ^2, a distribuição amostral de $\overline{X}$ será aproximadamente normal, com média μ e variância σ^2/n, se o tamanho da amostra for grande.

Roteiro

Distribuição amostral das proporções

Considere uma população em que a proporção de elementos portadores de certa característica é p. Logo, podemos definir uma v.a. X da seguinte maneira:

$$X = \begin{cases} 1 & \text{se o elemento for portador da característica} \\ 0 & \text{se o elemento for não portador da característica} \end{cases}$$

A v.a. X assim definida segue a distribuição de Bernoulli, i.e.,

$$\mu = E(X) = p, \; \sigma^2 = Var(X) = p(1-p)$$

Distribuição amostral da proporção (cont.)

Se retirarmos uma AAS dessa população, e indicarmos por Y_n o total de elementos portadores da característica estudada **na amostra**, então

$$Y_n = X_1 + X_2 + \ldots + X_n$$

onde:

- cada X_i tem distribuição de Bernoulli;
- X_i e X_j, $i \neq j$, são independentes.

Por outro lado, sabemos que Y_n assim definida segue uma distribuição binomial, i.e.,

$$Y_n \sim b(n,p)$$

em que p é a probabilidade de um elemento ser portador da característica.

Distribuição amostral das proporções (cont.)

Vamos definir a v.a. $\widehat{p}$ como sendo a proporção de elementos na amostra que são portadores da característica, i.e.,

$$\widehat{p} = \frac{Y_n}{n}$$

Então

$$\Pr(Y_n = k) = \Pr(Y_n/n = k/n) = \Pr(\widehat{p} = k/n)$$

ou seja, a distribuição amostral de $\widehat{p}$ é obtida da distribuição de Y_n.

Distribuição amostral da proporção (cont.)

Já que $\overline{X} = (X_1 + X_2 + \ldots + X_n)/n$, podemos escrever

$$Y_n = n\overline{X}$$

Mas, pelo TLC, $\overline{X}$ possui distribuição aproximadamente normal, com média p e variância $(p(1-p))/n$, ou seja,

$$\overline{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Distribuição amostral da proporção (cont.)

Observe que, na expressão anterior, a variável $\overline{X}$ é a própria variável $\widehat{p}$ e, desse modo, para n grande, o TLC nos permite considerar a distribuição amostral de $\widehat{p}$ como sendo aproximadamente normal:

$$\widehat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Conclusão: a proporção amostral, assim como a média amostral, segue a distribuição normal.

Distribuição amostral da proporção - exemplo

Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres. Colhemos uma AAS de $n = 10$ estudantes e definimos a v.a. $\widehat{p}$ = proporção de mulheres na amostra. Qual a probabilidade de que $\widehat{p}$ difira de p em menos de 0,01?

Distribuição amostral da proporção - exemplo (cont.)

Solução. A probabilidade que deve ser calculada é

$$\Pr(|\widehat{p} - p| < 0{,}01) = \Pr(-0{,}01 < \widehat{p} - p < 0{,}01)$$

Mas

$$\widehat{p} - p \sim N\left(0, \frac{p(1-p)}{n}\right)$$

e como $p = 0{,}3$, temos que

$$Var(\widehat{p}) = p(1-p)/n = (0{,}3)(0{,}7)/10 = 0{,}021$$

e, portanto, a probabilidade pedida é igual a

$$\Pr\left(\frac{-0{,}01}{\sqrt{0{,}021}} < Z < \frac{0{,}01}{\sqrt{0{,}021}}\right) = 0{,}056.$$