Estatística Para leigos

Seja para estudar para a prova, seja apenas para compreender os dados do cotidiano ao nosso redor, saber como e quando usar as técnicas de análise de dados e as fórmulas estatísticas será útil. Conseguir fazer conexões entre as técnicas estatísticas e as fórmulas talvez seja ainda mais importante. Isso desenvolve nossa confiança para resolver os problemas estatísticos e solidifica nossas estratégias para realizar projetos estatísticos.

EXAMINANDO OS INTERVALOS DE CONFIANÇA ESTATÍSTICOS

Em Estatística, um *intervalo de confiança* é um palpite calculado sobre alguma característica da população. Um intervalo de confiança contém uma estimativa inicial mais ou menos com uma *margem de erro* (a quantidade pela qual você espera que seus resultados variem, caso uma amostra diferente seja usada). A tabela a seguir mostra as fórmulas para os componentes dos intervalos de confiança mais comuns e as indicações para quando usá-los.

IC para	Estatística Amostral	Margem de Erro	Usar Quando
Média populacional *(μ)*	$\bar{x}$	$\pm z^* \frac{\sigma}{\sqrt{n}}$	X é normal, ou $n \geq 30$; σ conhecido
Média populacional *(μ)*	$\bar{x}$	$\bar{x} \pm t^*_{n-1} \frac{s}{\sqrt{n}}$	$n < 30$, e/ou σ desconhecido
Proporção populacional (p)	$\hat{p}$	$\pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$n\hat{p},\ n(1-\hat{p}) \geq 10$
Diferença de duas médias populacionais (μ_1 - μ_2)	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	As duas são distribuições normais n_1, $n_2 \geq 30$; σ_1, σ_2 conhecidos
Diferença de duas médias populacionais μ_1 - μ_2	$\bar{x}_1 - \bar{x}_2$	$\bar{x}_1 - \bar{x}_2 \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	n_1, $n_2 < 30$; e/ou $\sigma_1 = \sigma_2$ desconhecidos
Diferença de duas proporções ($p_1 - p_2$)	$\hat{p}_1 - \hat{p}_2$	$\pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$n\hat{p},\ n(1-\hat{p}) \geq 10$ para cada grupo

Estatística Para leigos

ENTENDENDO AS FÓRMULAS PARA AS ESTATÍSTICAS COMUNS

Após a coleta dos dados, o primeiro passo da análise é fazer algumas análises descritivas para termos uma noção geral sobre os dados. Por exemplo

- **Onde está o centro dos dados?**
- **Qual é a dispersão dos dados?**
- **Qual é o nível de correlação entre os dados e duas variáveis?**

As estatísticas descritivas mais comuns estão na tabela a seguir, junto com suas fórmulas e uma breve descrição sobre o que cada uma mede.

Estatística	Fórmula	Usada para
Média amostral	$\bar{x} = \frac{\sum x_i}{n}$	Medida do centro; é afetada por valores atípicos
Mediana	*n* ímpar: valor do meio dos dados organizados *n* par: média de dois valores do meio	Medida do centro; não é afetada por valores atípicos
Desvio-padrão amostral	$s = \sqrt{\sum \frac{(x - \bar{x})^2}{n-1}}$	Medida de variação; medida "média" a partir da média
Coeficiente de correlação	$r = \frac{1}{n-1}\left(\frac{\sum_x \sum_y (x - \bar{x})(y - \bar{y})}{s_x s_y}\right)$	Força e direção da relação linear entre *X* e *Y*

DESCOBRINDO O TAMANHO AMOSTRAL ESTATISTICAMENTE

Ao projetar um estudo, o tamanho amostral é uma consideração importante, pois quanto maior ele for, mais dados você terá e, assim, mais precisos serão seus resultados (considerando dados de alta qualidade). Se você souber o nível de precisão pretendido (ou seja, sua margem de erro desejada), poderá calcular o tamanho amostral necessário para alcançá-lo.

Para descobrir o tamanho amostral necessário para estimar a média de uma população (μ), use a seguinte fórmula:

$$n = \left(\frac{z^{*}\sigma}{MDE} \right)^{2}$$

Nessa fórmula, MDE representa a margem de erro desejada (que deve ser estabelecida com antecedência) e **σ** representa o desvio-padrão populacional. Caso **σ** seja desconhecido, você pode estimá-lo com o desvio-padrão amostral, s, a partir de um estudo-piloto; z^* é o valor crítico para o intervalo de confiança que você precisa.

VERIFICANDO OS VALORES ESTATÍSTICOS CRÍTICOS DOS INTERVALOS DE CONFIANÇA

Os valores críticos (valores z^*) são componentes importantes dos intervalos de confiança (a técnica estatística para estimar os parâmetros populacionais). O valor z^*, que aparece na fórmula da margem de erro, mede o número de erros-padrão que devem ser adicionados e subtraídos de modo a atingir seu nível de confiança desejado (a porcentagem de confiança que você quer). A tabela a seguir apresenta os níveis de confiança comuns e seus valores z* correspondentes.

Nível de Confiança	Valor z*
80%	1,28
85%	1,44
90%	1,64
95%	1,96
98%	2,33
99%	2,58

LIDANDO COM TESTES DE HIPÓTESES ESTATÍSTICOS

Usamos os testes de hipóteses para desafiar a veracidade dos argumentos a respeito de uma população (por exemplo, um argumento de que 40% dos brasileiros possuem um telefone celular). Para testar uma hipótese estatística, pegue uma amostra, colete os dados, forme uma estatística, padronize-a para formar uma estatística de teste (para que possa ser interpretada em uma escala padronizada) e decida se a estatística de teste refuta o argumento. A tabela a seguir traz detalhes importantes sobre os testes de hipóteses.

Teste para	Hipótese Nula (H_0)	Estatística de Teste	Distribuição	Usar Quando
Média populacional *(μ)*	$\mu = \mu_0$	$\frac{(\bar{x}-\mu_o)}{\sigma/\sqrt{n}}$	Z	Distribuição normal ou $n > 30$; σ desconhecido
Média populacional *(μ)*	$\mu = \mu_0$	$\frac{(\bar{x}-\mu_o)}{s/\sqrt{n}}$	t_{n-1}	$n > 30$, e/ou σ desconhecido
Proporção populacional (p)	$p = p_0$	$\frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$	Z	$n\hat{p},\ n(1-\hat{p}) \geq 10$
Diferença de duas médias *($\mu_1 - \mu_2$)*	$\mu_1 - \mu_2 = 0$	$\frac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	Z	As duas são distribuições normais ou n_1, $n_2 \geq 30$; σ_1, σ_2 conhecido
Diferença de duas médias *($\mu_1 - \mu_2$)*	$\mu_1 - \mu_2 = 0$	$\frac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	distribuição *t* com *gl* = o menor de n_1-1 e n_2-1	n_1, $n_2 < 30$ e/ou σ_1, σ_2 desconhecido
Diferença média μ_d (dados pareados)	$\mu_d = 0$	$\frac{(\bar{d}-\mu_d)}{s_d/\sqrt{n}}$	t_{n-1}	n < 30 pares de dados e/ou σ_d desconhecido
Diferença de duas proporções ($p_1 - p_2$)	$p_1 - p_2 = 0$	$\frac{(\hat{p}_1-\hat{p}_2)-0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$	Z	$n\hat{p},\ n(1-\hat{p}) \geq 10$ para cada grupo