Parte II – Inferência estatística

1. Objetivos da inferência estatística

Ana Maria Lopez Calvo de Feijoo

FEIJOO, AMLC. Objetivos da inferência estatística. In: *A pesquisa e a estatística na psicologia e na educação* [online]. Rio de Janeiro: Centro Edelstein de Pesquisas Sociais, 2010, pp. 31-38. ISBN: 978-85-7982-048-9. Available from SciELO Books <http://books.scielo.org>.

PARTE II
INFERÊNCIA ESTATÍSTICA

INTRODUÇÃO

O que se fez até agora, na Estatística descritiva, foi sumarizar resultados acerca de um grupo denominado amostra, que é representativo de um grupo maior, que se denomina *população.*

Às vezes se faz necessário extrair conclusões de uma determinada investigação científica sem se dispor de evidências suficientes. Neste caso, o pesquisador precisa saber quando dispõe de evidências suficientes para chegar a uma conclusão com determinado grau de confiança.

Ao se desejar concluir com evidências incompletas pode-se lançar mão de estatística. Chamam-se *deduções probabilísticas.* Ao se deduzir a partir da probabilidade, faz-se inferência estatística.

Para se recorrer à inferência estatística, faz-se uso do raciocínio indutivo e da teoria das probabilidades.

1. OBJETIVOS DA INFERÊNCIA ESTATÍSTICA

A inferência estatística aborda dois tipos fundamentais de problemas:

Estimativa dos parâmetros de uma população

A partir dos resultados obtidos na amostra, estimam-se os parâmetros da população.

Na maioria das vezes, torna-se impossível medir-se todos os membros de uma população; para tanto, extraem-se amostras dessa população. A composição das amostras acaba por comprometer os resultados, uma vez que as medidas estatísticas calculadas sofrem variações de amostra para amostra de uma dada população – erro amostral – ou seja, por pura obra do acaso, haverá sempre uma diferença entre os resultados da amostra e os da população da qual ela foi extraída. Isto significa que a média da amostra quase nunca será exatamente igual à média da população, e o desvio será idêntico ao desvio da população em raríssimas ocasiões, mesmo que o plano amostral tenha sido bem elaborado e executado.

Como, na maioria das vezes, utiliza-se apenas uma amostra, como se pode saber se esses resultados podem ou não ser inferidos à população, ou seja, como garantir os parâmetros de uma população?

As estatísticas, portanto, estão sujeitas ao que se designa *flutuações de amostragem.* Tais estatísticas, apenas sob determinadas condições, poderão vir a ser os parâmetros da população.

Para se saber até que ponto se pode acreditar na estimativa dos parâmetros com base nos dados obtidos da amostra, recorre-se ao erro-padrão da estatística utilizada.

Segundo Nick e Kellner (1971): "O erro-padrão é teoricamente o desvio-padrão da distribuição amostral das médias, mas praticamente é uma estimativa de erro que cometemos ao substituir o parâmetro desconhecido pela estimativa obtida através de uma amostra".

Erro-padrão da media ($\sigma_{\overline{X}}$)

Mede o grau de eficiência da estimativa, mede o grau em que a média é afetada por erros de medida.

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{N-1}}$$

Onde

$$S = \frac{\sqrt{\sum X^2}}{\sqrt{N-1}}$$

Usa-se N – 1 porque o desvio-padrão da amostra é menor que o desvio-padrão da população, e apenas é utilizado quando $N < 30$.

$\sigma_{\overline{X}}$ = erro-padrão da média

S= desvio-padrão da amostra

Na verdade, seria o desvio-padrão da população, mas como seria muito trabalhoso calculá-lo, utiliza-se N – 1 no denominador.

N= número total da amostra.

Ao empregarmos o erro-padrão, devemos estar atentos ao tamanho da amostra e à variabilidade da distribuição.

– quanto maior N, menor é o erro-padrão da média.

– quanto maior o erro-padrão da média, maior o desvio-padrão.

– a distribuição amostral é sempre menor que a variabilidade da distribuição total

Através desse cálculo pode-se estabelecer o intervalo de confiança, isto é, o intervalo de valores dentro do qual a verdadeira média populacional pode cair.

Intervalos de confiança

Na estimativa da média da população.

a) 68%: $\overline{X} \pm (1)$

$\overline{X}$= média amostral

$\sigma_{\overline{X}}$ = erro-padrão da média

(1) = valor do desvio-padrão

Ao escolher-se um intervalo de confiança de 68% significa que há 68 possibilidades em 100 da média estar correta.

b) 95%: $\overline{X} \pm (1{,}96)$

$\overline{X}$ = média amostral

$\sigma_{\overline{X}}$ = erro-padrão da média

(1,96) = valor de 2 desvios-padrão

Pode-se acertar 95% das vezes e errar 5% das vezes.

c) 99%: $\overline{X} \pm (2{,}58)$

$\overline{X}$= média amostral

$\sigma_{\overline{X}}$ = erro-padrão da média

(2,58) = valor de 3 desvios-padrão

Pode-se acertar 99% das vezes e errar 1% das vezes.

Este procedimento deve ser realizado quando se desejar fazer generalizações de uma amostra para uma população, ou seja, utilizar a média da amostra como parâmetro da população.

Erro-padrão da mediana (σ_{md}):

O erro-padrão da mediana é maior que o erro-padrão da média, o que mostra que a média é, em geral, mais fidedigna, menos sujeita às variações do que a mediana.

O erro-padrão da mediana é utilizado em estatísticas não paramétricas.

$$\sigma_{md} = \frac{1,253\ \sigma}{\sqrt{N}} \quad \text{ou} \quad \sigma_{md} = \frac{1,868Q}{\sqrt{N}}$$

Erro-padrão do desvio-padrão

$$\sigma_{\sigma} = \frac{0,71\ \sigma}{\sqrt{N}}$$

Erro-padrão do desvio semi-interquartil

Utilizado em estatísticas não paramétricas.

$$\sigma_{Q} = \frac{0,786Q}{\sqrt{N}}$$

Prova de hipóteses

A prova de hipóteses tem como objetivo testar a significância das diferenças. Para tanto, tenta-se provar uma hipótese, usando-se como intermediária a hipótese estatística. A análise estatística é, em parte, função da prova estatística empregada na análise. Por tal motivo, ao escolher-se a prova estatística, deve-se ficar atento ao fato de que as exigências para o uso de uma prova devem ser cumpridas.

A escolha da prova estatística deve estar condicionada aos seguintes fatores:

– Plano de amostragem

– Natureza da população da qual se extraiu a amostra

– Nível de mensuração das variáveis

– Variância das populações

Critérios para a escolha da prova estatística

Plano de amostragem

Consiste num processo de seleção que segue regras e operações para que elementos da população sejam incluídos na amostra.

Fatores tais como o número de elementos que contém a amostra, a extração de um elemento de uma amostra afetando ou não a composição de outra amostra e o método de amostragem interferem na escolha da prova estatística. Este processo necessita de um critério objetivo.

Sabe-se que quanto maior a amostra, maior é a probabilidade de se encontrar a normalidade. Ao aumentar o tamanho da amostra, aumenta a probabilidade desta assemelhar-se à população.

O recurso de se aumentar o tamanho da amostra para aumentar a precisão da estimativa da média da população tem seu grau de eficiência relacionado com a raiz quadrada do tamanho da amostra.

Só podemos considerar uma amostra como sendo representativa quando é constituída por elementos selecionados de acordo com uma técnica conhecida – processo de seleção – ou seja, regras e operações mediante as quais alguns membros da população são incluídos na amostra.

Processo de seleção:

– seleção probabilística

– seleção não probabilística

O método de amostragem que possibilita com maior grau de certeza a representativa da amostragem é o *aleatório*. Neste método, cada indivíduo da população possui a mesma chance de ser escolhido. Considera-se, também, que a seleção de um indivíduo não influenciará de forma alguma a escolha de outro.

A extração da amostra aleatória em psicologia apresenta três peculiaridades:

- A seleção sistemática dos componentes da amostra aproxima-se da aleatoriedade, uma vez que os membros estão arrolados. Assim, pode-se selecionar um nome de cada grupo. Os grupos vão depender do tamanho da amostra.
- Conhece-se a distribuição daquela característica que se deseja estudar. Por exemplo, sabe-se que inteligência distribui-se normalmente dentro da população; logo, uma amostra extraída aleatoriamente desta população também apresentará uma distribuição normal.
- A população, não sendo nitidamente definida, torna a amostragem inacessível. Quando isso ocorre, cabe uma verificação da adequação da amostra. Extraem-se diversas amostras aleatórias de aproximadamente o mesmo tamanho; caso tais amostras apresentem resultados muito diferentes, nenhuma delas é representativa.

Natureza da população da qual se extraiu a amostra

A população pode ser, em sua natureza, normal ou não normal. Muitos fenômenos na natureza distribuem-se de acordo com a normalidade, como, por exemplo, a altura, o peso e a inteligência. Outros fenômenos, no entanto, por natureza não se encaixam na distribuição normal, como, por exemplo, mortalidade, amamentação por classe socioeconômica.

Quando o fenômeno a ser estudado se caracteriza por uma distribuição normal, utiliza-se a estatística paramétrica, caso contrário recorre-se à estatística não paramétrica.

Nível de mensuração das variáveis

A classificação de medidas em determinados níveis foi feita por Stevens. Do mais baixo para o mais alto nível, temos as escalas: nominal, ordinal, intervalar e razão. São distinguidas em termos de critérios diferentes.

De acordo com a definição de mensuração, é possível atribuir-se números a objetos de acordo com certas regras.

Existem também diferenças no modo como podem ser feitas as operações estatísticas com números aplicados de níveis de medida.

Os níveis são atingidos quando aplicados a eles os postulados básicos para a média.

1° nível: NOMINAL – É aplicado o postulado:

1 – *Identidade:* o número é ele mesmo e nem mais um outro.

2º nível: ORDINAL – São aplicados os postulados:

1 – Identidade
2 – *Ordem:* os números permitem relação de grandeza, podendo-se colocar em ordem crescente ou decrescente.

3º nível: INTERVALAR – São aplicados os postulados:

1 – *Identidade*
2 – *Ordem*
3 – *Aditividade:* os intervalos entre os números se dão de forma constante: dessa forma, fica garantida a adição de dois valores, por exemplo, a + b = b + a.

Variância da população

Sempre que o teste exigir homogeneidade da variância, pode-se calcular as variâncias de suas amostras. Para tanto, dever-se-á utilizar o teste F.

$$F = \frac{S^2M}{S^2m}$$

S^2M = variância maior
S^2m= variância menor

Consultar a tabela da distribuição F, com graus de liberdade para o numerador e o denominador (N de cada grupo menos 1). Verificar se a diferença entre as variâncias é ou não significativa.

Quanto maior o numerador e menor o denominador com maior probabilidade, vamos obter um resultado significativo em termos estatísticos.

Averiguados todos esses critérios, procede-se à escolha da prova estatística, cujo objetivo está em verificar se a diferença entre as amostras ocorre por puro erro amostrai ou realmente por tratar-se de amostras que diferem quanto às características estudadas.