Universidade Federal do Paraná

Setor de Ciências Exatas

Departamento de Estatística

2ª lista de exercícios de Inferência Estatística (CE085)

Teste de Hipóteses

Marcelo Adriano Corrêa Maceno

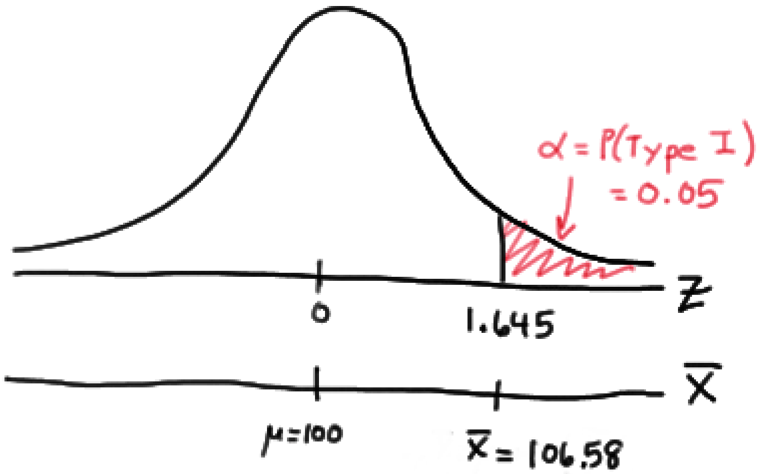
Eduardo Vargas Ferreira

Curitiba

Novembro de 2015

Exercício 1 – Considerando uma v.a x1, ... xn que segue uma distribuição normal N~(μ,4²). X representa o QI de uma amostra selecionada de adultos americanos. Pegando uma amostra n=16 estudantes e considerando o erro do tipo I α=0,05. Teste a hipótese H0: μ=100 contra a hipótese alternativa H1: μ >100. Qual é o poder do teste de hipótese se a verdadeira média populacional fosse μ=108.

Resposta: Sendo α=0,05, implica que devemos rejeitar a hipótese nula quando a estatística Z≥1,645, assim:



Verificando qual valor de média a partir da qual rejeitamos a hipótese H0:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Longrightarrow \bar{X} = \mu + z\left(\frac{\sigma}{\sqrt{n}}\right) \Longrightarrow \bar{X} = 100 + 1.645\left(\frac{16}{\sqrt{16}}\right) = 106.58$$

Assim, quando $\bar{x} \geq 106{,}58$ rejeitamos a hipótese H0. Isso implica que a função poder, que é a probabilidade de rejeitar a hipótese nula, quando μ=108 é:

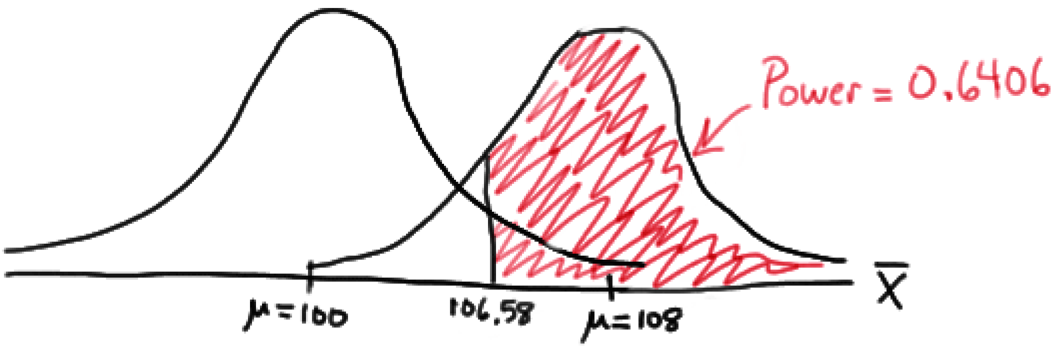
$$\text{Power} = P(\bar{X} \geq 106.58 \text{ when } \mu = 108) = P\left(Z \geq \frac{106.58 - 108}{16/\sqrt{16}}\right)$$

$$= P(Z \geq -0.36) = 1 - P(Z < -0.36)$$

$$= 1 - \Phi(-0.36)$$

$$= 1 - 0.3594 = 0.6406$$

Para uma melhor ilustração do que seria a função poder, segue:



Assim, determinamos que temos apenas 64,06 % de chance de rejeitar a hipótese nula H0: μ=100 em favor da hipótese alternativa H1: μ>100 se a verdadeira média populacional é na realidade μ=108.

Exercício 2 – Considerando os mesmos dados do Exercício 1, porém, com uma média populacional verdadeira μ=116. Qual é o poder do teste?

O α é igual do exercício 1, ou seja α=0,05. Assim, rejeitamos a hipótese nula H0 quando Z≥1,645, ou, equivalentemente, quando a média amostral observada é 106,58 ou maior. Isto significa que a probabilidade de rejeitar a hipótese, quando μ=116 é 0,9909, conforme abaixo:

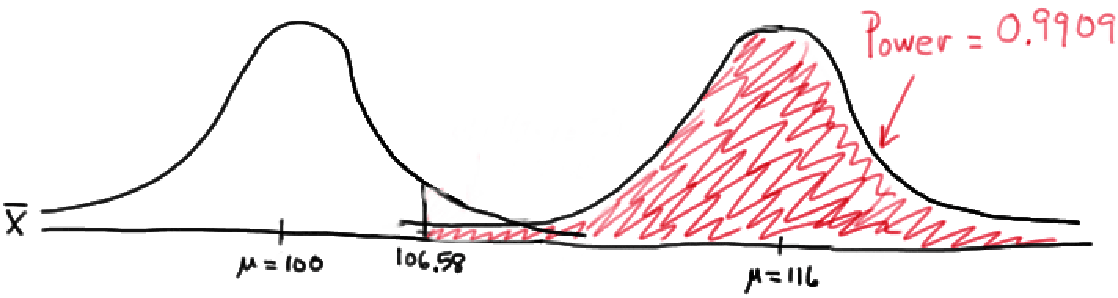
$$\text{Power} = P(\bar{X} \geq 106.58 \text{ when } \mu = 116) = P\left(Z \geq \frac{106.58 - 116}{16/\sqrt{16}}\right)$$

$$= P(Z \geq -2.36) = 1 - P(Z < -2.36)$$

$$= 1 - \Phi(-2.36)$$

$$= 1 - 0.0091 = 0.9909$$

De forma ilustrativa:



Resumindo, foi determinado que, neste caso, temos 99,09 % de chance de rejeitar a hipótese nula H0: μ=100 em preferência da hipótese alternativa H1: μ>100 se a média populacional é na realidade μ=116. Verifica-se que a probabilidade de rejeitar a hipótese H0 é maior do que quando considerou-se que a média populacional verdadeira é μ=108.

Exercício 3 – Uma companhia de processamento de alimento empacota mel em pequenos potes de vidro. Cada pote supostamente contém 10 mL de mel. Em um experimento prévio verificou-se que o volume X, o volume em mL de uma amostra da companhia de mel segue uma distribuição normal com uma variância 2. Verifique, em um nível de significância α de 0,05, a hipótese nula Ho: μ=10 contra a hipótese alternativa H1: μ≠10.

Resposta:

Estamos interessados em testar a hipótese nula H0: μ=10 contra a hipótese alternativa H1: μ≠10 para uma média normal, nosso espaço paramétrico é:

$$\Omega = \{\mu: -\infty < \mu < \infty\}$$

Encontrando λ:

$$\lambda = \frac{L(10)}{L(\bar{x})} = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left[-\frac{(x_i-10)^2}{2(2)}\right]}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left[-\frac{(x_i-\bar{x})^2}{2(2)}\right]}$$

O qual simplifica para:

$$\lambda = \frac{\exp\left[-\frac{1}{4}\sum_{i=1}^{n}(x_i-10)^2\right]}{\exp\left[-\frac{1}{4}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]}$$

Focando apenas na soma no numerador. Se “adicionarmos 0” de forma especial para a quantidade em parênteses:

$$\sum_{i=1}^{n}(x_i \overbrace{-\bar{x}+\bar{x}}^{\text{ADD } 0} -10)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + 2(\bar{x}-10)\overbrace{\boxed{\sum_{i=1}^{n}(x_i-\bar{x})}}^{0} + \overbrace{\boxed{\sum_{i=1}^{n}(\bar{x}-10)^2}}^{n(\bar{x}-10)^2}$$

Pode-se mostrar que a soma pode ser escrita como:

$$\sum_{i=1}^{n}(x_i-10)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-10)^2$$

Assim, a razão de verossimilhança se torna:

$$\lambda = \frac{\exp\left[-\frac{1}{4}\sum_{i=1}^{n}(x_i-\bar{x})^2-\frac{n}{4}(\bar{x}-10)^2\right]}{\exp\left[-\frac{1}{4}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]} = \frac{\cancel{\exp\left[-\frac{1}{4}\sum(x_i-\bar{x})^2\right]}\exp\left[-\frac{n}{4}(\bar{x}-10)^2\right]}{\cancel{\exp\left[-\frac{1}{4}\sum(x_i-\bar{x})^2\right]}}$$

O qual simplifica para:

$$\lambda = exp\left[-\frac{n}{4}(\bar{x}-10)^2\right]$$

O teste de razão de verossimilhança rejeita a hipótese nula quando o λ for pequeno, assim:

$$\lambda = exp\left[-\frac{n}{4}(\bar{x}-10)^2\right] \leq k$$

No qual k é escolhido para garantir, neste caso, que α=0,05. Tomando o log natural em ambos os lados da inequação, pode-se mostrar que λ≤k é equivalente a:

$$-\frac{n}{4}(\bar{x}-10)^2 \leq \ln k$$

Multiplicando ambos os lados por -4/n, é equivalente a:

$$(\bar{x}-10)^2 \geq -\frac{4}{n}\ln k$$

$$\frac{|\bar{X}-10|}{\sigma/\sqrt{n}} \geq \frac{\sqrt{-(4/n)\ln k}}{\sigma/\sqrt{n}} = k*$$

A parte esquerda da inequação a padronização da distribuição normal. Considerando essa padronização como Z, assim:

$$Z = \frac{\bar{X}-10}{\sigma/\sqrt{n}}$$

Segue uma distribuição quando H0: μ=10. Podemos determina um k* apropriado usando a tabela de distribuição normal padronizada. Mostramos que o teste de razão de verossimilhança a hipótese nula H0: μ=10 em preferência da hipótese H1: μ≠10 para todas as médias amostrais as quais seguem:

$$\frac{|\bar{X}-10|}{\sqrt{2}/\sqrt{n}} \geq z_{0.025} = 1.96$$

Fazendo isto garante-se que a probabilidade de cometer um erro do tipo I é de 0,05.

Exercício 4 – Suponha que X é uma única observação de uma população com função de densidade de probabilidade dado por:

$$f(x) = \theta x^{\theta - 1}$$

Para 0<x<1. Encontre o teste com melhor região crítica, ou seja, o teste mais poderoso, com nível de significância α=0,05 para testar a hipótese nula H0: θ=3 contra a hipótese alternativa H1: θ=2.

Resposta: Devido a ambas a hipóteses serem hipóteses simples, podemos aplicar o lema de Neyman Pearson para encontrar o teste mais poderoso. O lema nos diz que a razão de verossimilhança sob a hipótese e a hipótese alternativa deve ser menor do que uma constante k. Assim, por estarmos lidando com apenas uma observação de X, a razão de verossimilhança fica:

$$\frac{L(\theta_0)}{L(\theta_\alpha)} = \frac{3x^{3-1}}{2x^{2-1}} = \frac{3}{2}x \leq k$$

Isto é, segundo o lema a região de rejeição para o teste mais poderoso é:

$$\frac{3}{2}x \leq k$$

ou alternativamente, sendo (2/3k) uma nova constante k*, a região de rejeição para o teste mais poderoso é:

$$x < \frac{2}{3}k = k^*$$

Agora, tem que ser encontrado este fator k*. Queremos que α seja igual a 0,05. Para que isto aconteça:

$$\alpha = P(X < k^* \text{ when } \theta = 3) = \int_0^{k^*} 3x^2 dx = 0.05$$

Fazendo a integração, chegamos a:

$$\left[x^3\right]_{x=0}^{x=k^*} = (k^*)^3 = 0.05$$

E resolvendo k*:

$$k^* = (0.05)^{1/3} = 0.368$$

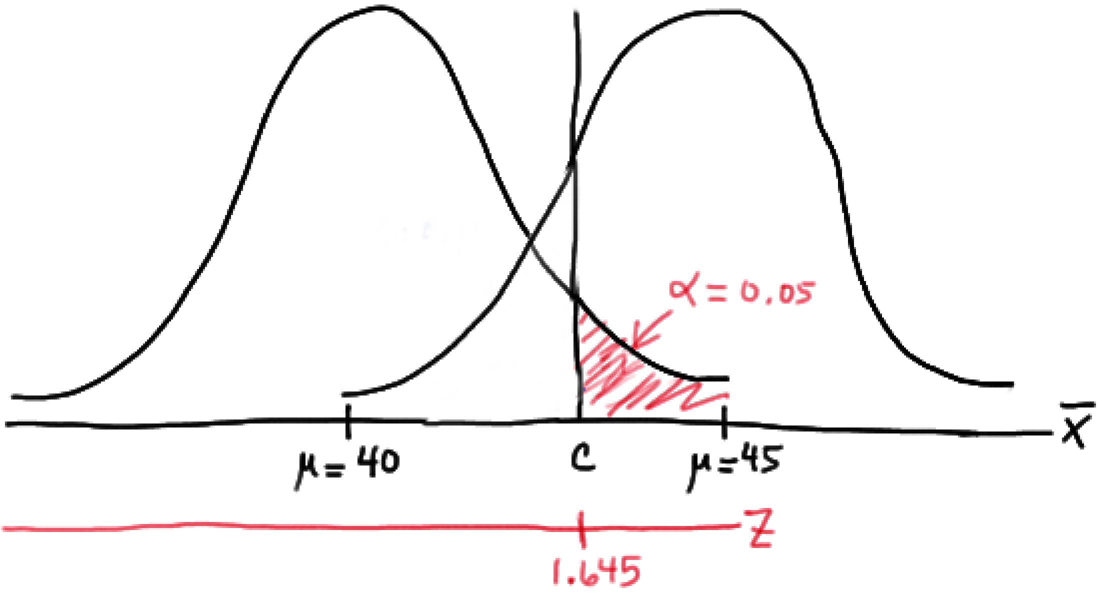
Isto é, pelo lema de Neyman Pearson a região de rejeição do teste mais poderoso H0: θ=3 contra H1: θ=2 é:

$$x < 0.368$$

Isto é, entre todos os testes possíveis para testar H0: θ=3 contra H1: θ=2, baseado em apenas uma observação e com nível de significância α=0,05, este teste tem o maior valor possível para a função poder sob a hipótese alternativa, isto é, quando θ=2.

Exercício 5 – Sendo X a quantidade de milho obtido em uma pequena plantação medido como ton/hectare. Sabendo que esta variável aleatória segue uma distribuição normal com média μ e desvio padrão σ=6. Um pesquisador está tentando aumentar a quantidade de milho obtida na plantação para acima de 40 ton/hectare. Assim, ele está interessado em testar a hipótese H0: μ=40 contra a hipótese alternativa H1: μ>40. Encontre o número de amostras n que ele precisa obter para que o poder do teste seja igual a 0,90, sendo a média alternativa μ=45.

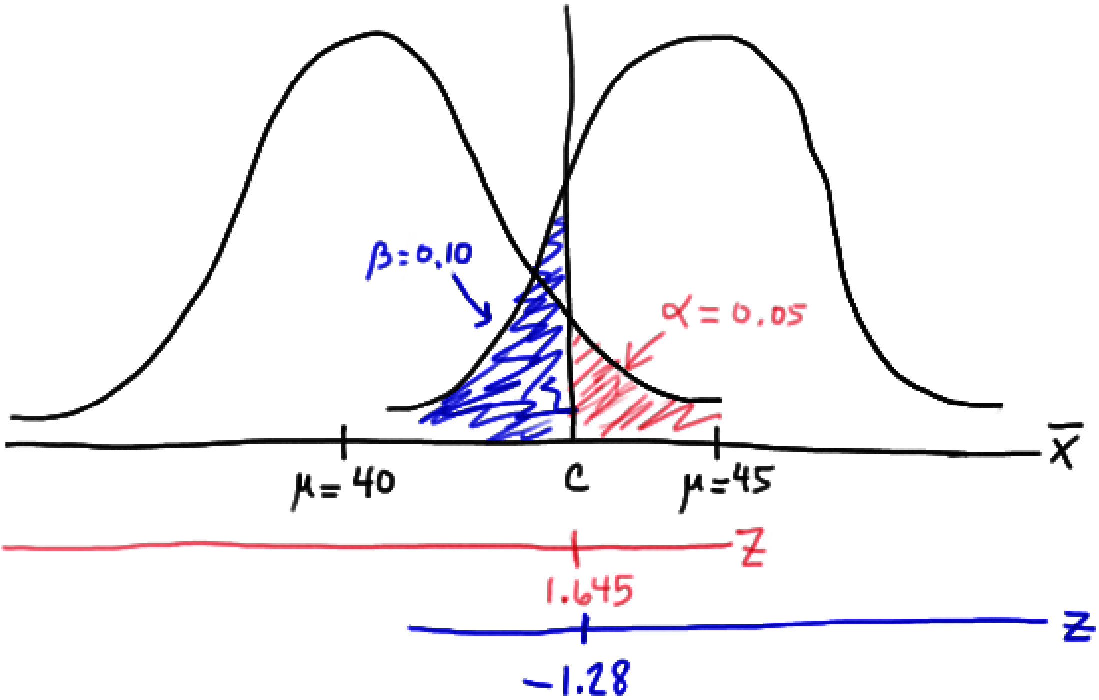
Precisamos encontrar um valor c, tal que se a média amostral for maior do que c, rejeitamos a hipótese nula, assim:



Para que o teste tenha α=0,05, a equação seguinte precisa ser satisfeita:

$$c = 40 + 1.645\left(\frac{6}{\sqrt{n}}\right)$$

Mas, esta não é única condição que c precisa seguir, pois c também precisa ser definido para que o poder seja igual a 0,90. Ou, alternativamente, que a probabilidade do erro do tipo II seja 0,10. Isto acontecerá se houver uma chance de 10 % de nosso teste estatístico estar perto de c, como ilustrado abaixo:



Esta ilustração sugere que, para que nossa hipótese tenha poder de 0,90, a equação seguinte tem que ser satisfeita:

$$c = 45 - 1.28\left(\frac{6}{\sqrt{n}}\right)$$

Assim, temos duas equações para c e duas variáveis. Resolvendo as equações:

$$40 + 1.645\left(\frac{6}{\sqrt{n}}\right) = 45 - 1.28\left(\frac{6}{\sqrt{n}}\right)$$

$$(1.645 + 1.28)\frac{6}{\sqrt{n}} = 45 - 40$$

$$\frac{17.55}{\sqrt{n}} = 5$$

$$\sqrt{n} = \frac{17.55}{5} = 3,51$$

$$n = 3.51^2 = 12.3 \approx 13$$

Assim, para este caso n tem que ser igual a 13, resolvendo a equação de c:

$$c = 40 + 1.645\left(\frac{6}{\sqrt{13}}\right) = 42.737$$

O pesquisador tem que coletar 13 amostras e rejeitar a hipótese nula H0: μ=40 se a média destas 13 amostras for maior do que 42,737 ton/hectare. O pesquisador terá 5 % de chance de cometer o erro do tipo I e 10 % de chance de cometer o erro do tipo II se a média populacional for na verdade 45 ton/hectare.