

FIAP

NBA

MBA em DATA SCIENCE & ARTIFICIAL INTELLIGENCE

APPLIED STATISTICS



Dra. Regina Tomie Ivata Bernal

Cientista de Dados na área da Saúde

Formação Acadêmica:

Estatístico - UFSCar

Mestre em Saúde Pública – FSP/USP

Doutor em Ciências – Epidemiologia - FSP/USP

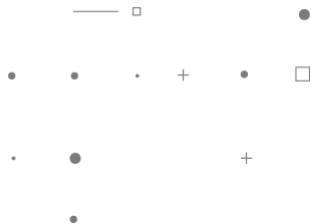
Atividades Profissionais:

Professora de pós-graduação na FIAP

Consultora externa da SVS/MS

Cientista de Dados em Saúde

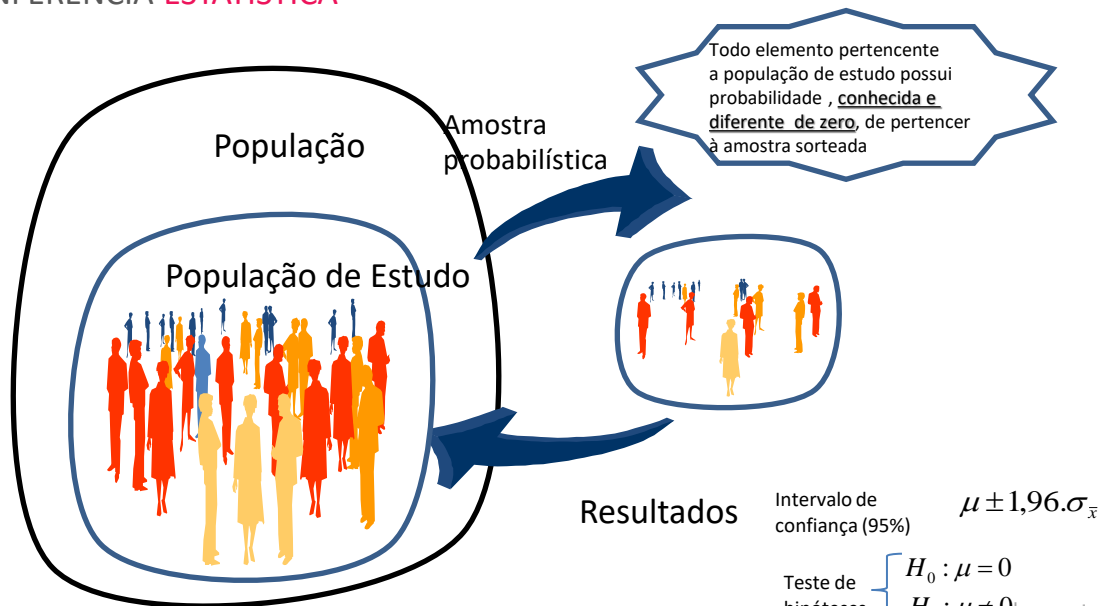
profregina.bernal@fiap.com.br
reginabernal@terra.com.br

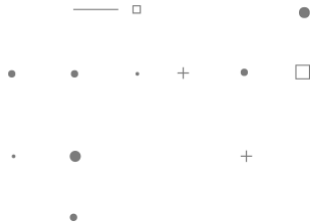


TESTE DE HIPÓTESES



INFERÊNCIA ESTATÍSTICA





TESTE DE HIPÓTESES

O que são Testes de Hipótese? - Bioestatística #9

<https://www.youtube.com/watch?v=FZHqrVyc104>



*o ruído e o ruído e o ruído e o ruído e o ruído e o
ruído e o ruído e o ruído e o ruído e o ruído e o ruído
e o ruído e o ruído e o ruído e o ruído e o ruído e o
ruído e o ruído e o ruído e o ruído e o ruído e o ruído
e o ruído e o ruído e o ruído e o ruído e o ruído e o
ruído e o ruído e o ruído e o ruído e o ruído e o ruído
e o ruído e o ruído e o ruído e o ruído e o ruído e o
ruído e o ruído e o ruído e o ruído e o ruído e o ruído
e o ruído e o ruído e o ruído e o insight e o ruído e o
ruído e o ruído e o ruído e o ruído e o ruído e o ruído
e o ruído e o ruído e o ruído e o ruído e o ruído e o
ruído e o ruído e o ruído e o ruído e o ruído e o ruído*

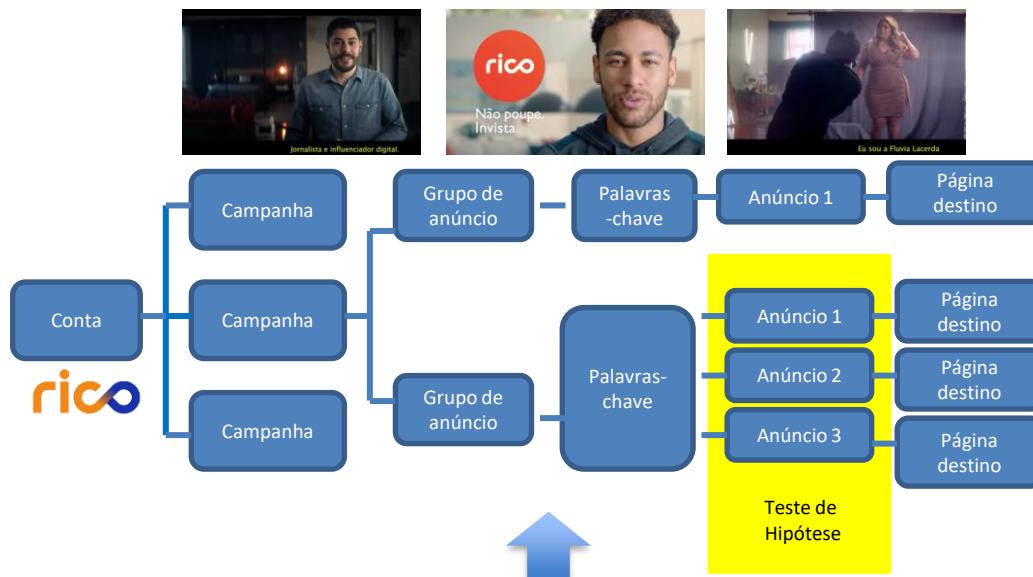
[illegible]

TESTE DE HIPÓTESES

Aplicado em Marketing

Planejamento de campanhas

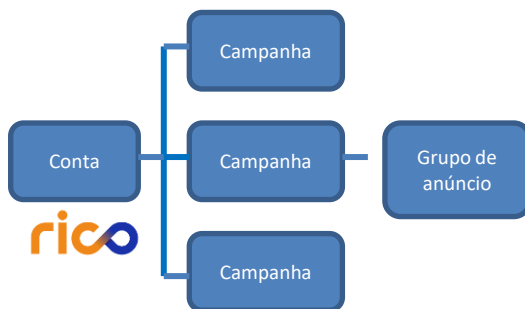
EXEMPLO



Não poupe. Invista. Abra a sua conta na Rico

Planejamento de campanhas

EXEMPLO



Página destino

Página destino

Página destino

Página destino

Teste de Hipótese



Não poupe. Invista.
Abra a sua conta na Rico

TESTE A/B

Origem: Wikipédia, a enciclopédia livre.

Teste A/B é um método de teste de [design](#) através do qual comparam-se elementos aleatórios com duas variantes, A e B, em que estes são o controle e o tratamento de uma experiência controlada, com o objetivo de melhorar a percentagem de aprovação. Estas experiências são muito utilizadas em desenvolvimento web e de marketing, e até mesmo em formas tradicionais de publicidade. Teste A/B também se designa por experiência aleatória controlada, experiência online controlada e teste de divisão. Em web design, o teste A/B é utilizado para identificar alterações nas páginas web que podem provocar mudanças positivas ou negativas no interesse dos utilizadores. Como o nome já diz, duas versões são comparadas, as quais são idênticas exceto por uma variante que pode impactar o comportamento do utilizador. A versão A pode ser a versão utilizada atualmente (controle), enquanto a Versão B é a modificada (tratamento). Podem ser vistas melhorias significativas através de testes de elementos como copiar o texto, layouts, imagens e cores, mas nem sempre. Os testes multivariados ou teste de balde são semelhantes ao teste A/B, mas estes testes abordam mais de duas versões diferentes ao mesmo tempo.

[1]

Referências

- ↑ «Split Testing Guide for Online Stores» [arquivado](#). *webics.com.au* (em inglês). Consultado em 17 de setembro de 2018

Fonte: https://pt.wikipedia.org/wiki/Teste_A/B

TESTE DE HIPÓTESE

Origem: Wikipédia, a enciclopédia livre.

Teste de hipóteses, **teste estatístico** ou **teste de significância**^[1] é um procedimento estatístico que permite tomar uma decisão (aceitar ou rejeitar a hipótese nula H_0) entre duas ou mais hipóteses (hipótese nula H_0 ou hipótese alternativa H_1), utilizando os dados observados de um determinado experimento.^[2] Há diversos métodos para realizar o teste de hipóteses, dos quais se destacam o método de Fisher (teste de significância),^[3] o método de Neyman–Pearson^[4] e o método de Bayes.^[5]

Por meio da teoria da probabilidade, é possível inferir sobre quantidades de interesse de uma população a partir de uma amostra observada de um experimento científico. Por exemplo, estimar pontualmente e de forma intervalar um parâmetro de interesse, testar se uma determinada teoria científica deve ser descartada, verificar se um lote de remédios deve ser devolvido por falta de qualidade, entre outros. Por meio do rigor matemático, a inferência estatística pode ser utilizada para auxiliar a tomada de decisões nas mais variadas áreas.^[6]

Os testes de hipóteses são utilizados para determinar quais resultados de um estudo científico podem levar à rejeição da hipótese nula H_0 a um nível de significância pré-estabelecido. O estudo da teoria das probabilidades e a determinação da estatística de teste correta são fundamentais para a coerência de um teste de hipótese. Se as hipóteses do teste de hipóteses não forem assumidas de maneira correta, o resultado será incorreto e a informação será incoerente com a questão do estudo científico. Os tipos conceituais de erro (erro do tipo I e erro do tipo II) e os limites paramétricos ajudam a distinguir entre a hipótese nula H_0 e a hipótese alternativa H_1 .^[7]

Fonte: https://pt.wikipedia.org/wiki/Testes_de_hipóteses

TESTE DE HIPÓTESES

Aplicado em Pesquisas

TESTE DE HIPÓTESES

EXEMPLO

Desejando-se conhecer a média de gasto anual com medicamentos na cidade Y, selecionou-se uma amostra aleatória de 100 adultos maiores de 40 anos. Teste a hipótese de que o gasto anual dessa população é inferior ao gasto médio de R\$ 120,00 a.a. com nível de significância de 5%?

Hipótese estatística:

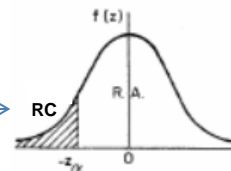
$$H_0 : \mu = 120$$

$$H_1 : \mu < 120$$

Critério de decisão:

$$\alpha = 0,05$$

$$z_{0,05} = 1,64$$



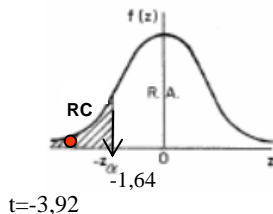
Sortear uma amostra aleatória de 100 adultos

TESTE DE HIPÓTESES

EXEMPLO

Resultados da pesquisa:

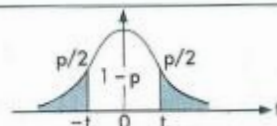
| | N | Mean | Std. Deviation | Std. Error Mean |
|-------|-----|-------|----------------|-----------------|
| gasto | 100 | 95,10 | 63,333 | 6,333 |



| Test Value = 120 | | | | | | |
|------------------|--------|----|-----------------|-----------------|---|--------|
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| gasto | -3,932 | 99 | ,000 | -24,900 | -37,47 | -12,33 |

Ao nível de 5% de significância, há evidências para rejeição de H_0 . Portanto, o gasto médio anual de medicamentos na população de adultos maiores de 40 anos residentes na cidade Y é inferior a R\$ 120.

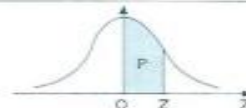
Tabela V — Distribuição t de Student
 Corpo da tabela dá os valores t_c tais que $P(-t_c < t < t_c) = 1 - p$.
 Para $v > 120$, usar a aproximação normal.



| Graus de liberdade v | | | | | | | | | | | | | | | | Graus de liberdade v |
|---------------------------|------------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|---------|---------|---------------------------|
| | $p = 90\%$ | 80% | 70% | 60% | 50% | 40% | 30% | 20% | 10% | 5% | 4% | 2% | 1% | 0,2% | 0,1% | |
| 1 | 0,158 | 0,325 | 0,510 | 0,727 | 1,000 | 1,376 | 1,963 | 3,078 | 6,314 | 12,706 | 15,894 | 31,821 | 63,657 | 318,309 | 636,619 | 1 |
| 2 | 0,142 | 0,289 | 0,445 | 0,617 | 0,816 | 1,061 | 1,386 | 1,886 | 2,920 | 4,303 | 4,849 | 6,965 | 9,925 | 22,327 | 31,598 | 2 |
| 3 | 0,137 | 0,277 | 0,424 | 0,584 | 0,765 | 0,978 | 1,250 | 1,638 | 2,353 | 3,182 | 3,482 | 4,541 | 5,841 | 10,214 | 12,924 | 3 |
| 4 | 0,134 | 0,271 | 0,414 | 0,569 | 0,741 | 0,941 | 1,190 | 1,533 | 2,132 | 2,776 | 2,998 | 3,747 | 4,604 | 7,173 | 8,610 | 4 |
| 5 | 0,132 | 0,267 | 0,408 | 0,559 | 0,727 | 0,920 | 1,156 | 1,476 | 2,015 | 2,571 | 2,756 | 3,365 | 4,032 | 5,893 | 6,869 | 5 |
| 6 | 0,131 | 0,265 | 0,404 | 0,553 | 0,718 | 0,906 | 1,134 | 1,440 | 1,943 | 2,447 | 2,612 | 3,143 | 3,707 | 5,208 | 5,959 | 6 |
| 7 | 0,130 | 0,263 | 0,402 | 0,549 | 0,711 | 0,896 | 1,119 | 1,415 | 1,895 | 2,365 | 2,517 | 2,998 | 3,499 | 4,785 | 5,408 | 7 |
| 8 | 0,130 | 0,262 | 0,399 | 0,546 | 0,706 | 0,889 | 1,108 | 1,397 | 1,860 | 2,306 | 2,449 | 2,896 | 3,355 | 4,501 | 5,041 | 8 |
| 9 | 0,129 | 0,261 | 0,398 | 0,543 | 0,703 | 0,883 | 1,100 | 1,383 | 1,833 | 2,262 | 2,398 | 2,821 | 3,250 | 4,297 | 4,781 | 9 |
| 10 | 0,129 | 0,260 | 0,397 | 0,542 | 0,700 | 0,879 | 1,093 | 1,372 | 1,812 | 2,228 | 2,359 | 2,764 | 3,169 | 4,144 | 4,587 | 10 |
| 11 | 0,129 | 0,260 | 0,396 | 0,540 | 0,697 | 0,876 | 1,088 | 1,363 | 1,796 | 2,201 | 2,328 | 2,718 | 3,106 | 3,025 | 4,437 | 11 |
| 12 | 0,128 | 0,259 | 0,395 | 0,539 | 0,695 | 0,873 | 1,083 | 1,356 | 1,782 | 2,179 | 2,303 | 2,681 | 3,055 | 3,930 | 4,318 | 12 |
| 13 | 0,128 | 0,259 | 0,394 | 0,538 | 0,694 | 0,870 | 1,079 | 1,350 | 1,771 | 2,160 | 2,282 | 2,650 | 3,012 | 3,852 | 4,221 | 13 |
| 14 | 0,128 | 0,258 | 0,393 | 0,537 | 0,692 | 0,868 | 1,076 | 1,345 | 1,761 | 2,145 | 2,264 | 2,624 | 2,977 | 3,787 | 4,140 | 14 |
| 15 | 0,128 | 0,258 | 0,393 | 0,536 | 0,691 | 0,866 | 1,074 | 1,341 | 1,753 | 2,131 | 2,248 | 2,602 | 2,947 | 3,733 | 4,073 | 15 |
| 16 | 0,128 | 0,258 | 0,392 | 0,535 | 0,690 | 0,865 | 1,071 | 1,337 | 1,746 | 2,120 | 2,235 | 2,583 | 2,921 | 3,686 | 4,015 | 16 |
| 17 | 0,128 | 0,257 | 0,392 | 0,534 | 0,689 | 0,863 | 1,069 | 1,333 | 1,740 | 2,110 | 2,224 | 2,567 | 2,898 | 3,646 | 3,965 | 17 |
| 18 | 0,127 | 0,257 | 0,392 | 0,534 | 0,688 | 0,862 | 1,067 | 1,330 | 1,734 | 2,101 | 2,214 | 2,552 | 2,878 | 3,610 | 3,922 | 18 |
| 19 | 0,127 | 0,257 | 0,391 | 0,533 | 0,688 | 0,861 | 1,066 | 1,328 | 1,729 | 2,093 | 2,205 | 2,539 | 2,861 | 3,579 | 3,883 | 19 |
| 20 | 0,127 | 0,257 | 0,391 | 0,533 | 0,687 | 0,860 | 1,064 | 1,325 | 1,725 | 2,086 | 2,197 | 2,528 | 2,845 | 3,552 | 3,850 | 20 |
| 21 | 0,127 | 0,257 | 0,391 | 0,532 | 0,686 | 0,859 | 1,063 | 1,323 | 1,721 | 2,080 | 2,189 | 2,518 | 2,831 | 3,527 | 3,819 | 21 |
| 22 | 0,127 | 0,256 | 0,390 | 0,532 | 0,686 | 0,858 | 1,061 | 1,321 | 1,717 | 2,074 | 2,183 | 2,508 | 2,819 | 3,505 | 3,792 | 22 |
| 23 | 0,127 | 0,256 | 0,390 | 0,532 | 0,685 | 0,858 | 1,060 | 1,319 | 1,714 | 2,069 | 2,177 | 2,500 | 2,807 | 3,485 | 3,768 | 23 |
| 24 | 0,127 | 0,256 | 0,390 | 0,531 | 0,685 | 0,857 | 1,059 | 1,318 | 1,711 | 2,064 | 2,172 | 2,492 | 2,797 | 3,467 | 3,745 | 24 |
| 25 | 0,127 | 0,256 | 0,390 | 0,531 | 0,684 | 0,856 | 1,058 | 1,316 | 1,708 | 2,060 | 2,166 | 2,485 | 2,787 | 3,450 | 3,725 | 25 |
| 26 | 0,127 | 0,256 | 0,390 | 0,531 | 0,684 | 0,856 | 1,058 | 1,315 | 1,706 | 2,056 | 2,162 | 2,479 | 2,779 | 3,435 | 3,707 | 26 |
| 27 | 0,127 | 0,256 | 0,389 | 0,531 | 0,684 | 0,855 | 1,057 | 1,314 | 1,703 | 2,052 | 2,158 | 2,473 | 2,771 | 3,421 | 3,690 | 27 |
| 28 | 0,127 | 0,256 | 0,389 | 0,530 | 0,684 | 0,855 | 1,056 | 1,313 | 1,701 | 2,048 | 2,154 | 2,467 | 2,763 | 3,408 | 3,674 | 28 |
| 29 | 0,127 | 0,256 | 0,389 | 0,530 | 0,683 | 0,854 | 1,055 | 1,311 | 1,699 | 2,045 | 2,150 | 2,462 | 2,756 | 3,396 | 3,659 | 29 |
| 30 | 0,127 | 0,256 | 0,389 | 0,530 | 0,683 | 0,854 | 1,055 | 1,310 | 1,697 | 2,042 | 2,147 | 2,457 | 2,750 | 3,385 | 3,646 | 30 |
| 35 | 0,126 | 0,255 | 0,388 | 0,529 | 0,682 | 0,852 | 1,052 | 1,306 | 1,690 | 2,030 | 2,133 | 2,438 | 2,724 | 3,340 | 3,591 | 35 |
| 40 | 0,126 | 0,255 | 0,388 | 0,529 | 0,681 | 0,851 | 1,050 | 1,303 | 1,684 | 2,021 | 2,123 | 2,423 | 2,704 | 3,307 | 3,551 | 40 |
| 50 | 0,126 | 0,254 | 0,387 | 0,528 | 0,679 | 0,849 | 1,047 | 1,299 | 1,676 | 2,009 | 2,109 | 2,403 | 2,678 | 3,261 | 3,496 | 50 |
| 60 | 0,126 | 0,254 | 0,387 | 0,527 | 0,679 | 0,848 | 1,045 | 1,296 | 1,671 | 2,000 | 2,099 | 2,390 | 2,660 | 3,232 | 3,460 | 60 |
| 120 | 0,126 | 0,254 | 0,386 | 0,526 | 0,677 | 0,845 | 1,041 | 1,289 | 1,658 | 1,980 | 2,076 | 2,358 | 2,617 | 3,160 | 3,373 | 120 |
| ∞ | 0,126 | 0,253 | 0,385 | 0,524 | 0,674 | 0,842 | 1,036 | 1,282 | 1,645 | 1,960 | 2,054 | 2,326 | 2,576 | 3,090 | 3,291 | ∞ |
| | $p = 90\%$ | 80% | 70% | 60% | 50% | 40% | 30% | 20% | 10% | 5% | 4% | 2% | 1% | 0,2% | 0,1% | |

DISTRIBUIÇÃO NORMAL

Tabela III — Distribuição Normal Padrão
 $Z \sim N(0, 1)$
 Corpo da tabela dá a probabilidade p , tal que $p = P(0 < Z < Z_c)$



$$P(Z > 0) = ?$$

| parte inteira e primeira decimal de Z_c | Segunda decimal de Z_c | | | | | | | | | | parte inteira e primeira decimal de Z_c |
|---|--------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 0,0 | p = 0 | 00399 | 00798 | 01197 | 01595 | 01994 | 02392 | 02790 | 03188 | 03586 | 0,0 |
| 0,1 | 03983 | 04380 | 04776 | 05172 | 05567 | 05962 | 06356 | 06749 | 07142 | 07535 | 0,1 |
| 0,2 | 07926 | 08317 | 08706 | 09095 | 09483 | 09871 | 10257 | 10642 | 11026 | 11409 | 0,2 |
| 0,3 | 11791 | 12172 | 12552 | 12930 | 13307 | 13683 | 14058 | 14431 | 14803 | 15173 | 0,3 |
| 0,4 | 15542 | 15910 | 16276 | 16640 | 17003 | 17364 | 17724 | 18082 | 18439 | 18793 | 0,4 |
| 0,5 | 19146 | 19497 | 19847 | 20194 | 20540 | 20884 | 21226 | 21566 | 21904 | 22240 | 0,5 |
| 0,6 | 22575 | 22907 | 23237 | 23565 | 23891 | 24215 | 24537 | 24857 | 25175 | 25490 | 0,6 |
| 0,7 | 25804 | 26115 | 26424 | 26730 | 27035 | 27337 | 27637 | 27935 | 28230 | 28524 | 0,7 |
| 0,8 | 28814 | 29103 | 29389 | 29673 | 29955 | 30234 | 30511 | 30785 | 31057 | 31327 | 0,8 |
| 0,9 | 31594 | 31859 | 32121 | 32381 | 32639 | 32894 | 33147 | 33398 | 33646 | 33891 | 0,9 |
| 1,0 | 34134 | 34375 | 34614 | 34850 | 35083 | 35314 | 35543 | 35769 | 35993 | 36214 | 1,0 |
| 1,1 | 36433 | 36650 | 36864 | 37076 | 37286 | 37493 | 37698 | 37900 | 38100 | 38298 | 1,1 |
| 1,2 | 38493 | 38686 | 38877 | 39065 | 39251 | 39435 | 39617 | 39796 | 39973 | 40147 | 1,2 |
| 1,3 | 40320 | 40490 | 40658 | 40824 | 40988 | 41149 | 41309 | 41466 | 41621 | 41774 | 1,3 |
| 1,4 | 41924 | 42073 | 42220 | 42364 | 42507 | 42647 | 42786 | 42922 | 43056 | 43189 | 1,4 |
| 1,5 | 43319 | 43448 | 43574 | 43699 | 43822 | 43943 | 44062 | 44179 | 44295 | 44408 | 1,5 |
| 1,6 | 44520 | 44630 | 44738 | 44845 | 44950 | 45053 | 45154 | 45254 | 45352 | 45449 | 1,6 |
| 1,7 | 45543 | 45637 | 45728 | 45818 | 45907 | 45994 | 46080 | 46164 | 46246 | 46327 | 1,7 |
| 1,8 | 46407 | 46485 | 46562 | 46638 | 46712 | 46784 | 46856 | 46926 | 46995 | 47062 | 1,8 |
| 1,9 | 47128 | 47193 | 47257 | 47320 | 47381 | 47441 | 47500 | 47558 | 47615 | 47670 | 1,9 |
| 2,0 | 47725 | 47778 | 47831 | 47882 | 47932 | 47982 | 48030 | 48077 | 48124 | 48169 | 2,0 |
| 2,1 | 48214 | 48257 | 48300 | 48341 | 48382 | 48422 | 48461 | 48500 | 48537 | 48574 | 2,1 |
| 2,2 | 48610 | 48645 | 48679 | 48713 | 48745 | 48778 | 48809 | 48840 | 48870 | 48899 | 2,2 |
| 2,3 | 48928 | 48956 | 48983 | 49010 | 49036 | 49061 | 49086 | 49111 | 49134 | 49158 | 2,3 |
| 2,4 | 49180 | 49202 | 49224 | 49245 | 49266 | 49286 | 49305 | 49324 | 49343 | 49361 | 2,4 |
| 2,5 | 49379 | 49396 | 49413 | 49430 | 49446 | 49461 | 49477 | 49492 | 49506 | 49520 | 2,5 |
| 2,6 | 49534 | 49547 | 49560 | 49573 | 49585 | 49598 | 49609 | 49621 | 49632 | 49643 | 2,6 |
| 2,7 | 49653 | 49664 | 49674 | 49683 | 49693 | 49702 | 49711 | 49720 | 49728 | 49736 | 2,7 |
| 2,8 | 49744 | 49752 | 49760 | 49767 | 49774 | 49781 | 49788 | 49795 | 49801 | 49807 | 2,8 |
| 2,9 | 49813 | 49819 | 49825 | 49831 | 49836 | 49841 | 49846 | 49851 | 49856 | 49861 | 2,9 |
| 3,0 | 49865 | 49869 | 49874 | 49878 | 49882 | 49886 | 49889 | 49893 | 49897 | 49900 | 3,0 |
| 3,1 | 49903 | 49906 | 49910 | 49913 | 49916 | 49918 | 49921 | 49924 | 49926 | 49929 | 3,1 |
| 3,2 | 49931 | 49934 | 49936 | 49938 | 49940 | 49942 | 49944 | 49946 | 49948 | 49950 | 3,2 |
| 3,3 | 49952 | 49953 | 49955 | 49957 | 49958 | 49960 | 49961 | 49962 | 49964 | 49965 | 3,3 |
| 3,4 | 49966 | 49968 | 49969 | 49970 | 49971 | 49972 | 49973 | 49974 | 49975 | 49976 | 3,4 |
| 3,5 | 49977 | 49978 | 49978 | 49979 | 49980 | 49981 | 49981 | 49982 | 49983 | 49983 | 3,5 |
| 3,6 | 49984 | 49985 | 49985 | 49986 | 49986 | 49987 | 49987 | 49988 | 49988 | 49989 | 3,6 |
| 3,7 | 49989 | 49990 | 49990 | 49990 | 49991 | 49991 | 49992 | 49992 | 49992 | 49992 | 3,7 |

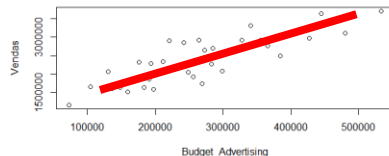
TESTE DE HIPÓTESES

Aplicado em Modelos de Regressão
Linear

Faça a previsão das vendas (R\$) mensal no período de 12 meses da empresa XYZ a partir dos dados disponíveis de Vendas (R\$) e Budget Advertising (R\$) da empresa.

EXEMPLO
Saída Python

| | Data | ano | Vendas | Budget_Advertising |
|----|--------|------|---------|--------------------|
| 1 | Jan/16 | 2016 | 1160081 | 72800 |
| 2 | fev/16 | 2016 | 1622540 | 123392 |
| 3 | mar/16 | 2016 | 1597260 | 135761 |
| 4 | abr/16 | 2016 | 1640675 | 140064 |
| 5 | mai/16 | 2016 | 1511270 | 159746 |
| 6 | Jun/16 | 2016 | 1634073 | 183353 |
| 7 | Jul/16 | 2016 | 1856971 | 190722 |
| 8 | ago/16 | 2016 | 1585566 | 197802 |
| 9 | set/16 | 2016 | 2041672 | 248891 |
| 10 | out/16 | 2016 | 1933557 | 256353 |
| 11 | nov/16 | 2016 | 2076910 | 298805 |
| 12 | dez/16 | 2016 | 1740202 | 268925 |



```
> mc = cor(dadosquant);mc
                                Vendas Budget_Advertising
Vendas                        1.000                      0.852
Budget_Advertising            0.852                      1.000
```

Modelo de regressão linear simples

$$Y = b_0 + b_1 \cdot X_1 \quad (\text{teórico})$$

```
import statsmodels.api as sm
from scipy import stats
from statsmodels.compat import lzip
from scipy.stats.stats import pearsonr
```

```
# Ordinary Least Square (OLS)
X_ = sm.add_constant(X)
est = sm.OLS(y, X_)
est2 = est.fit()
print(est2.summary())
```

Adiciona o intercepto
no modelo.



```
=====
                        OLS Regression Results
=====
Dep. Variable:          Vendas      R-squared:                0.725
Model:                  OLS         Adj. R-squared:           0.717
Method:                 Least Squares   F-statistic:             89.75
Date:                   Mon, 21 Jun 2021   Prob (F-statistic):      4.58e-11
Time:                   23:18:47         Log-Likelihood:         -509.57
No. Observations:       36             AIC:                   1023.
Df Residuals:           34             BIC:                   1026.
Df Model:                1
Covariance Type:        nonrobust
=====
                                coef    std err          t      P>|t|      [0.025    0.975]
-----
const                1.061e+06    1.52e+05     6.988    0.000    7.52e+05    1.37e+06
Budget_Advertising    4.9641      0.524     9.473    0.000    3.899      6.029
=====
Omnibus:                 1.236   Durbin-Watson:           0.781
Prob(Omnibus):           0.539   Jarque-Bera (JB):         1.112
Skew:                    0.256   Prob(JB):                 0.574
Kurtosis:                2.309   Cond. No.                 7.54e+05
=====
```

Faça a previsão das vendas (R\$) mensal no período de 12 meses da empresa XYZ a partir dos dados disponíveis de Vendas (R\$) e Budget Advertising (R\$) da empresa.

EXEMPLO
Saída Python

Hipótese estatística:

$$H_0 : B_0 = 0$$

$$H_1 : B_0 \neq 0$$

Critério de decisão:

$$n = 36$$

$$gl = n - 1$$

$$\alpha = 0,05$$

$$t_{0,05} = 2.030$$

(Tabela t-Student)

| OLS Regression Results | | | | | | |
|------------------------|------------------|---------------------|----------|---------|------------------------|----------|
| Dep. Variable: | Vendas | R-squared: | 0.725 | | | |
| Model: | OLS | Adj. R-squared: | 0.717 | | | |
| Method: | Least Squares | F-statistic: | 89.75 | | | |
| Date: | Mon, 21 Jun 2021 | Prob (F-statistic): | 4.58e-11 | | | |
| Time: | 23:18:47 | Log-Likelihood: | -509.57 | | | |
| No. Observations: | 36 | AIC: | 1023. | | | |
| Df Residuals: | 34 | BIC: | 1026. | | | |
| Df Model: | 1 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| | coef | std err | t | p-value | Intervalo de confiança | |
| | | | | P> t | [0.025 | 0.975] |
| const | 1.061e+06 | 1.52e+05 | 6.988 | 0.000 | 7.52e+05 | 1.37e+06 |
| Budget_Advertising | 4.9641 | 0.524 | 9.473 | 0.000 | 3.899 | 6.029 |
| Omnibus: | 1.236 | Durbin-Watson: | 0.781 | | | |
| Prob(Omnibus): | 0.539 | Jarque-Bera (JB) | 1.112 | | | |
| Skew: | 0.256 | Prob(JB): | 0.574 | | | |
| Kurtosis: | 2.309 | Cond. No. | 7.54e+05 | | | |

b0
b1

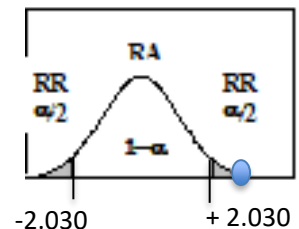
$$t_{observado} = 6.988$$

$$\text{Vendas} = b_0 + b_1 * \text{Budget}$$

(modelo ajustado)

$$\text{Vendas} = 1060550 + 4.964 * \text{Budget}$$

Teste Bilateral



Faça a previsão das vendas (R\$) mensal no período de 12 meses da empresa XYZ a partir dos dados disponíveis de Vendas (R\$) e Budget Advertising (R\$) da empresa.

EXEMPLO
Saída RStudio

```
modelo <- lm(Vendas ~ Budget_Advertising)
summary(modelo)
```

Projeção para 2019

Vendas Budget

| | | | |
|--------|------|------------------|--------|
| jan/19 | 2019 | 1.512.274 | 91000 |
| fev/19 | 2019 | | 154240 |
| mar/19 | 2019 | | 169702 |
| abr/19 | 2019 | | 185081 |
| mai/19 | 2019 | | 199683 |
| jun/19 | 2019 | | 229192 |
| jul/19 | 2019 | | 238403 |
| ago/19 | 2019 | | 247253 |
| set/19 | 2019 | | 311114 |
| out/19 | 2019 | | 320442 |
| nov/19 | 2019 | | 373507 |
| dez/19 | 2019 | | 336157 |

Modelo de regressão linear simples

```
> summary(modelo)
```

Call:

```
lm(formula = vendas ~ Budget_Advertising)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-655330 -256271 -30444  234875  743028
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1060550.396   151771.308    6.99 0.000000046312 ***
Budget_Advertising    4.964      0.524    9.47 0.000000000046 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 350000 on 34 degrees of freedom
Multiple R-squared:  0.725,    Adjusted R-squared:  0.717
F-statistic: 89.7 on 1 and 34 DF,  p-value: 0.0000000000458
```

$$\text{Vendas janeiro/19} = 1060550 + 4.964 * 91000 = 1.512.274$$

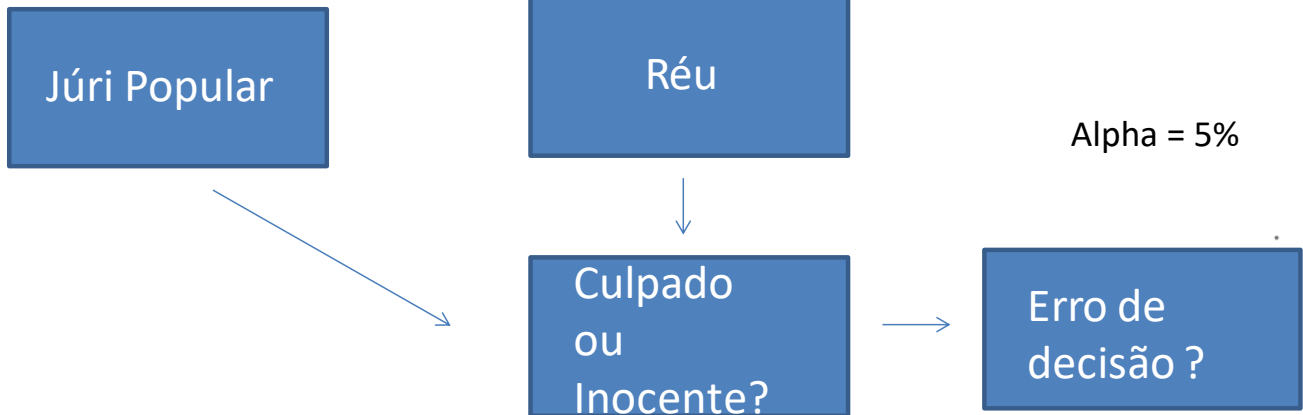
Teste de Hipóteses

Exemplo:

H_0 : Inocente

H_1 : Culpado

Alpha = 5%



Teste de Hipóteses

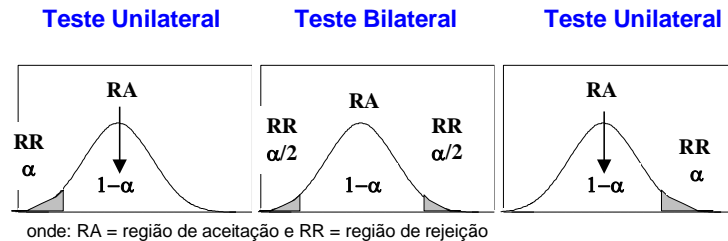
De acordo com a formulação das hipóteses, os testes podem ser monocaudal ou bicaudal.

- Monocaudal:

- $H_0: \mu = \mu_0$
- $H_1: \mu > \mu_0$
- $H_0: \mu = \mu_0$
- $H_1: \mu < \mu_0$

- Bicaudal

- $H_0: \mu = \mu_0$
- $H_1: \mu \neq \mu_0$



TESTE DE HIPÓTESES

1) A Empresa AAA deseja avaliar o impacto do novo produto no faturamento médio da empresa. Elabore as hipóteses H_0 e H_1 .

H_0 :

H_1 :

2) A indústria farmacêutica XYZ tem um medicamento BBB que apresenta um tempo médio de 10min para efeito do remédio. A empresa fez uma alteração na fórmula desse produto e ela deseja avaliar se lança ou não o novo produto no mercado. Elabore as hipóteses H_0 e H_1 .

H_0 :

H_1 :

Teste de Hipóteses

□ p-valor

É uma estimativa do nível de significância observada na amostra. Indica a probabilidade de ocorrer valores da estatística mais extremos do que o observado, sob a hipótese H_0 ser verdadeira.

Regra de decisão:

Se $p\text{-valor} < \alpha$ então rejeito H_0 , caso contrário não rejeito.

TESTE DE HIPÓTESES

Origem: Wikipédia, a enciclopédia livre.

São fundamentais os seguintes conceitos para um teste de hipóteses:^[7]

- **Hipótese nula** (H_0): é a hipótese assumida como verdadeira para a construção do teste. É a teoria, o efeito ou a alternativa que se está interessado em testar.
- **Hipótese alternativa** (H_1): é considerada quando a hipótese nula não tem evidência estatística.
- **Erro do tipo I** (α): é a probabilidade de se rejeitar a hipótese nula quando ela é verdadeira.
- **Erro do tipo II**: é a probabilidade de se rejeitar a hipótese alternativa quando ela é verdadeira.

| | Hipótese nula H_0 é verdadeira | Hipótese nula H_0 é falsa |
|-------------------------------------|----------------------------------|-----------------------------|
| Hipótese nula H_0 é rejeitada | Erro do tipo I | Não há erro |
| Hipótese nula H_0 não é rejeitada | Não há erro | Erro do tipo II |

Fonte: https://pt.wikipedia.org/wiki/Testes_de_hipóteses



TESTE DE HIPÓTESES

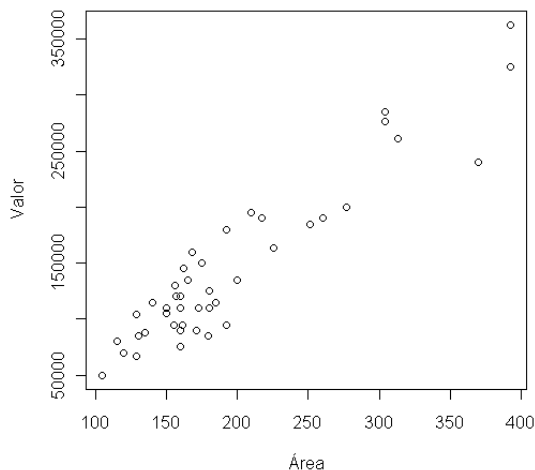


- ANÁLISE DE ASSOCIAÇÃO

Analisa o comportamento conjunto de duas variáveis qualitativas apresentada em tabela bivariada.

- Teste Qui-Quadrado (Variáveis Qualitativas)
- Correlação de Pearson (Variáveis Quantitativas)

Existe correlação entre o valor do imóvel e a área?



Teste Correlação de Pearson

$H_0: r = 0$ (ausência de correlação)

$H_1: r \neq 0$ (presença de correlação)

Erro de decisão: 0,05 ou 5%

R Console

```
> corr_t <- cor.test(Valor, Área, method="pearson", alternative="two.sided")
> corr_t
```

```
Pearson's product-moment correlation
```

```
data: Valor and Área
```

```
t = 17.0563, df = 41, p-value < 2.2e-16
```

```
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
```

```
0.8845899 0.9651600
```

```
sample estimates:
```

```
cor
```

```
0.9362024
```

Conclusão:

**Rejeito H_0 , portanto
há associação.**

CORRELAÇÃO DE PEARSON

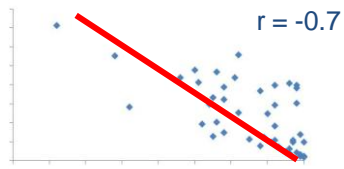
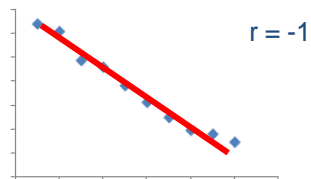
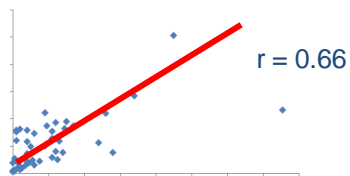
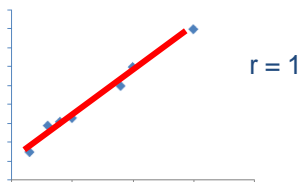
Correlação indica a força e a direção do relacionamento linear entre duas **variáveis aleatórias**. No uso estatístico geral, correlação se refere à medida da relação entre duas variáveis, embora correlação não implique **causalidade**. Nesse sentido geral, existem vários coeficientes medindo o grau de correlação, adaptados à natureza dos dados.

Vários coeficientes são utilizados para situações diferentes. O mais conhecido é o **coeficiente de correlação de Pearson**, o qual é obtido dividindo a **covariância** de duas variáveis pelo produto de seus **desvios padrão**. Apesar do nome, ela foi apresentada inicialmente por **Francis Galton**, em meados do século XVII.

Coeficiente de correlação de Pearson, em geral é expresso por (R ou ρ).

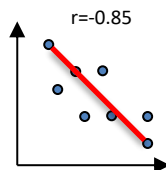
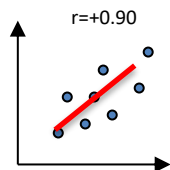
CORRELAÇÃO DE PEARSON

Análise de correlação



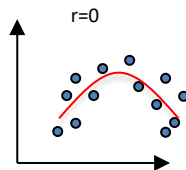
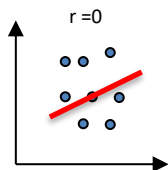
CORRELAÇÃO DE PEARSON

Análise de correlação



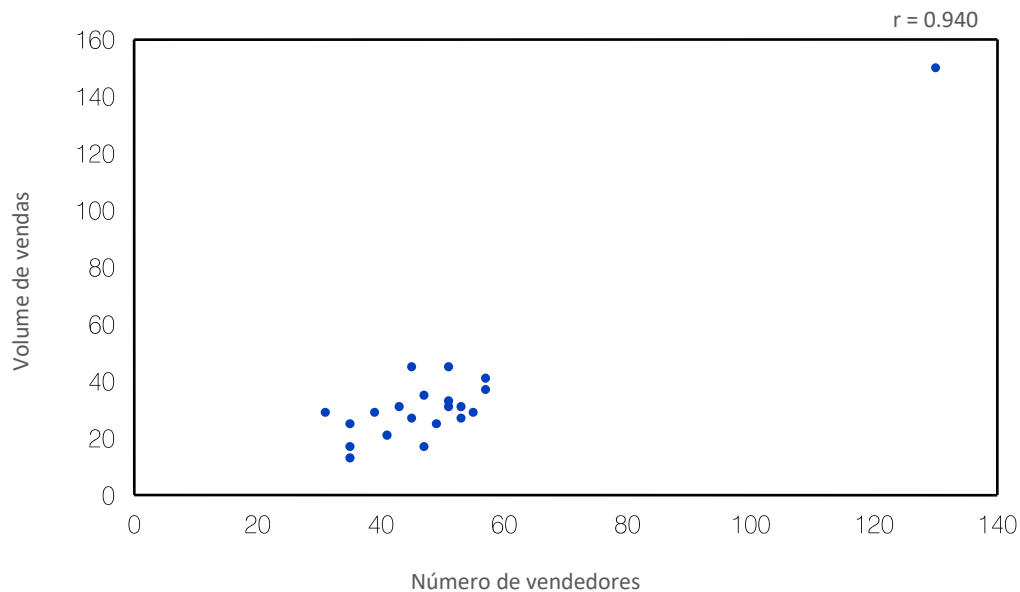
Correlação Linear Simples
(r de Pearson)

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 * \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

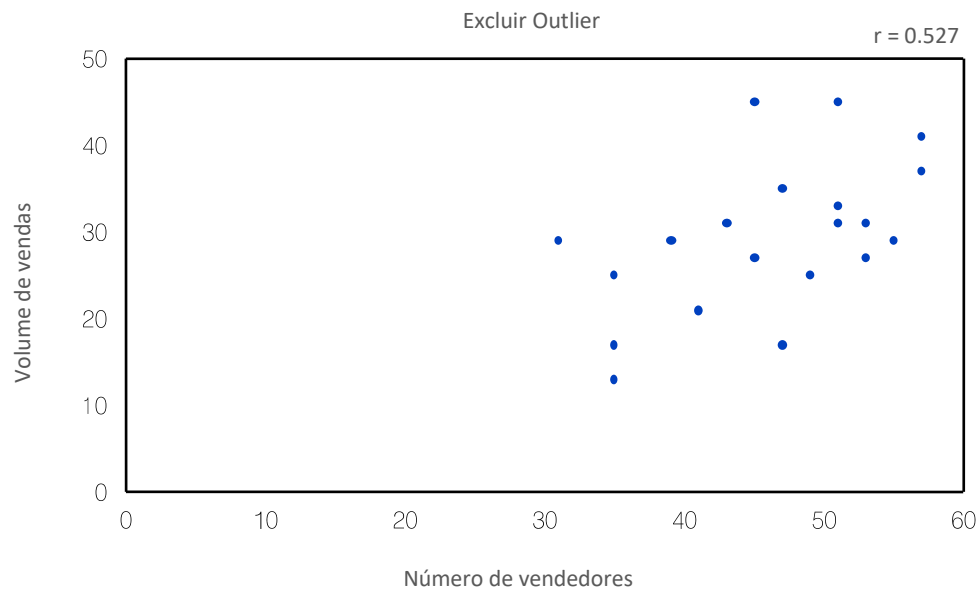


Para avaliar-se a correlação entre variáveis, é importante conhecer a magnitude ou força tanto quanto a significância da correlação.

Volume de vendas X Número de vendedores

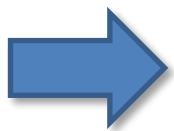


Volume de vendas X Número de vendedores



• • • + • □

- Associação⁺ entre dois fatores e quando queremos saber se um causa o outro ?
- Big data muitos resultados estatisticamente significativos que não fazem sentido causal
- variável de confusão quando há muitas variáveis na análise

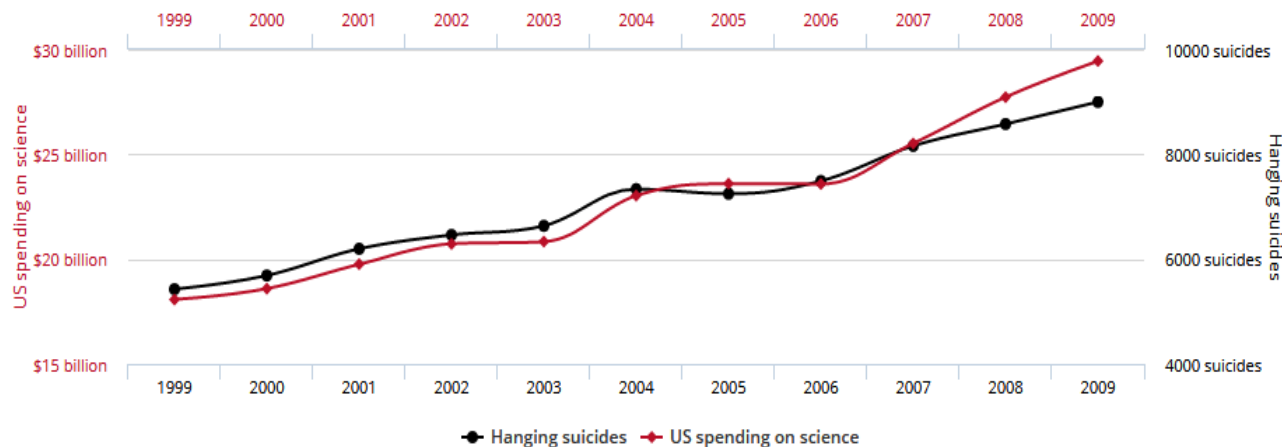


Uma relação estatística existente entre duas variáveis, mas onde não existe nenhuma relação causa-efeito entre elas. Essa relação estatística pode ocorrer por pura coincidência⁺ ou por causa⁺ de uma terceira variável.

• • • •
□ • • • •

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation

Correlation: 99.79% ($r=0.99789126$)

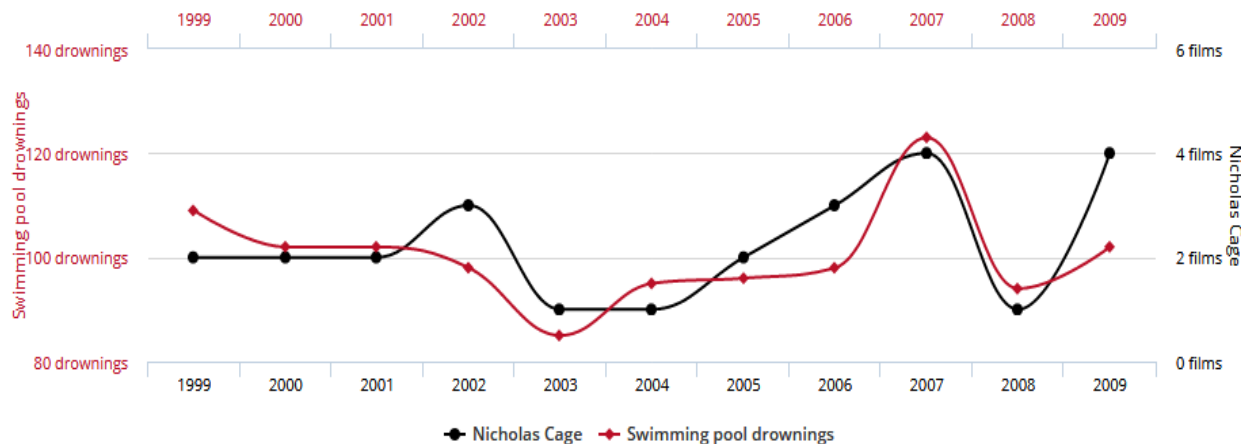


Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

tylervigen.com

Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in

Correlation: 66.6% ($r=0.666004$)

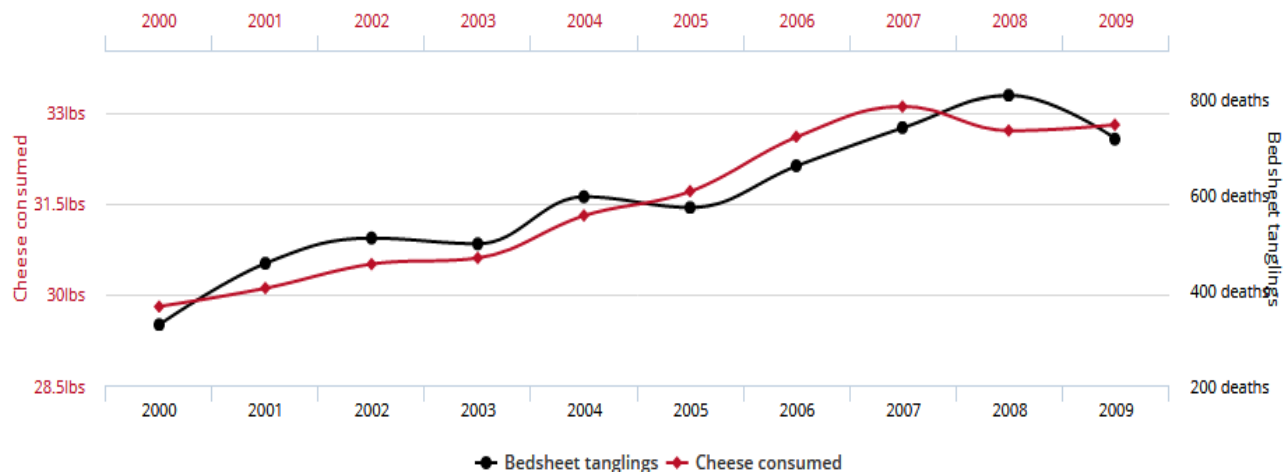


Data sources: Centers for Disease Control & Prevention and Internet Movie Database

tylervigen.com

Per capita cheese consumption correlates with Number of people who died by becoming tangled in their bedsheets

Correlation: 94.71% ($r=0.947091$)

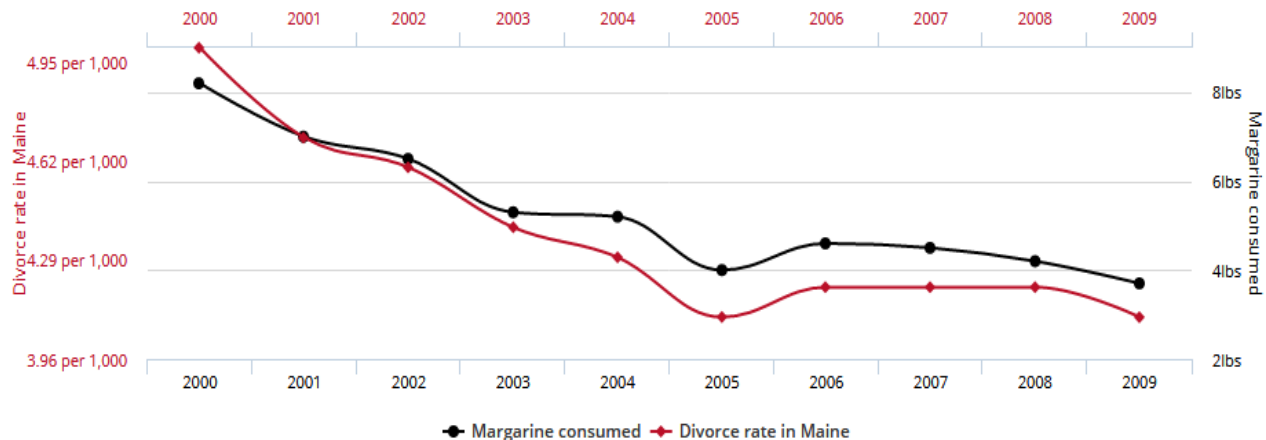


Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

tylervigen.com

Divorce rate in Maine correlates with Per capita consumption of margarine

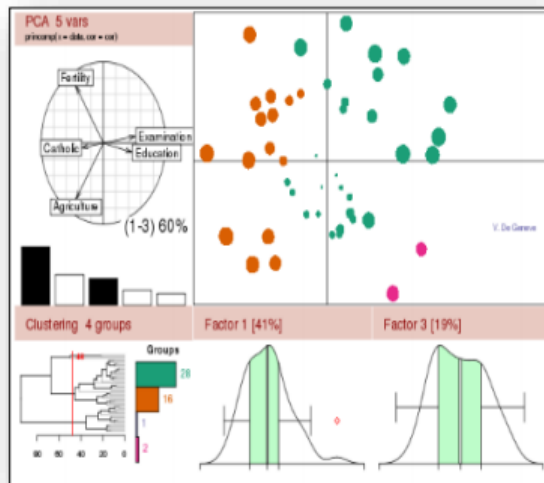
Correlation: 99.26% ($r=0.992558$)



Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

tylervigen.com

ANÁLISE DE CORRELAÇÃO DE PEARSON



Exercícios
Bike Sharing

- ANÁLISE DE ASSOCIAÇÃO

Analisa o comportamento conjunto de duas variáveis qualitativas apresentada em tabela bivariada.

- Teste Qui-Quadrado (Variáveis Qualitativas)
- Correlação de Pearson (Variáveis Quantitativas)

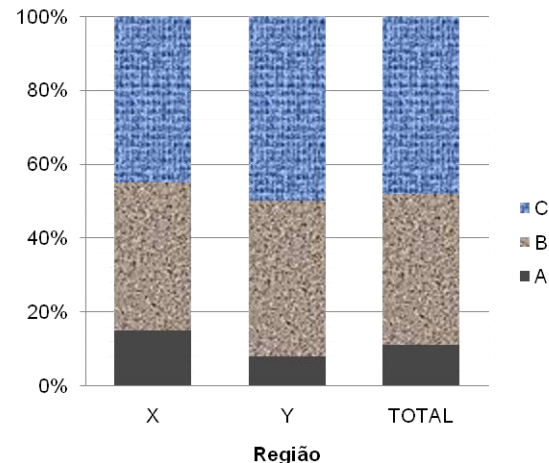
Existe associação entre as vendas de produto e região?

Exemplo:

Tabela 1 – Distribuição de vendas segundo produto e região. 2018

| Produto | Região | | | | Total | |
|---------|--------|-----|------|-----|-------|-----|
| | X | | Y | | | |
| | N | % | N | % | N | % |
| A | 300 | 15 | 200 | 8 | 500 | 11 |
| B | 800 | 40 | 1000 | 42 | 1800 | 41 |
| C | 900 | 45 | 1200 | 50 | 2100 | 48 |
| Total | 2000 | 100 | 2400 | 100 | 4400 | 100 |

Existe associação entre as vendas de produto e região?



Existe associação entre as vendas de produto e região?

| | | REGIAO | | TOTAL |
|---------|---|--------|------|-------|
| | | X | Y | |
| Produto | A | 300 | 200 | 500 |
| | B | 800 | 1000 | 1800 |
| | C | 900 | 1200 | 2100 |
| | | 2000 | 2400 | 4400 |

Chi Square for R by C Table

Chi Square= 49.12
 Degrees of Freedom= 2
 p-value= <0.0000001

Cochran recommends accepting the chi square if

1. No more than 20% of cells have expected < 5.
2. No cell has an expected value < 1.

In this table:

None of 6 cells have expected values < 5.

No cells have expected values < 1.

Using these criteria, this chi square can be accepted.

Expected value = row total*column total/grand total

Rosner, B. Fundamentals of Biostatistics. 5th ed. Duxbury Thompson Learning. 2000; p. 395

Teste de independência qui-quadrado

H_0 : independentes

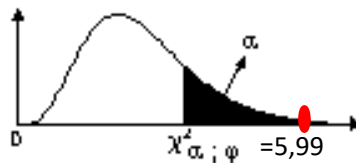
H_1 : dependentes

$\alpha = 5\%$

Conclusão:

Rejeito H_0 , portanto há associação.

Graus de liberdade=



ϕ = graus de liberdade

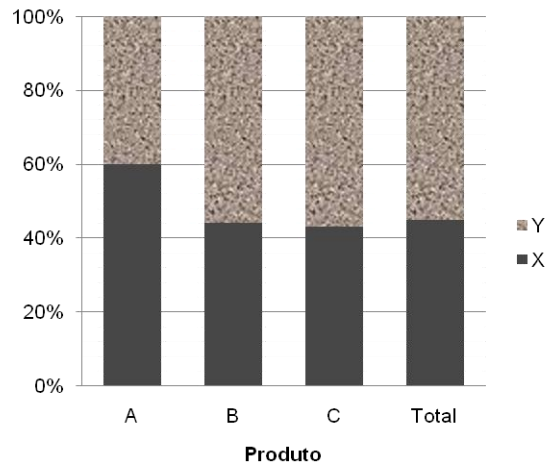
Existe associação entre as vendas de produto e região?

Exemplo:

Tabela 2 – Distribuição de vendas segundo região e produto. 2018

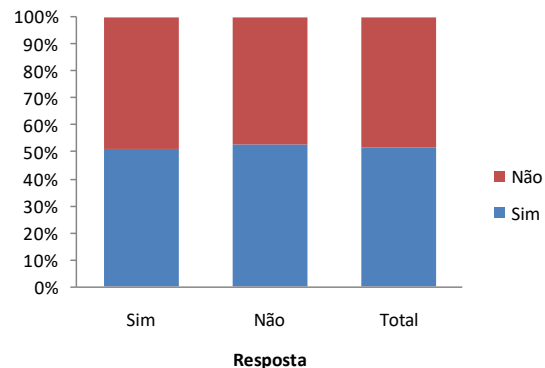
| Produto | Região | | | | Total | |
|---------|--------|----|------|----|-------|-----|
| | X | | Y | | | |
| | N | % | N | % | N | % |
| A | 300 | 60 | 200 | 40 | 500 | 100 |
| B | 800 | 44 | 1000 | 56 | 1800 | 100 |
| C | 900 | 43 | 1200 | 57 | 2100 | 100 |
| Total | 2000 | 45 | 2400 | 55 | 4400 | 100 |

Fonte:zzzz



Será que uma carta de pré-notificação afeta a taxa de resposta dos médicos participantes da pesquisa?

| Resposta | Carta | | | | Total | |
|----------|-------|-------|------|-------|-------|-------|
| | Sim | | Não | | | |
| | N | % | N | % | N | % |
| Sim | 2570 | 51.2 | 2645 | 52.6 | 5215 | 51.9 |
| Não | 2448 | 48.8 | 2384 | 47.4 | 4832 | 48.1 |
| Total | 5018 | 100.0 | 5029 | 100.0 | 10047 | 100.0 |



Será que uma carta de pré-notificação afeta a taxa de resposta dos médicos participantes da pesquisa?

2 x 2 Table Statistics

| | | Single Table Analysis | | |
|----------|-----|-----------------------|------|-------|
| | | Carta | | |
| | | Sim | Nao | |
| Resposta | Sim | 2570 | 2645 | 5215 |
| | Nao | 2448 | 2384 | 4832 |
| | | 5018 | 5029 | 10047 |

Chi Square and Exact Measures of Association

| Test | Value | p-value(1-tail) | p-value(2-tail) |
|----------------------------|-------|-----------------|-----------------|
| Uncorrected chi square | 1.914 | 0.08332 | 0.1666 |
| Yates corrected chi square | 1.859 | 0.08644 | 0.1729 |
| Mantel-Haenszel chi square | 1.914 | 0.08333 | 0.1667 |
| Fisher exact | | ?(P) | ? |
| Mid-P exact | | ?(P) | ? |

All expected values (row total*column total/grand total) are ≥ 5
OK to use chi square.

Conclusão:
Não rejeito H₀, portanto
Não há associação.

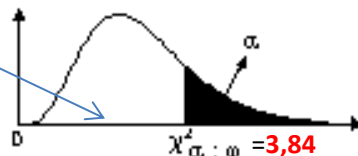
Teste de independência qui-quadrado

H₀: independentes

H₁: dependentes

$\alpha = 5\%$

Graus de liberdade = 1



ϕ = graus de liberdade

TESTE DE HIPÓTESES NÃO PARAMÉTRICO

Teste Qui-Quadrado (Independência)

Distribuições bivariadas de frequências para variáveis qualitativas são apresentadas em tabelas de contingência, que facilitam a análise estatística da possível relação entre duas características observadas em determinada população. A estatística chamada qui-quadrado sintetiza as diferenças entre as frequências observadas de uma tabela bivariada e as correspondentes frequências esperadas.

Definindo as hipóteses H_0 e H_1

H_0 : as variáveis são independentes

H_1 : as variáveis não são independentes, isto é, apresentam algum grau de associação entre si.

TESTE DE HIPÓTESES NÃO PARAMÉTRICO

VANTAGENS

Não é necessário fazer suposições sobre a distribuição da população da qual tenham sido extraídos os dados para a análise.

Aplicáveis a variáveis não contínuas (variáveis categóricas nominais e ordinais).

Simplicidade do ponto de vista de cálculo.

Aplicabilidade a pequenas amostras. Estudo Piloto (ex: amostras de pessoas portadoras de uma certa doença)

➔ As técnicas não-paramétricas são em geral menos eficazes que as paramétricas quando aplicadas a dados onde é possível usar as técnicas paramétricas.

TESTE DE HIPÓTESES NÃO PARAMÉTRICO

A estatística Qui-Quadrado de Pearson é calculada pela expressão

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Sob H_0 , tem distribuição Qui-Quadrado, com $(r-1)(c-1)$ graus de liberdade, sendo r o número de linhas e c o número de colunas

Onde: O_{ij} = número de casos observados classificados na linha i da coluna j ;
 E_{ij} = número de casos esperados, sob H_0 , na linha i da coluna j ;

$\sum_{i=1}^r \sum_{j=1}^k$ Indica somatório sobre todas as células

Distribuição do Qui-Quadrado - χ^2_n

Os valores tabelados correspondem aos pontos x tais que: $P(\chi^2_n \leq x)$

| | P($\chi^2_{n-1} \leq x$) | | | | | | | | | | | | | | |
|----|----------------------------|----------|----------|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----|--|
| n | 0,005 | 0,01 | 0,025 | 0,05 | 0,1 | 0,25 | 0,5 | 0,75 | 0,9 | 0,95 | 0,975 | 0,99 | 0,995 | | |
| 1 | 3,93E-05 | 0,000157 | 0,000982 | 0,003932 | 0,016 | 0,102 | 0,455 | 1,323 | 2,706 | 3,841 | 5,024 | 6,635 | 7,879 | 1 | |
| 2 | 0,010 | 0,020 | 0,054 | 0,102 | 0,211 | 0,475 | 1,386 | 2,773 | 4,605 | 5,991 | 7,378 | 8,558 | 10,597 | 2 | |
| 3 | 0,072 | 0,115 | 0,216 | 0,352 | 0,584 | 1,213 | 2,366 | 4,108 | 6,251 | 7,815 | 9,348 | 11,345 | 12,838 | 3 | |
| 4 | 0,207 | 0,297 | 0,484 | 0,711 | 1,064 | 1,923 | 3,357 | 5,385 | 7,779 | 9,488 | 11,143 | 13,277 | 14,860 | 4 | |
| 5 | 0,412 | 0,554 | 0,831 | 1,145 | 1,610 | 2,675 | 4,351 | 6,626 | 9,236 | 11,070 | 12,832 | 15,086 | 16,750 | 5 | |
| 6 | 0,676 | 0,872 | 1,237 | 1,635 | 2,204 | 3,455 | 5,348 | 7,841 | 10,645 | 12,592 | 14,449 | 16,812 | 18,548 | 6 | |
| 7 | 0,989 | 1,239 | 1,690 | 2,167 | 2,833 | 4,255 | 6,346 | 9,037 | 12,017 | 14,067 | 16,013 | 18,475 | 20,278 | 7 | |
| 8 | 1,344 | 1,647 | 2,180 | 2,733 | 3,490 | 5,071 | 7,344 | 10,219 | 13,362 | 15,507 | 17,535 | 20,090 | 21,955 | 8 | |
| 9 | 1,735 | 2,088 | 2,700 | 3,325 | 4,168 | 5,899 | 8,343 | 11,389 | 14,684 | 16,919 | 19,023 | 21,666 | 23,589 | 9 | |
| 10 | 2,156 | 2,558 | 3,247 | 3,940 | 4,865 | 6,737 | 9,342 | 12,549 | 15,987 | 18,307 | 20,483 | 23,209 | 25,188 | 10 | |
| 11 | 2,603 | 3,053 | 3,816 | 4,575 | 5,578 | 7,584 | 10,341 | 13,701 | 17,275 | 19,675 | 21,920 | 24,725 | 26,757 | 11 | |
| 12 | 3,074 | 3,571 | 4,404 | 5,226 | 6,304 | 8,438 | 11,340 | 14,845 | 18,549 | 21,026 | 23,337 | 26,217 | 28,300 | 12 | |
| 13 | 3,565 | 4,107 | 5,009 | 5,892 | 7,041 | 9,299 | 12,340 | 15,984 | 19,812 | 22,362 | 24,736 | 27,688 | 29,819 | 13 | |
| 14 | 4,075 | 4,660 | 5,629 | 6,571 | 7,790 | 10,165 | 13,339 | 17,117 | 21,064 | 23,685 | 26,119 | 29,141 | 31,319 | 14 | |
| 15 | 4,601 | 5,229 | 6,262 | 7,261 | 8,547 | 11,037 | 14,339 | 18,245 | 22,307 | 24,996 | 27,488 | 30,578 | 32,801 | 15 | |
| 16 | 5,142 | 5,812 | 6,908 | 7,962 | 9,312 | 11,912 | 15,338 | 19,369 | 23,542 | 26,296 | 28,845 | 32,000 | 34,267 | 16 | |
| 17 | 5,697 | 6,408 | 7,564 | 8,672 | 10,085 | 12,792 | 16,338 | 20,489 | 24,769 | 27,587 | 30,191 | 33,409 | 35,718 | 17 | |
| 18 | 6,265 | 7,015 | 8,231 | 9,390 | 10,865 | 13,675 | 17,338 | 21,605 | 25,989 | 28,869 | 31,526 | 34,805 | 37,156 | 18 | |
| 19 | 6,844 | 7,633 | 8,907 | 10,117 | 11,651 | 14,562 | 18,338 | 22,718 | 27,204 | 30,144 | 32,852 | 36,191 | 38,582 | 19 | |
| 20 | 7,434 | 8,260 | 9,591 | 10,851 | 12,443 | 15,452 | 19,337 | 23,828 | 28,412 | 31,410 | 34,170 | 37,566 | 39,997 | 20 | |
| 21 | 8,034 | 8,897 | 10,283 | 11,591 | 13,240 | 16,344 | 20,337 | 24,935 | 29,615 | 32,671 | 35,479 | 38,932 | 41,401 | 21 | |
| 22 | 8,643 | 9,542 | 10,982 | 12,338 | 14,041 | 17,240 | 21,337 | 26,039 | 30,813 | 33,924 | 36,781 | 40,289 | 42,796 | 22 | |
| 23 | 9,260 | 10,196 | 11,689 | 13,091 | 14,848 | 18,137 | 22,337 | 27,141 | 32,007 | 35,172 | 38,076 | 41,638 | 44,181 | 23 | |
| 24 | 9,886 | 10,856 | 12,401 | 13,848 | 15,659 | 19,037 | 23,337 | 28,241 | 33,196 | 36,415 | 39,364 | 42,980 | 45,558 | 24 | |
| 25 | 10,520 | 11,524 | 13,120 | 14,611 | 16,473 | 19,939 | 24,337 | 29,339 | 34,382 | 37,652 | 40,646 | 44,314 | 46,928 | 25 | |
| 26 | 11,160 | 12,198 | 13,844 | 15,379 | 17,292 | 20,843 | 25,336 | 30,435 | 35,563 | 38,885 | 41,923 | 45,642 | 48,290 | 26 | |
| 27 | 11,808 | 12,878 | 14,573 | 16,151 | 18,114 | 21,749 | 26,336 | 31,528 | 36,741 | 40,113 | 43,195 | 46,963 | 49,645 | 27 | |
| 28 | 12,461 | 13,565 | 15,308 | 16,928 | 18,939 | 22,657 | 27,336 | 32,620 | 37,916 | 41,337 | 44,461 | 48,278 | 50,994 | 28 | |
| 29 | 13,121 | 14,256 | 16,047 | 17,708 | 19,768 | 23,567 | 28,336 | 33,711 | 39,087 | 42,557 | 45,722 | 49,588 | 52,335 | 29 | |
| 30 | 13,787 | 14,953 | 16,791 | 18,493 | 20,599 | 24,478 | 29,336 | 34,800 | 40,256 | 43,773 | 46,979 | 50,892 | 53,672 | 30 | |
| 40 | 20,707 | 22,164 | 24,433 | 26,509 | 29,051 | 33,660 | 39,335 | 45,616 | 51,805 | 55,758 | 59,342 | 63,691 | 66,766 | 40 | |
| 50 | 27,991 | 29,707 | 32,357 | 34,764 | 37,689 | 42,942 | 49,335 | 56,334 | 63,167 | 67,505 | 71,420 | 76,154 | 79,490 | 50 | |

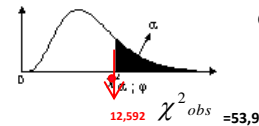
TESTE DE HIPÓTESES NÃO PARAMÉTRICO

Exemplo:

A Associação de Imprensa do Estado de São Paulo fez um levantamento com 1300 leitores, para verificar se a preferência por leitura de um determinado jornal é independente do nível de instrução do indivíduo. Os resultados obtidos foram:

| Observado | A | B | C | D | Total |
|---------------|---------|---------|---------|---------|-------|
| 1. Grau | 10 20% | 8 16% | 5 10% | 27 54% | 50 |
| 2. Grau | 90 20% | 162 36% | 125 28% | 73 16% | 450 |
| Universitário | 200 25% | 250 31% | 220 28% | 130 16% | 800 |
| Total | 300 23% | 420 32% | 350 27% | 230 18% | 1300 |

| Esperado | A | B | C | D | Total |
|---------------|-----|-----|-----|-----|-------|
| 1. Grau | 12 | 16 | 13 | 9 | 50 |
| 2. Grau | 104 | 145 | 121 | 80 | 450 |
| Universitário | 185 | 258 | 215 | 142 | 800 |
| Total | 300 | 420 | 350 | 230 | 1300 |



$$Gl = (\text{linha} - 1) * (\text{colunas} - 1) = (3 - 1) * (4 - 1) = 6$$

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Rejeitamos H_0 ao nível de 5%, isto é os dados trazem evidência de uma forte dependência entre o fatores:
Grau de escolaridade e preferência de jornal

TESTE DE HIPÓTESES NÃO PARAMÉTRICO

Exemplo:

Após uma pesquisa de satisfação estamos interessados em verificar se a preferência pela operadora(OP) estava associada com o fator regional.

| Estado e Operadora | OP1 | OP2 | OP3 | OP4 | Total | |
|--------------------|---------|---------|---------|---------|-------|------|
| SP | 214 33% | 237 37% | 78 12% | 119 18% | 648 | 100% |
| Sul | 51 17% | 102 34% | 126 42% | 22 7% | 301 | 100% |
| RJ | 111 18% | 304 51% | 139 23% | 48 8% | 602 | 100% |
| Total | 376 24% | 643 42% | 343 22% | 189 12% | 1551 | 100% |

→ Se tivesse independência todos os estados teriam a mesma distribuição

→ Número esperado em SP: $648 * 0.24 = 157$;
Sul: $301 * 0.24 = 73$; RJ: $602 * 0.24 = 146$;

H_0 : São independentes

H_a : Não são independentes

$\alpha = 5\%$

Graus de liberdade= $(I-1)(C-1) = 3*2 = 6$



TESTE DE HIPÓTESES NÃO PARAMÉTRICO

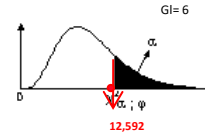
Exemplo:

Valores Observados

| Estado e Operadora | OP1 | OP2 | OP3 | OP4 | Total |
|--------------------|-----|-----|-----|-----|-------|
| SP | 214 | 237 | 78 | 119 | 648 |
| Sul | 51 | 102 | 126 | 22 | 301 |
| RJ | 111 | 304 | 139 | 48 | 602 |
| Total | 376 | 643 | 343 | 189 | 1.551 |

Valores Esperado

| Estado e Operadora | OP1 | OP2 | OP3 | OP4 | Total |
|--------------------|-----|-----|-----|-----|-------|
| SP | 157 | 269 | 143 | 79 | 648 |
| Sul | 73 | 125 | 67 | 37 | 301 |
| RJ | 146 | 250 | 133 | 73 | 602 |
| Total | 376 | 643 | 343 | 189 | 1.551 |

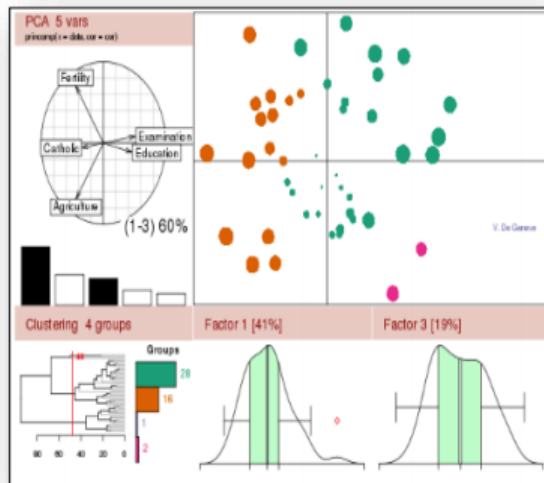


$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2_{obs} = (214-157)^2/157 + \dots + (48-73)^2/73 = 173.24$$

Rejeitamos H_0 ao nível de 5%, isto é os dados trazem evidência de uma forte dependência entre o fatores: Operadora de Celular e Região

ANÁLISE DE ASSOCIAÇÃO

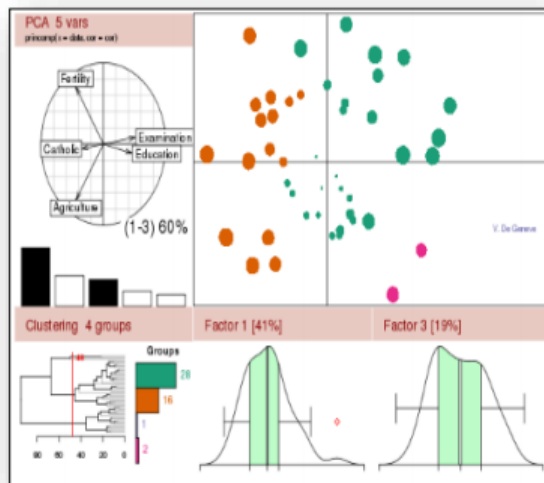


Exercícios
Bike Sharing

• OUTROS TESTES DE HIPÓTESES NÃO PARAMÉTRICO

| Teste | |
|----------------------|---|
| Spearman | Usado para verificar a associação entre duas variáveis qualitativas ordinais |
| Kolmogorov-Smirnov | Comparar duas populações |
| Mann-Witney | Comparar duas amostras independentes |
| Wilcoxon Signed-Rank | Usado para comparar duas amostras pareadas ou dependentes |
| Kruskal-Wallis | Comparar três ou mais grupos |
| Sinal | Usado para verificar a associação entre duas variáveis qualitativas ordinais de amostras pareadas |
| | |

DISTRIBUIÇÃO DE REGRESSÃO LINEAR MULTIPLA

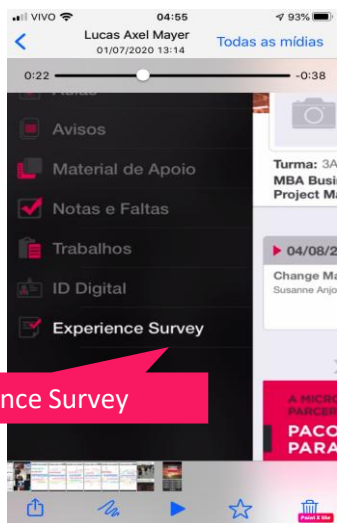


Exercícios

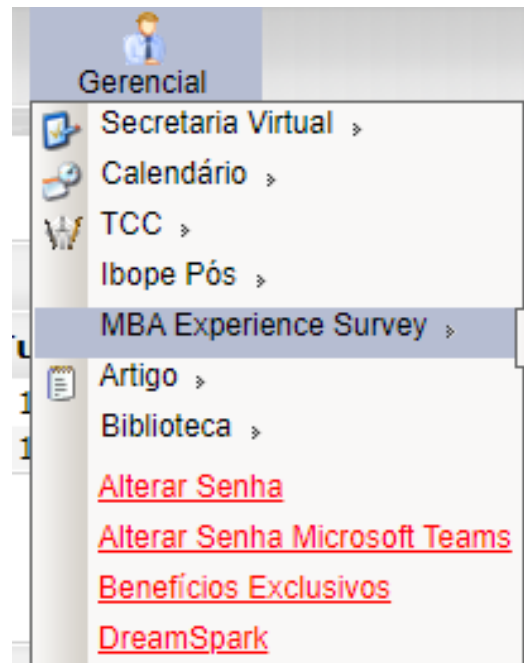
O que você achou da aula de hoje?

Pelo aplicativo da FIAP

(Entrar no FIAPP, e no menu clicar em Experience Survey)



Experience Survey



A grande finalidade do
conhecimento não é conhecer,
mas agir.

T. Huxley

OBRIGADO



/ Regina T. I. Bernal

FIAP

Copyright © 2024 | Professora Dra. Regina Tomie Ivata Bernal
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente
proibido sem consentimento formal, por escrito, do professor/autor.

FIAP