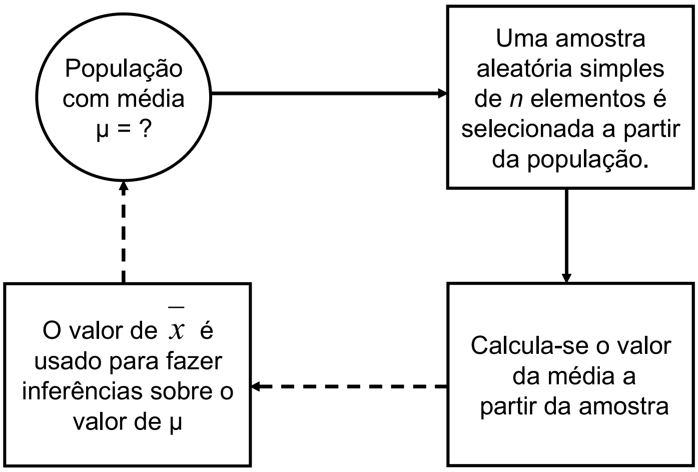
Distribuição amostral de $\bar{x}$

Um dos procedimentos estatísticos mais comuns é o uso de uma média da amostra ($\bar{x}$) para fazer inferências sobre uma população de média μ. Esse processo é apresentado na figura abaixo.



Para cada uma das repetições que possamos fazer no processo descrito acima, a estimativa de $\bar{x}$ apresentará um valor diferente. Logo, podemos dizer que a própria estimativa da média de uma variável aleatória x, também é uma variável aleatória. Assim, a distribuição de probabilidade da variável aleatória $\bar{x}$ é chamada de distribuição amostral da média da amostra.

Distribuição amostral de $\bar{x}$
A distribuição amostral de $\bar{x}$ é a distribuição de todos os valores passíveis da média da amostra.

O propósito de se estudar a distribuição amostral de $\bar{x}$ é descrever as propriedades da própria distribuição. Veremos que o conhecimento da distribuição amostral de $\bar{x}$ nos possibilitará, mais adiante, fazer declarações de probabilidade sobre os erros de amostragem evolvidos quando a média de uma amostra ($\bar{x}$) é usada para estimar a média da população (μ).

Valor Esperado de $\bar{x}$

A média da variável aleatória $\bar{x}$ é o valor esperado de $\bar{x}$, ou seja, é a média de todas as médias possíveis para uma amostras de tamanho n de uma população. É importante saber que o valor esperado de $\bar{x}$ é igual ao valor da média da população (μ). Assim, temos:

Valor Esperado de $\bar{x}$
$E(\bar{x}) = \mu$
onde,
$E(\bar{x})$ = o valor esperado de $\bar{x}$;
μ = a média da população.

Isso nos mostra que, embora não saibamos quanto o tamanho de uma amostra está próxima do tamanho da população, pelo menos sabemos que a média de todas as médias possíveis das amostras, ou seja, o valor esperado de $\bar{x}$ será igual a média da população (μ).

Desvio-Padrão de $\bar{x}$

O desvio-padrão da distribuição amostral de $\bar{x}$, também chamado de erro-padrão, depende de a população ser finita ou infinita. As duas expressões para o desvio-padrão de $\bar{x}$ são apresentadas a seguir:

Desvio-Padrão de $\bar{x}$

População Finita ($N < 30n$)

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

População Infinita ou $N \geq 30n$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

onde,

$\sigma_{\bar{x}}$ = o desvio padrão da distribuição amostral de $\bar{x}$;
σ = o desvio-padrão da população;
n = o tamanho da amostra;
N = o tamanho da população.

Teorema do Limite Central

A etapa final do estudo da distribuição amostral de $\bar{x}$ é a forma da distribuição de probabilidade de $\bar{x}$. Faremos então duas considerações, quando à distribuição da população é desconhecida, e quando essa distribuição é conhecida e *normalmente* distribuída.

Quando a distribuição é desconhecida, contamos com um dos mais importantes teoremas da estatística: *o teorema do Limite Central.*

Teorema do Limite Central

Ao selecionar amostras aleatórias simples de tamanho n a partir de uma população, a distribuição amostral de $\bar{x}$ pode ser aproximada pela distribuição *normal* de probabilidade na medida em que o tamanho da amostra se torna maior

Na prática, temos:

A distribuição amostral de $\bar{x}$ pode ser aproximada por uma distribuição *normal* sempre que o tamanho da amostra for grande, ou seja, $n \geq 30$. Para amostras pequenas, usamos a distribuição *t-Student*, com grau de liberdade *n-1* (ver http://miltonborba.org/EST/Distrib_t.pdf

Porém, quando soubemos de antemão que a população está *normalmente* distribuída, não importa o tamanho da amostra, pois a distribuição de $\bar{x}$ sempre será uma distribuição *normal*. Assim, temos:

Sempre que a população tem uma distribuição normal, a distribuição de amostragem de $\bar{x}$ tem uma distribuição normal de probabilidade para qualquer tamanho de amostra.

Exercícios sobre distribuição amostral das médias.

1 – Uma população tem uma média de 200 e um desvio padrão de 50. Uma amostra aleatória simples de tamanho 100 será tomada e a média da amostra $\bar{x}$ será usada para estimar a média da população.

a) Qual é o valor esperado de $\bar{x}$? R: $E(\bar{x}) = \mu = 200$

b) Qual é o desvio-padrão de $\bar{x}$? R: 5

2 – Uma população tem uma média de 200 e um desvio-padrão de 50. suponha que uma amostra aleatória simples de tamanho 100 seja selecionada e $\bar{x}$ seja usado para estimar μ.

a) Qual é a probabilidade de que a média da amostra selecionada estar dentro de ± 5 da média da população? R: 0,6827

b) Qual é a probabilidade de que a média da amostra estar dentro de ± 10 da média da população? R: 0,9545

3 – Assuma que o desvio-padrão da população seja $\sigma = 25$. Calcule o erro-padrão da média ($\sigma\bar{x}$), para tamanhos de amostra de 50, 100, 150, e 200. O que você pode dizer sobre o tamanho do erro-padrão da média quando o tamanho da amostra é aumentado.

$n = 50$ R: 3,535 → 4

$n = 100$ R: 2,500 → 2,5

$n = 150$ R: 2,041 → 2,0

$n = 200$ R: 1,768 → 1,8

4 – Suponha que uma amostra aleatória simples de tamanho 50 seja selecionada de uma população com $\sigma = 10$. Encontre o valor do erro-padrão da média em cada um dos seguintes casos (se apropriado, use o fator de correção da população finita).

a) o tamanho da população é infinito. R: 1,414 → 1,4

b) o tamanho da população é de $N = 50.000$. R: 1,414 → 1,4

c) o tamanho da população é $N = 5.000$ R: 1,414 → 1,4

d) o tamanho da população é $N = 500$ R: 1,343 → 1,3

5 – uma população tem uma média de 400 e um desvio-padrão de 50. A distribuição de probabilidade da população é desconhecida. Um pesquisador usará amostras aleatórias simples de 10, 20, 30 e 40 itens para coletar dados sobre a população. Com qual dessas alternativas de tamanho de amostra seremos capazes de usar a distribuição normal de probabilidade para descrever a distribuição amostral de $\bar{x}$? Explique.

6 – Uma população tem uma média de 100 e um desvio-padrão de 16. Qual é a probabilidade de uma média da amostra estar dentro de ± 2 da média da população para cada um dos seguintes tamanhos de amostra?

a) $n = 50$. R: 0,6232

b) $n = 100$. R: 0,7887

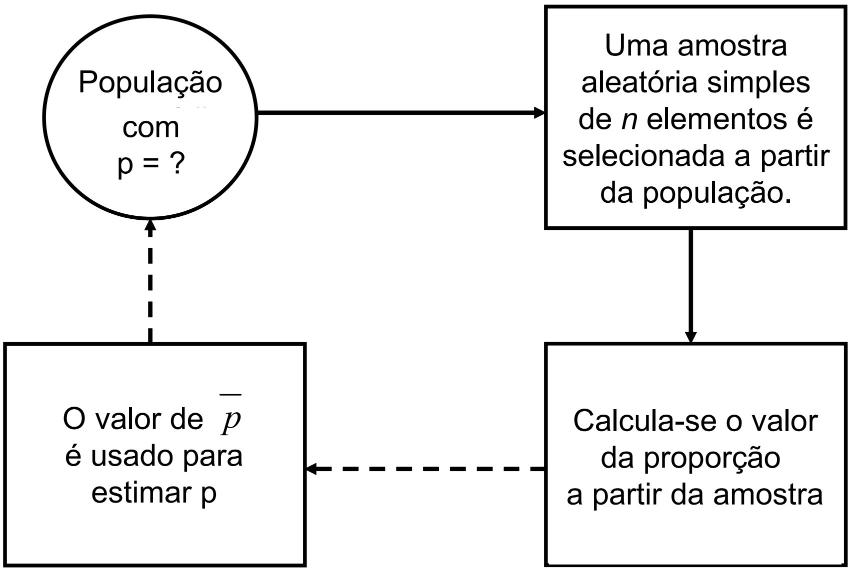
c) $n = 200$. R: 0,9229

d) $n = 400$. R: 0,9876

e) qual é a vantagem de um tamanho maior de amostra?

Distribuição amostral de $\bar{p}$

Em muitas situações usamos a proporção de uma amostra para fazer inferências sobre as proporções de uma população. A proporção de uma amostra é denotada por $\bar{p}$. Esse processo é relatado na figura abaixo.



Em cada repetição do processo podemos antecipar a obtenção de um valor diferente para a proporção da amostra. A distribuição de probabilidade para todos os valores possíveis da proporção da amostra é chamada de distribuição amostral da proporção $\bar{p}$.

Distribuição amostral de $\bar{p}$

A distribuição amostral de $\bar{p}$ é a distribuição de probabilidade de todos os valores possíveis da proporção da amostra $\bar{p}$.

Para determinar a distância que a proporção da amostra $\bar{p}$ está da proporção da população *p,* nós necessitamos entender as propriedades da distribuição amostral da variável aleatória $\bar{p}$: o valor esperado, o desvio-padrão de $\bar{p}$ e também a forma da sua distribuição amostral.

Valor Esperado de $\bar{p}$

A média da variável aleatória $\bar{p}$ é o valor esperado de $\bar{p}$, ou seja, é a média de todas as proporções possíveis para amostras de tamanho *n* de uma população. É importante saber que o valor esperado de $\bar{p}$ é igual ao valor da proporção da população (*p*). Assim, temos:

Valor Esperado de $\bar{p}$

$E(\bar{p}) = p$

onde,

$E(\bar{p})$ = o valor esperado de $\bar{p}$;

p = a proporção da população.

Isso nos mostra que, embora não saibamos quanto o tamanho de uma amostra está próxima do tamanho da população, pelo menos sabemos que a média de todas as proporções possíveis das amostras será igual à proporção da população (p).

Desvio-Padrão de $\bar{p}$

O desvio-padrão da distribuição amostral de $\bar{p}$, também chamado de *erro-padrão da proporção*, depende de a população ser finita ou infinita. As duas expressões para o desvio-padrão de $\bar{p}$ são apresentadas a seguir:

Desvio-Padrão de $\bar{p}$

População Finita ($N < 30n$)

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$$

População Infinita ou $N \geq 30n$.

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

onde,

$\sigma_{\bar{p}}$ = o desvio padrão da distribuição amostral de $\bar{p}$;

p = a proporção da população;

n = o tamanho da amostra;

N = o tamanho da população.

Teorema do Limite Central

Agora que conhecemos a média (ou o valor esperado) e o desvio-padrão de $\bar{p}$, devemos considerar a forma da distribuição amostral de $\bar{p}$. Para isso fazemos uso do Teorema Central do Limite.

Teorema do Limite Central

A distribuição amostral de $\bar{p}$ pode ser aproximada por uma distribuição *normal* sempre o tamanho da amostra for grande ($np \geq 5$ e $n(1-p) \geq 5$)

Exercícios sobre distribuição amostral das proporções.

1 – Uma amostra aleatória simples de tamanho 100 é selecionada de uma população com $p = 0{,}4$.

a) Qual o valor esperado de $\overline{p}$? R.: 0,4

b) Qual o desvio-padrão de $\overline{p}$? R.: 0,05

2 – A proporção de uma população é 0,40. Uma amostra aleatória simples de tamanho 200 será tomada e a proporção $\overline{p}$ será usada para estimar a proporção da população.

a) Qual é a probabilidade de que a proporção da amostra esteja dentro de ± 0,03 da proporção da população? R.: 0,6135

b) Qual é a probabilidade de que a proporção da amostra esteja dentro de ± 0,05 da proporção da população? R.: 0,8510

3 – Assuma que a proporção da população seja 0,55. Calcule o erro-padrão da proporção para tamanhos de amostras de 100, 200, 500 e 1000. R.: 0,05, 0,04, 0,022, 0,016 (respectivamente).

4 – A proporção de uma população é 0,30. Qual é a probabilidade de que a proporção da amostra esteja dentro de ± 0,04 da proporção da população para cada um dos seguintes tamanhos de amostra:

a) $n = 100$ R.: 0,6173
b) $n = 200$ R.: 0,7830
c) $n = 500$ R.: 0,9490
d) $n = 1000$ R.: 0,9942
e) Qual é a vantagem de um tamanho de amostra maior?

<u>REVISÃO</u>: Estimativa de Parâmetros

1 - O tempo médio de viagem para o trabalho para os habitantes de Joinville é de 31,5 minutos. Considere que a média da população seja $\mu = 31{,}5$ minutos e o desvio padrão da população seja $\sigma = 12$ minutos. Uma amostra de 50 residentes de Joinville é selecionada.

a) Qual é a probabilidade de que a média da amostra esteja dentro de ± 1 minuto para a média da população? R.: 0,4443

b) Qual é a probabilidade de que a média da amostra esteja dentro de ± 3 minutos para a média da população? R.: 0,9229

2 - O Depto. de Estatística sobre o Trabalho relatou que a média salarial horária para indivíduos em cargos executivos, administrativos e gerenciais é de R$ 24,07. Considere que a média da população seja μ = R$ 24,07 e que o desvio-padrão seja σ = R$ 4,80. Uma amostra de 120 indivíduos dos cargos executivos, administrativos e gerenciais será selecionada.

a) Qual é a probabilidade de que a média da amostra esteja dentro de ± 0,50 centavos da média da população? R.: 0,7462

b) Qual é a probabilidade de que a média da amostra esteja dentro de ± R$ 1,00 da média da população? R.: 0,9774

3 - De acordo com um jornal, o número médio de dias por ano que as pessoas que viajam a negócios estão na estrada é de 115. O desvio-padrão é de 60 dias por ano. Considere que esses resultados se aplicam à população de viajantes a negócios, e que uma amostra de 50 viajantes será selecionada da população.

a) Qual é o valor do erro-padrão da média? R.: 8

b) Qual é a probabilidade de que a média da amostra seja mais de 120 dias por ano? R.: 0,2778

c) Qual é a probabilidade de que a média da amostra seja mais de 135 dias por ano? R.: 0,0092

d) Qual é a probabilidade de que a média da amostra esteja dentro de ± 5 dias da média da população? R.: 0,4443

e) Como a probabilidade mudaria no item (d) se o tamanho da amostra fosse aumentado para 100? R.: 0,5953

4 - Qual é o fator mais importante para as pessoas que viajam a negócios quando estão hospedados em um hotel? De acordo com o mesmo jornal da questão anterior, 74% desses viajantes declaram que ter um quarto para fumante é o fator mais importante. Considere uma amostra de 200 viajantes será seleciona.

a) Qual é a probabilidade de que a proporção da amostra esteja dentro de ± 0,04 da proporção da população? R.: 0,8028

b) Qual é a probabilidade de que a proporção da amostra esteja dentro de ± 0,02 da proporção da população? R.: 0,4810

5 - Uma campanha de produção não é aceitável para embarque a clientes se uma amostra de 1000 itens contém 5% ou mais de itens defeituosos. Se uma campanha de produção tem uma proporção de defeitos da população de p = 8%, qual é a probabilidade de que $\bar{p}$ será pelo menos 0,05? R.: 0,9984

E se p fosse 10? R.: 1,0000

6 - A proporção de clientes de uma seguradora de automóveis que receberam pelo menos uma multa nos últimos 5 anos é de 0,15. A probabilidade de que a proporção da amostra esteja dentro de ± 0,03 da proporção da população é de 0,6965. Qual o tamanho da amostra? R: 150

Material elaborado por Rodrigo Coral e Carla Regina Kuss Ferreira
Atualizado por Milton Procópio de Borba