MAE116 – Noções de Estatística

Grupos B e D – II Semestre de 2020

Lista de Exercícios 8 – Teste de Hipóteses I – CASA (gabarito)

Exercício 1.

Sabe-se que em certa região do litoral, 20% dos peixes apresentam câncer de fígado causado pela poluição. O Ministério da Pesca e Aquicultura está preocupado de que essa proporção tenha aumentado após um vazamento de óleo próximo a essa região e resolve realizar um teste com base em uma amostra aleatória de peixes dessa região, após o vazamento.

Solução

(a) Formule este problema como um problema de teste de hipóteses (estabeleça as hipóteses nula e alternativa, e especifique o parâmetro de interesse).

O parâmetro de interesse é a proporção de peixes que apresentam câncer de fígado causado pela poluição após o vazamento de óleo. Vamos denotá-lo por p.

As hipóteses de interesse são:

$$\mathrm{H}_0 : p = 0,20 \quad \text{versus} \quad \mathrm{H}_1 : p > 0,20.$$

As hipóteses possuem a seguinte interpretação:

H_0 : A proporção de peixes que apresentam câncer de fígado causado pela poluição após o vazamento não difere da proporção antes do vazamento.

H_1 : O vazamento aumentou a incidência de câncer de fígado nos peixes.

(b) Construa a região crítica correspondente ao nível de significância de 5%, para uma amostra de 100 peixes. Se 25 dos 100 peixes analisados apresentam câncer de fígado, qual é a estimativa pontual do parâmetro? Qual é a conclusão do teste?

Recorde que a região crítica dos testes estudados em aula depende da hipótese alternativa formulada. Para testar as hipóteses

$$\mathrm{H}_0 : p = 0,20 \quad \text{versus} \quad \mathrm{H}_1 : p > 0,20$$

a região crítica é da forma

$$\mathrm{RC} = \{\widehat{p} : \widehat{p} \geqslant a\},$$

em que $\widehat{p}$ é a proporção amostral e a é um valor de referência que pode ser obtido da distribuição amostral de $\widehat{p}$, para um dado nível de significância α.

Sendo a probabilidade do erro tipo I igual a $\alpha = 0,05$, temos que

$$\mathrm{P}(\widehat{p} \in \mathrm{RC}, \text{sendo } \mathrm{H}_0 \text{ verdadeira}) = \mathrm{P}(\widehat{p} \geq a, \text{sendo } p = 0,20) = 0,05.$$

Utilizando o Teorema Limite Central para aproximar a distribuição amostral de $\widehat{p}$, temos que

$$\widehat{p} \sim \text{Normal}\left(p, \frac{p(1-p)}{n}\right), \quad \text{aproximadamente,}$$

e, consequentemente,

$$Z = \frac{\widehat{p} - p}{\sqrt{\dfrac{p \times (1-p)}{n}}} \sim \text{Normal}\,(0,1)\,, \quad \text{aproximadamente.}$$

Assim, sob a hipótese H_0,

$$\begin{aligned}
&\mathrm{P}(\widehat{p} \geq a, \text{sendo } p = 0,20) = 0,05 \\
&\Rightarrow \mathrm{P}\left(\underbrace{\frac{\widehat{p} - 0,20}{\sqrt{\dfrac{0,20 \times (1-0,20)}{100}}}}_{Z} \geq \underbrace{\frac{a - 0,20}{\sqrt{\dfrac{0,20 \times (1-0,20)}{100}}}}_{z}\right) = 0,05 \\
&\Rightarrow \mathrm{P}\,(Z \geq z) = 0,05,
\end{aligned}$$

em que $Z \sim \mathrm{N}(0,1)$, e z é uma constante tal que $\mathrm{P}\,(Z \geq z) = 0,05$. Utilizando a tabela fornecida no e-disciplinas obtemos que $z = 1,64$.

Obtemos então que

$$\frac{a - 0,20}{\sqrt{\frac{0,20 \times (1-0,20)}{100}}} = 1,64,$$

logo

$$a = 0,20 + 1,64 \times \sqrt{\frac{0,20 \times (1-0,20)}{100}} = 0,20 + 1,64 \times 0,04 = 0,2656.$$

Decisão: $\mathrm{RC} = \{\widehat{p} : \widehat{p} \geq 0,2656\}$. Como $\widehat{p}_{obs} = \frac{25}{100} = 0,25$ não está dentro da região crítica, então decidimos não rejeitar H_0, a um nível de significância de 5%.

Conclusão: Ou seja, a um nível de significância de 5%, concluímos que não há evidências suficientes de que o vazamento aumentou a proporção de peixes com câncer de fígado.

(c) Supondo agora que a amostra é de 200 peixes, construa a região crítica para o nível de significância de 5%. Se 50 dos 200 peixes analisados apresentam câncer de fígado, qual é a estimativa pontual do parâmetro? Qual é a conclusão do teste?

Uma vez que o nível de significância é o mesmo, podemos aproveitar o procedimento do item anterior, e ir diretamente para a obtenção do valor de a, referente a região crítica

$$\text{RC} = \{\widehat{p} : \widehat{p} \geqslant a\}\,.$$

Temos que

$$\frac{a - 0,20}{\sqrt{\frac{0,20\times(1-0,20)}{200}}} = 1,64,$$

logo

$$a = 0,20 + 1,64 \times \sqrt{\frac{0,20 \times (1-0,20)}{200}} = 0,20 + 1,64 \times 0,0283 = 0,2464.$$

Decisão: RC $= \{\widehat{p} : \widehat{p} \geq 0,2464\}$. Como $\widehat{p}_{obs} = \frac{50}{200} = 0,25$ está dentro da região crítica, então decidimos rejeitar H_0, a um nível de significância de 5%.

Conclusão: A um nível de significância de 5%, concluímos que existem evidências suficientes de que o vazamento aumentou a proporção de peixes com câncer de fígado.

(d) Compare os resultados dos itens (b) e (c) e discuta.

As estimativas pontuais são iguais a $0,25$ nos itens (b) e (c), bem como os níveis de significância (0,05). O que difere são os tamanhos das amostras. Com uma amostra maior, há mais evidência, logo a mesma estimativa pontual foi significativa para rejeitar H_0.

Exercício 2

Vazamentos de tanques de gasolina subterrâneos em postos de gasolina podem prejudicar o meio ambiente. Em levantamento feito há dois anos, constatou-se que 30% desses tanques apresentavam vazamento. Após um programa de conscientização de donos de postos, a Secretaria de Meio Ambiente selecionou uma amostra de 80 postos e constatou que 18 apresentavam vazamento.

Solução

(a) Formule as hipóteses adequadas, especificando o parâmetro a ser testado.

O parâmetro de interesse é a proporção de tanques de gasolina subterrâneos que apresentam vazamento após o programa de conscientização. Vamos denotá-lo por p.

As hipóteses de interesse são:

$$\mathrm{H}_0 : p = 0,30 \quad \text{versus} \quad \mathrm{H}_1 : p < 0,30.$$

As hipóteses possuem a seguinte interpretação:

H_0 : A proporção de tanques de gasolina subterrâneos que apresentam vazamento após o programa de conscientização não difere da proporção antes do programa de conscientização.

H_1 : A proporção de tanques de gasolina subterrâneos que apresentam vazamento após o programa de conscientização diminuiu, ou seja, o programa de conscientização surtiu efeito.

(b) Pode-se concluir que o programa teve efeito, para α= 6%?

Com base na hipótese alternativa formulada no item (a), para testar as hipóteses

$$\mathrm{H}_0 : p = 0,30 \quad \text{versus} \quad \mathrm{H}_1 : p < 0,30$$

a região crítica é da forma

$$\mathrm{RC} = \{\widehat{p} : \widehat{p} \leq a\},$$

em que $\widehat{p}$ é a proporção amostral e a é um valor de referência que pode ser obtido da distribuição amostral de $\widehat{p}$, para um dado nível de significância α.

Sendo a probabilidade do erro tipo I igual a $\alpha = 0,06$, temos que

$$\mathrm{P}(\widehat{p} \in \mathrm{RC}, \text{ sendo } \mathrm{H}_0 \text{ verdadeira}) = \mathrm{P}(\widehat{p} \leq a, \text{ sendo } p = 0,30) = 0,06.$$

Utilizando o Teorema Limite Central para aproximar a distribuição amostral de $\widehat{p}$, temos que

$$\widehat{p} \sim \text{Normal}\left(p, \frac{p(1-p)}{n}\right), \quad \text{aproximadamente},$$

e, consequentemente,

$$Z = \frac{\widehat{p} - p}{\sqrt{\dfrac{p \times (1-p)}{n}}} \sim \text{Normal}\,(0,1), \quad \text{aproximadamente}.$$

Assim, sob a hipótese H_0,

$$
\begin{aligned}
&\mathrm{P}(\widehat{p} \leq a, \text{ sendo } p = 0,30) = 0,06 \\
&\Rightarrow \mathrm{P}\left(\underbrace{\frac{\widehat{p} - 0,30}{\sqrt{\frac{0,30 \times (1-0,30)}{80}}}}_{Z} \leq \underbrace{\frac{a - 0,30}{\sqrt{\frac{0,30 \times (1-0,30)}{80}}}}_{z}\right) = 0,06 \\
&\Rightarrow \mathrm{P}\left(Z \leq z\right) = 0,06,
\end{aligned}
$$

em que $Z \sim \mathrm{N}(0,1)$, e z é uma constante tal que $\mathrm{P}\left(Z \leq z\right) = 0,06$. Utilizando a tabela fornecida no e-disciplinas obtemos que $z = -1,56$.

Obtemos então que

$$\frac{a - 0,30}{\sqrt{\frac{0,30\times(1-0,30)}{80}}} = -1,56,$$

logo

$$a = 0,30 - 1,56 \times \sqrt{\frac{0,30 \times (1-0,30)}{80}} = 0,30 - 1,56 \times 0,0512 = 0,2201.$$

Decisão: RC $= \{\widehat{p} : \widehat{p} \leq 0,2201\}$. Como $\widehat{p}_{obs} = \frac{18}{80} = 0,225$ não está dentro da região crítica, então decidimos não rejeitar H_0, a um nível de significância de 6%.

Conclusão: Ou seja, a um nível de significância de 6%, concluímos que não há evidências suficientes para afirmar que o programa de conscientização surtiu efeito.

(c) Pode-se concluir que o programa surtiu efeito, para α= 9%?

Aproveitando o procedimento realizado no item anterior, temos agora que, sob a hipótese H_0,

$$
\begin{aligned}
&\mathrm{P}(\widehat{p} \leq a, \text{ sendo } p = 0,30) = 0,09 \\
&\Rightarrow \mathrm{P}\left(\underbrace{\frac{\widehat{p} - 0,30}{\sqrt{\frac{0,30 \times (1-0,30)}{80}}}}_{Z} \leq \underbrace{\frac{a - 0,30}{\sqrt{\frac{0,30 \times (1-0,30)}{80}}}}_{z}\right) = 0,09 \\
&\Rightarrow \mathrm{P}\left(Z \leq z\right) = 0,09,
\end{aligned}
$$

em que $Z \sim \mathrm{N}(0,1)$, e z é uma constante tal que $\mathrm{P}\left(Z \leq z\right) = 0,09$. Utilizando a tabela fornecida no e-disciplinas obtemos que $z = -1,34$.

Obtemos então que

$$\frac{a - 0,30}{\sqrt{\frac{0,30\times(1-0,30)}{80}}} = -1,34,$$

logo

$$a = 0,30 - 1,34 \times \sqrt{\frac{0,30 \times (1-0,30)}{80}} = 0,30 - 1,34 \times 0,0512 = 0,2314.$$

Decisão: RC $= \{\widehat{p} : \widehat{p} \leq 0,2314\}$. Como $\widehat{p}_{obs} = \frac{18}{80} = 0,225$ está dentro da região crítica, então decidimos rejeitar H_0, a um nível de significância de 9%.

Conclusão: Ou seja, a um nível de significância de 9%, concluímos que há evidências suficientes para afirmar que o programa de conscientização surtiu efeito.

(d) Compare os resultados dos itens (b) e (c). Discuta

As conclusões dos testes nos itens (b) e (c) diferem. Do item (b) para o item (c) o tamanho da amostra permaneceu o mesmo, mas o nível de significância aumentou, fazendo com que o tamanho da região crítica aumentasse. Logo, ao nível de significância de 6% não conseguimos rejeitar H_0, mas ao nível de significância de 9% a hipótese nula foi rejeitada.

Exercício 3

Segundo dados da Anatel (www.anatel.gov.br/acessos-tv-por-assinatura), em 2015, 42% dos domicílios da região sudeste tinham TV por assinatura. Nos últimos anos, houve um aumento no interesse em canais de esporte de TV por assinatura; mas por outro lado, filmes e séries são acessados via "streaming". Portanto, não se sabe se o porcentual de domicílios que têm TV por assinatura em 2018 aumentou ou diminuiu comparado com 2015. Para saber se houve alteração nessa porcentagem, uma pesquisa será realizada selecionando-se ao acaso 400 domicílios na região sudeste.

Solução

(a) Defina as hipóteses estatísticas adequadas ao problema. Especifique o parâmetro a ser testado e seu significado no problema.

O parâmetro de interesse é a proporção de domicílios da região sudeste com TV por assinatura em 2018. Vamos denotá-lo por p.

As hipóteses de interesse são:

$$\text{H}_0 : p = 0,42 \quad \text{versus} \quad \text{H}_1 : p \neq 0,42.$$

As hipóteses possuem a seguinte interpretação:

H_0 : A proporção de domicílios da região sudeste com TV por assinatura em 2018 é a mesma que a proporção em 2015.

H_1 : A proporção de domicílios da região sudeste com TV por assinatura em 2018 é diferente da proporção de domicílios em 2015.

(b) Qual é a região crítica correspondente ao nível de significância de 10%?

Com base na hipótese alternativa formulada no item (a), para testar as hipóteses

$$\text{H}_0 : p = 0,42 \quad \text{versus} \quad \text{H}_1 : p \neq 0,42$$

a região crítica é da forma

$$\text{RC} = \left\{\widehat{p} : \widehat{p} \leq a_1 \text{ ou } \widehat{p} \geq a_2\right\},$$

em que $\widehat{p}$ é a proporção amostral e a_1 e a_2 são valores de referência que podem ser obtidos da distribuição amostral de $\widehat{p}$, para um dado nível de significância α.

Sendo a probabilidade do erro tipo I igual a $\alpha = 0,10$, temos que

$$\text{P}(\widehat{p} \in \text{RC, sendo H}_0 \text{ verdadeira}) = \text{P}(\widehat{p} \leq a_1 \text{ ou } \widehat{p} \geq a_2, \text{ sendo } p = 0,42) = 0,10.$$

Utilizando o Teorema Limite Central para aproximar a distribuição amostral de $\widehat{p}$, temos que

$$\widehat{p} \sim \text{Normal}\left(p, \frac{p(1-p)}{n}\right), \quad \text{aproximadamente},$$

e, consequentemente,

$$Z = \frac{\widehat{p} - p}{\sqrt{\frac{p \times (1-p)}{n}}} \sim \text{Normal}\left(0,1\right), \quad \text{aproximadamente}.$$

Assim, sob a hipótese H_0,

$$\text{P}(\widehat{p} \leq a_1 \text{ ou } \widehat{p} \geq a_2, \text{ sendo } p = 0,42) = 0,10$$

$$\Rightarrow \text{P}\left(\underbrace{\frac{\widehat{p} - 0,42}{\sqrt{\frac{0,42 \times (1-0,42)}{400}}}}_{Z} \leq \underbrace{\frac{a_1 - 0,42}{\sqrt{\frac{0,42 \times (1-0,42)}{400}}}}_{z_1} \text{ ou } \underbrace{\frac{\widehat{p} - 0,42}{\sqrt{\frac{0,42 \times (1-0,42)}{400}}}}_{Z} \geq \underbrace{\frac{a_2 - 0,42}{\sqrt{\frac{0,42 \times (1-0,42)}{400}}}}_{z_2}\right) = 0,10$$

$$\Rightarrow \text{P}\left(Z \leq z_1 \text{ ou } Z \geq z_2\right) = 0,10,$$

em que $Z \sim \text{N}(0,1)$, e z_1 e z_2 são constantes tais que $\text{P}\left(Z \leq z_1 \text{ ou } Z \geq z_2\right) = 0,10$. Note que z_1 e z_2 também são tais que

$$\text{P}\left(z_1 < Z < z_2\right) = 0,90,$$

e que devido a simetria da distribuição normal, podemos utilizar pesos iguais nas caudas de forma que que $z_1 = -z_2$, e portanto chegamos a

$$\text{P}\left(-z_2 < Z < z_2\right) = 0,90.$$

Utilizando a tabela fornecida no e-disciplinas obtemos que $z_2 = 1,64$. Assim,

$$\frac{a_1 - 0,42}{\sqrt{\frac{0,42\times(1-0,42)}{400}}} = -1,64, \quad \text{e} \quad \frac{a_2 - 0,42}{\sqrt{\frac{0,42\times(1-0,42)}{400}}} = 1,64,$$

logo

$$a_1 = 0,42 - 1,64 \times \sqrt{\frac{0,42\times(1-0,42)}{400}} = 0,42 - 1,64 \times 0,02468 = 0,3795,$$

e

$$a_2 = 0,42 + 1,64 \times \sqrt{\frac{0,42\times(1-0,42)}{400}} = 0,42 + 1,64 \times 0,02468 = 0,4605.$$

Portanto, a região crítica é dada por

$$\text{RC} = \left\{\widehat{p} : \widehat{p} \leq 0,3795 \text{ ou } \widehat{p} \geq 0,4605\right\}.$$

(c) Qual será a conclusão se forem observados 150 domicílios com TV por assinatura na amostra?

Decisão: Como $\widehat{p}_{obs} = \frac{150}{400} = 0,375$ está dentro da região crítica, então decidimos rejeitar H_0 a um nível de significância de 10%.

Conclusão: Ao nível de significância de 10%, concluímos que há evidências suficientes de que a proporção de domicílios da região sudeste com TV por assinatura em 2018 é diferente da proporção de domicílios com TV por assinatura em 2015. Como a proporção estimada foi menor do que 42%, temos evidências suficientes que a proporção de domicílios com TV por assinatura em 2018 diminuiu em comparação à proporção em 2015.

(d) Obtenha um intervalo de confiança de 90% de confiança para a proporção de domicílios na região sudeste que tem TV por assinatura em 2018, considerando a informação no item (c)

No item (c) obtemos que a estimativa pontual para p é dada por

$$\widehat{p}_{obs} = \frac{150}{400} = 0,375$$

Utilizando o Teorema Limite Central para aproximar a distribuição amostral do estimador da proporção, um intervalo de confiança aproximado para p, com coeficiente de confiança γ é dado por

$$\text{IC}(p;\gamma) = \left[\widehat{p} - z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}};\ \ \widehat{p} + z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right],$$

em que z é o quantil da distribuição normal padrão tal que $\text{P}(-z \leq Z \leq z) = \gamma$, sendo γ o coeficiente de confiança e $Z \sim \text{N}(0,1)$.

Para $\gamma = 0,90$, temos que

$$\text{P}(-z \leq Z \leq z) = 0,90,$$

e do item (b), $z = 1,64$. Logo, o intervalo de confiança para p com 90% de confiança é dado por

$$\begin{aligned}
\text{IC}(p;90\%) &= \left[0,375 - 1,64\sqrt{\frac{0,375(1-0,375)}{400}};\ \ 0,375 + 1,64\sqrt{\frac{0,375(1-0,375)}{400}}\right] \\
&= [0,375 - 0,0397;\ \ 0,375 + 0,0397] \\
&= [0,3353;\ \ 0,4147]\,.
\end{aligned}$$

Note que o intervalo estimado é formado por valores inferiores a 0,42.