

Teste de hipótese para diferença de proporção: Um estudo sobre o consumo de rádio em 2018 e 2019, na cidade de Caruaru-PE

Damião F. Santos¹, Julianny A. Sá², Ewerton D. L. Dantas³, João Batista S. Filho⁴, Jonatha F. Farias⁵

¹Mestre em Estatística - UNB.

²Graduanda em Bacharelado em Estatística - UEPB. E-mail: juliannyaraujodesa@gmail.com

³Bacharel em Ciências Econômicas - UFCG. E-mail: ewertondyego@yahoo.com.br

⁴Bacharel em Estatística - UEPB. E-mail: jb_sf@hotmail.com

⁵Bacharel em Estatística - UEPB. E-mail: farias.jonatha22@gmail.com

Resumo: *Há quase um século a história do rádio no Brasil deu-se início com a primeira transmissão. Nos dias atuais, a mídia rádio detém um poder de anúncio e investimento que atraem algumas empresas interessadas em divulgar sua marca e como forma de constatar o poder de alcance utilizam-se de pesquisas de audiência. Dessa forma, a partir de duas pesquisas de audiência realizadas em Caruaru-PE, tem-se como objetivo desse trabalho a realização de um comparativo sobre a proporção de consumo de rádio na referida cidade, entre os anos de 2018 e 2019, utilizando-se de análise descritiva, intervalo de confiança e teste de hipóteses para diferença de proporção. O estudo foi realizado ao longo de uma semana em novembro de 2018 e em uma semana em setembro de 2019, em 27 localidades da referida cidade. Dessa forma, ao utilizar apenas análise descritiva, observou-se algumas diferenças numéricas, mas com o uso do intervalo de confiança e do teste z mostrou que essa diferença não é estatisticamente significativa.*

Palavras-chave: Pesquisa de audiência de rádio; Intervalo de confiança para diferença de proporção; Teste z para diferença de proporção .

Abstract: *Almost a century ago the history of the radio in Brazil has begun with your first transmission. Nowadays, radio holds a advertising that attracts some companies interested in promoting your brand and as a means of ascertaining the reach power, audience surveys are used. In this way, from two radio audience surveys accomplished in Caruaru-PE, this article goal is a comparison about the radio consumption proportion in this city, between 2018 and 2019. By using descriptive analysis, confidence interval and hypothesis test to measure the difference of proportion. This study was made in a entire week in November 2018 and another one in September 2019, in 27 places of the city. In this way, by using descriptive analysis it was observed some numerical differences, but using confidence interval and z test proves that difference has no statistical difference.*

Keywords: Radio audience survey; Confidence interval for proportion difference; z test for proportion difference.

¹Autor correspondente: d.flaviostate@gmail.com.

Introdução

A história do rádio no Brasil, segundo Haussen (2004) começa em 1922 e continua até os dias de hoje, diariamente atingindo milhões de pessoas com transmissões de notícias, músicas e propagandas, fazendo parte do cotidiano do brasileiro há quase 100 anos.

O modelo comercial dessa mídia é definido à partir do financiamento junto ao setor privado de algumas empresas que anunciam seus produtos ou serviços nesse meio e com isso, atingem o público consumidor de rádio. Para validar esses resultados, são realizadas pesquisas de audiência que trarão não somente o nível de consumo de maneira geral, mas também o perfil e hábitos dos consumidores.

De acordo com o CENP (2019), em um sistema do Conselho Executivo de Normas-Padrão que traz informações acerca do investimento em mídia via agências de publicidade, atualmente reúne dados de 78 agências, informa acerca do volume financeiro total, bem como da participação de cada um desses meios. O mesmo mostra que em 2018, a TV aberta foi responsável por R\$ 9.642.777.000 e o rádio deteve R\$ 659.877.000 de todo o investimento do ano.

Pensando nisso, foram realizadas duas pesquisas de audiência na cidade de Caruaru-PE em 2018 e em 2019, sendo observado o nível de consumo de rádio ou a proporção de ouvintes de rádio. Dessa forma, o objetivo desse trabalho é analisar se existe ou não diferença estatisticamente significativa entre o consumo observado nas duas pesquisas.

Material e métodos

Neste estudo analisou-se o consumo de rádio na cidade de Caruaru-PE nos anos de 2018 e 2019. O consumo de rádio trata-se da terminologia utilizada em relação a proporção (\hat{p}) de pessoas que ouvem rádio durante o dia, independente de quanto tempo ocorreu o evento. Os dados refere-se a um estudo realizado pela empresa 6 Sigma Pesquisa e Consultoria Estatística - LTDA, em 27 bairros/localidades da cidade de Caruaru-PE, sendo coletadas ao longo de uma semana, 2.667 informações em novembro 2018 e 2.814 em setembro de 2019.

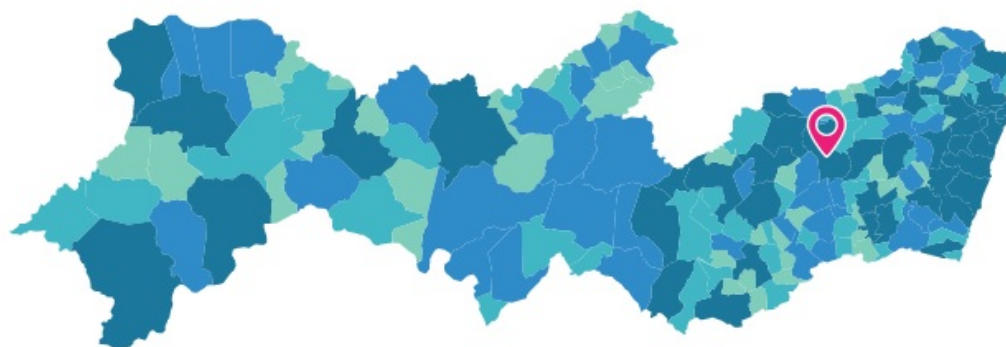


Figura 1: Localização do município de Caruaru no estado de Pernambuco.

Fonte: IBGE (2019)

No intuito de verificar se houve alguma diferença significativa - não importando que seja negativa ou positiva - entre o consumo de rádio na referida cidade entre setembro de 2019 e novembro de 2018, neste trabalho propõe-se o uso de um teste de hipóteses para a diferença de proporção, sendo essa, verificada ao longo dos sete dias da semana. Antes disso, será construído o intervalo de confiança para a diferença das proporções, que servirá como uma outra forma de testar as hipóteses à seguir:

$$\begin{cases} H_0 : \hat{p}_2 - \hat{p}_1 = 0 & \text{(Não existe diferença estatisticamente significativa),} \\ H_1 : \hat{p}_2 - \hat{p}_1 \neq 0 & \text{(Existe diferença estatisticamente significativa).} \end{cases}$$

Para a aplicação de tais técnicas, utilizou-se o *software* livre R (CORE TEAM. 2019).

Intervalo de confiança para a diferença de proporção

Ao estimar a diferença de proporção de um determinado evento de interesse em uma amostra com relação a proporção do mesmo evento em uma segunda amostra, torna-se necessário a utilização de métodos de inferência estatística, como a construção de um intervalo de confiança para tal estimativa. Esse intervalo de confiança (IC) se baseia na distribuição Binomial, que ao analisar uma amostra suficientemente grande e que os pressupostos da aproximação normal sejam respeitados, é possível utilizar-se a distribuição $N(0, 1)$ para a obtenção de tal intervalo.

Dessa forma, o intervalo de confiança para a diferença entre duas proporções é dado por:

$$IC(\hat{p}_2 - \hat{p}_1) = \hat{p}_2 - \hat{p}_1 \pm z_\alpha \sqrt{\left(\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2} \right)}, \quad (1)$$

em que α é o nível de significância utilizado no teste, $\hat{p}_1 = \frac{x_1}{n_1}$ e $\hat{p}_2 = \frac{x_2}{n_2}$, sendo x_1 e x_2 o número de elementos pertencente ao evento de interesse (consumo de rádio) nas duas amostras analisadas e n_1 e n_2 o tamanho das duas amostras. Usualmente é utilizado o nível de significância $\alpha = 0,05$, que também será adotado nesse trabalho. Ressalta-se que ao observar a presença de valor zero dentro do IC, sugere-se que não há diferença estatisticamente significativa entre as duas proporções.

Teste de Hipótese para a diferença de proporção

Uma outra maneira de verificar se existe ou não, diferença estatisticamente significativa entre duas proporções de amostras suficientemente grande e independentes, é a construção do teste z , que é dado pela seguinte expressão:

$$z = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_2} + \frac{1}{n_1} \right)}}, \quad (2)$$

em que $\hat{p} = \frac{(x_1 + x_2)}{(n_1 + n_2)}$.

De acordo com Vieira (2010), sob a hipótese de nulidade, a estatística $z \sim N(0, 1)$ se $n\hat{p} > 5$ e $n(1 - \hat{p}) > 5$, dessa forma, o teste z é recomendado apenas para amostras suficientemente grande, de tal modo que, a especificação seja atendida. A regra de decisão para o teste bilateral, com nível de significância $\alpha = 0,05$ é dada por:

$$\begin{cases} \text{Rejeita-se } H_0 : \hat{p}_2 - \hat{p}_1 = 0 & \text{se } z \leq -1,96 \text{ ou se } z \geq 1,96, \\ \text{Não rejeita-se } H_0 : \hat{p}_2 - \hat{p}_1 = 0 & \text{se } -1,96 < z < 1,96. \end{cases}$$

Além disso, ao utilizar o software R por meio da função `prop.test()`, obtém-se o “valor-p”, que trata-se de uma probabilidade da ocorrência de uma diferença tão grande ou maior do que a observada. A regra de decisão com o uso do “valor-p” é dada da seguinte maneira:

$$\begin{cases} \text{Rejeita-se } H_0 : \hat{p}_2 - \hat{p}_1 = 0 & \text{se } \text{valor-p} \leq \alpha, \\ \text{Não rejeita-se } H_0 : \hat{p}_2 - \hat{p}_1 = 0 & \text{se } \text{valor-p} > \alpha, \end{cases}$$

sendo α o nível de significância previamente estabelecido.

Resultados e discussões

Inicialmente é utilizada uma análise descritiva com o uso de gráfico de colunas empilhadas na Figura 2, com a distribuição comparativa da proporção de consumo de rádio durante o ano de 2018 e 2019, sendo segmentada pelos sete dias da semana.

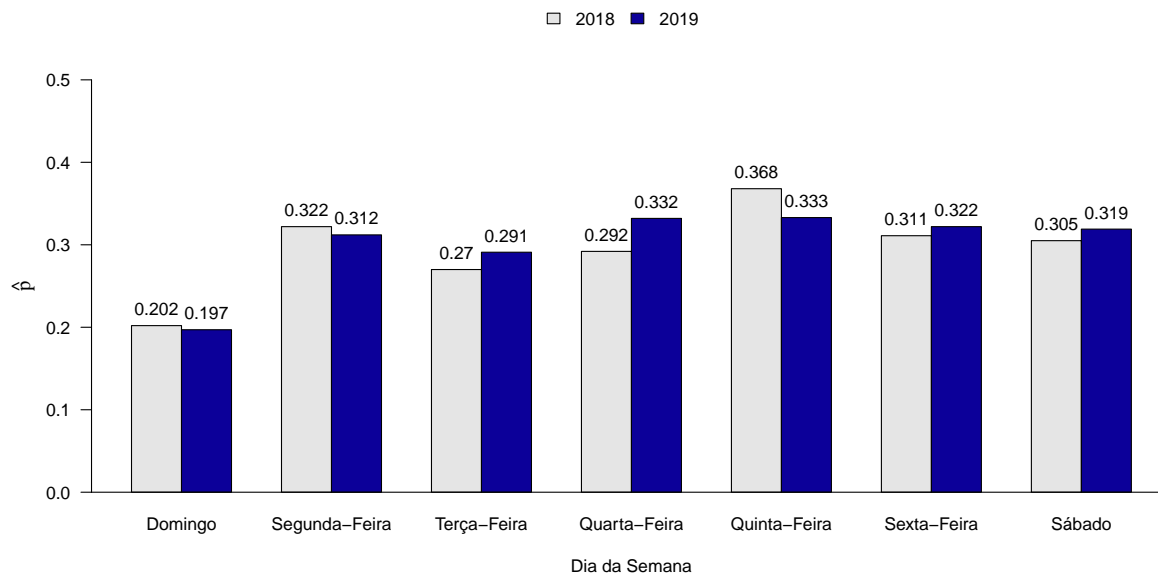


Figura 2: Proporção de consumo de rádio em Caruaru-PE nos anos de 2018 e 2019.

Em seguida é calculada a diferença entre a proporção de consumo durante o ano de 2019 em relação ao ano de 2018 e apresentada na Figura 3. Observa-se de forma descritiva, que houve 3 dias ao longo da semana em que o consumo de 2019 em relação ao de 2018 caiu, mas também, houve 4 dias que o consumo cresceu.

No entanto, apenas de forma descritiva não há como concluir que houve diferença ou não, para isso utilizou-se os intervalos de 95% confiança para tais diferenças e também aplicou-se o teste z ao nível de significância $\alpha = 0,05$.

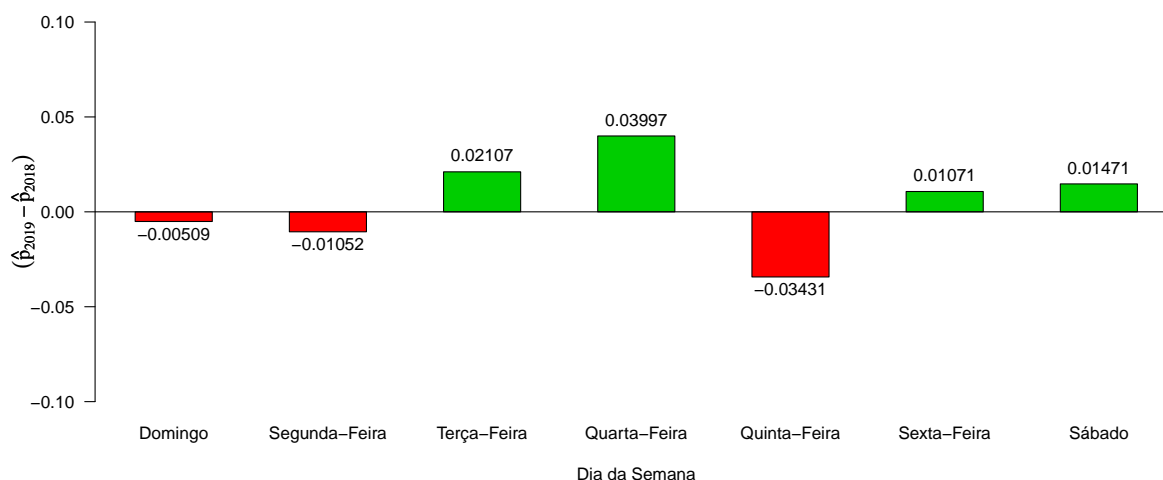


Figura 3: Diferença entre a proporção de consumo de rádio em 2019 e 2018 em Caruaru-PE.

Como observado na Tabela 1 em todos os sete dias analisados, não houve nenhuma diferença estatisticamente significativa, utilizando-se o nível de significância $\alpha = 0,05$. Essa conclusão se dá, primeiramente, ao observar que o valor zero está contido em todos os intervalos de 95% de confiança. Em seguida, ao observar que $-1,96 < z < 1,96$ e $\text{valor-p} > 0,05$, obtém a mesma conclusão.

Tabela 1: Teste de hipótese para a diferença de proporção de consumo de rádio entre 2019 e 2018 na cidade de Caruaru-PE.

Dia da semana	Consumo	2018	2019	$\hat{p}_{2019} - \hat{p}_{2018}$	$IC_{95\%}$	z	Valor-p
Domingo	\hat{p}	77 (0,2021)	79 (0,1970)	-0,00509	[-0,05096;0,06114]	-0,17811	0,85863
	$1-\hat{p}$	304 (0,7979)	322 (0,8030)				
Segunda-Feira	\hat{p}	118 (0,3224)	126 (0,3119)	-0,01052	[-0,05531;0,07635]	-0,31342	0,75396
	$1-\hat{p}$	248 (0,6776)	278 (0,6881)				
Terça-Feira	\hat{p}	98 (0,2700)	117 (0,2910)	0,02107	[-0,08477;0,04263]	0,64745	0,51734
	$1-\hat{p}$	265 (0,7300)	285 (0,7090)				
Quarta-Feira	\hat{p}	124 (0,2918)	138 (0,3317)	0,03997	[-0,10250;0,02260]	1,25124	0,21085
	$1-\hat{p}$	301 (0,7082)	278 (0,6683)				
Quinta-Feira	\hat{p}	150 (0,3676)	134 (0,3333)	-0,03431	[-0,03130;0,09998]	-1,02329	0,30617
	$1-\hat{p}$	258 (0,6324)	268 (0,6667)				
Sexta-Feira	\hat{p}	115 (0,3108)	127 (0,3215)	0,01071	[-0,07663;0,05521]	0,31826	0,75028
	$1-\hat{p}$	255 (0,6892)	268 (0,6785)				
Sábado	\hat{p}	108 (0,3051)	126 (0,3198)	0,01471	[-0,08121;0,05178]	0,43330	0,66480
	$1-\hat{p}$	246 (0,6949)	268 (0,6802)				

É importante destacar que, apesar da diferença numérica entre as proporções analisadas na quarta-feira e na quinta-feira ser relativamente maior do que nos outros dias, isso não foi suficiente para rejeitar a hipótese nula, apenas gerou uma probabilidade menor de ocorrência de uma diferença tão grande ou maior do que a observada.

Conclusão

Observou-se através das técnicas de inferência estatística que não houve diferença estatisticamente significativas entre a proporção de consumo de rádio no ano de 2019 em comparação com 2018. Ao utilizar a análise gráfica dos dados e apesar de observar diferença numérica maior nas observações da quarta-feira e quinta-feira, foi primordial o uso do intervalo de confiança e do teste z para identificar que essa diferença não é estatisticamente significativa. Desta forma, os objetivos foram alcançados, pois, na prática, esse resultado mostrou que mesmo com 10 meses de diferença entre a primeira coleta (novembro de 2018) e a segunda (setembro de 2019) não houve nenhuma diferença não aceitável nos níveis de consumo de rádio na cidade de Caruaru-PE.

Agradecimentos

Os autores agradecem à 6 Sigma Pesquisa e Consultoria Estatística - LTDA, pela disponibilização do banco de dados referente à esse estudo.

Referências Bibliográficas

CENP. CENP-Meios. Disponível em: <https://www.cenp.com.br/cenp-meios?id=5>. Acesso em: 16 set. 2019.

HAUSSEN, D.F. *Rádio brasileiro: uma história de cultura, política e integração*. Rádio sintonia do futuro, v.1, p. 51-62, 2004.

IBGE. IBGE-cidades @. Disponível em:

<https://cidades.ibge.gov.br/brasil/pe/caruaru/panorama>. Acesso em: 16 set. 2019.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL <https://www.R-project.org/>.

VIEIRA, SONIA, 1942 - *Bioestatística: tópicos avançados*. 3^a ed. Rio de Janeiro: Elsevier, 2010. 58p.