

# Teste09

Guilherme Pazian

13 de junho de 2017

```
library(dplyr)

## Warning: Installed Rcpp (0.12.10) different from Rcpp used to build dplyr (0.12.11).
## Please reinstall dplyr to avoid random crashes or undefined behavior.

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

dados <- read.csv("Names.csv")
```

## Questão A

Defina a taxa de retorno de chamada como a proporção de CV's que geraram uma chamada telefonica do futuro empregador. Quais foram as taxas de retorno de chamada para os brancos e afroamericanos? Calcule um intervalo de confiança de 95% para a diferença das taxas de retorno. Essa diferença é estatisticamente significativa? A diferença é grande na prática?

```
#taxa de retorno para brancos
brancos=filter(dados, black==0)
retornobrancos=sum(brancos$call_back)
num_brancos=nrow(brancos)
taxa_b=retornobrancos/num_brancos

#taxa de retorno para afroamericanos
negros=filter(dados, black==1)
retornonegros=sum(negros$call_back)
num_negros=nrow(negros)
taxa_n=retornonegros/num_negros
```

As taxas de retorno foram:

- Brancos: 0.0965092
- Afroamericanos: 0.0644764

IC:

Seja  $Y_b$  : número de retornos para brancos  $Y_b \sim \text{Binomial}(n_b, p_b)$  Pelo teorema central do limite:  $Y_b \sim N(p_b, n_b p_b (1 - p_b))$  onde  $n_b = 2435$

Pela propriedade de distribuição assintótica do estimador de máxima verossimilhança, temos:  $\hat{p}_b \sim N(p_b, I(p))$  onde  $I(p_b)^{-1} = \frac{n_b}{p_b a(1-p_b)}$  é a informação de fisher.

Seja  $Y_a$  : número de retornos para afroamericanos  $Y_a \sim \text{Binomial}(n_a, p_a)$

$Y_a \sim N(p_a, n_a p_a (1 - p_a))$  onde  $n_a = 2435$

Pela propriedade de distribuição assintótica do estimador de máxima verossimilhança, temos:  $\hat{p}_a \sim N(p_a, I(a))$  onde  $I(p_a)^{-1} = \frac{n_a}{p_a(1-p_a)}$  é a informação de Fisher.

Por hipótese as distribuições são independentes.

Então

$$(p_b - p_a) \sim N(p_b - p_a, \text{var}(p_b) + \text{var}(p_a)) \Rightarrow (p_b - p_a) \sim N(p_b - p_a, \frac{n_b}{p_b(1-p_b)} + \frac{n_a}{p_a(1-p_a)})$$

Um  $Ic((p_b - p_a), 95\%) = (\hat{p}_b - \hat{p}_a - 1.96 \frac{n_b}{p_b(1-p_b)}; \hat{p}_b - \hat{p}_a + 1.96 \frac{n_b}{p_b(1-p_b)})$

```
Var_dif <- 1/(num_brancois/(taxa_b*(1-taxa_b)) + num_negros/(taxa_n*(1-taxa_n)))
LI_IC_A <- (taxa_b - taxa_n) - qnorm(.975) * sqrt(Var_dif)
LS_IC_A <- (taxa_b - taxa_n) - qnorm(.975) * sqrt(Var_dif)
```

Um  $Ic((p_b - p_a), 95\%) = [0.0245; 0.0245]$

## Questão 2

Considere os brancos; é diferente a taxa de retorno para homens e mulheres? E para os afroamericanos?

```
#brancos
homens_b=filter(brancos, female==0)
mulheres_b=filter(brancos, female==1)
retornobrancos_H=sum(homens_b$call_back==1)
retornobrancos_M=sum(mulheres_b$call_back==1)
num_brancois_H=nrow(homens_b)
num_brancois_M=nrow(mulheres_b)

taxa_b_H=retornobrancos_H/num_brancois_H
taxa_b_M=retornobrancos_M/num_brancois_M

#afroamericanos
homens_n=filter(negros, female==0)
mulheres_n=filter(negros, female==1)

retornonegros_H=sum(homens_n$call_back==1)
retornonegros_M=sum(mulheres_n$call_back==1)
num_negros_H=nrow(homens_n)
num_negros_M=nrow(mulheres_n)

taxa_n_H=retornonegros_H/num_negros_H
taxa_n_M=retornonegros_M/num_negros_M
```

As taxas de retorno para brancos:

- Homens: 0.0886957
- Mulheres: 0.0989247

As taxas de retorno para negros foram:

- Homens: 0.0582878
- Mulheres: 0.0662778

Podemos testar as hipóteses de que  $H_0 : p_{bh} - p_{bm} = 0$  Vs  $H_1 p_{bh} - p_{bm} \neq 0$

$H_0 : p_{ah} - p_{am} = 0$  Vs  $H_1 p_{ah} - p_{am} \neq 0$  com um teste de hipóteses normal. (como anteriormente para o teste de hipóteses)

```
var_dif_HM_branco <- 1/(num_branco_M/(taxa_b_M*(1-taxa_b_M)) + num_branco_H/(taxa_b_H*(1-taxa_b_H)))
teste_HM_branco <- (retornobranco_H - retornobranco_M)/sqrt(var_dif_HM_branco)
p_valor_HM_branco <- pnorm(teste_HM_branco)
```

0 é o p-valor para o teste HM\_branco

```
var_dif_HM_negro <- 1/(num_negro_M/(taxa_n_M*(1-taxa_n_M)) + num_negro_H/(taxa_n_H*(1-taxa_n_H)))
teste_HM_negro <- (retornonegro_H - retornonegro_M)/sqrt(var_dif_HM_negro)
p_valor_HM_negro <- pnorm(teste_HM_negro)
```

0 é o p-valor para o teste HM\_negro

### Questão 3

Qual a diferença na taxa de retorno nos CV's de qualificação elevada comparados com os CV's de qualificação baixa? Qual a diferença entre os CV's de alta qualificação e os de baixa qualificação para os brancos? e para os afroamericanos? Essas diferenças são estatisticamente significativas e são importantes na prática?

```
retorno=sum(dados$call_back)
num_cvs=nrow(dados)
```

### Questão 4

Os autores do estudo afirmam que a etnia foi asignada aleatoriamente a cada CV. Há evidência de asignação não aleatória?

```
ajuste=lm(data = dados, black~female+high+chicago+ofjobs+yearsexp+honors+volunteer+military+empholes+workinschool+
summary(ajuste)
```

```
##
## Call:
## lm(formula = black ~ female + high + chicago + ofjobs + yearsexp +
##      honors + volunteer + military + empholes + workinschool +
##      email + computerskills + specialskills + college + expminreq +
##      eoe + manager + supervisor + secretary + offsupport + salesrep +
##      retailsales + req + expreq + comreq + +educreq + compreq +
##      orgreq + manuf + transcom + bankreal + trade + busservice +
##      othservice + missind, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55727 -0.50254  0.01167  0.49634  0.56976
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4561582  0.0459800   9.921  <2e-16 ***
## female         0.0202326  0.0203191   0.996    0.319
## high          -0.0287215  0.0408071  -0.704    0.482
## chicago        -0.0010772  0.0198367  -0.054    0.957
## ofjobs         -0.0037604  0.0073022  -0.515    0.607
## yearsexp        0.0009921  0.0017237   0.576    0.565
## honors         -0.0162917  0.0344611  -0.473    0.636
```

```

## volunteer      0.0158859  0.0244266  0.650  0.515
## military       0.0287609  0.0276774  1.039  0.299
## empholes      -0.0015041  0.0185957 -0.081  0.936
## workinschool  -0.0073201  0.0187240 -0.391  0.696
## email         0.0069290  0.0351206  0.197  0.844
## computerskills 0.0508133  0.0220130  2.308  0.021 *
## specialskills  0.0005960  0.0168662  0.035  0.972
## college       0.0132414  0.0178118  0.743  0.457
## expminreq0     0.0044599  0.3606361  0.012  0.990
## expminreq0,5   0.0125450  0.3134638  0.040  0.968
## expminreq1     0.0001252  0.2606751  0.000  1.000
## expminreq10    -0.0062322  0.2827937 -0.022  0.982
## expminreq2     0.0023340  0.2580564  0.009  0.993
## expminreq3     0.0001608  0.2580996  0.001  1.000
## expminreq4     0.0026934  0.3123338  0.009  0.993
## expminreq5     0.0045138  0.2595982  0.017  0.986
## expminreq6     0.0008377  0.3132470  0.003  0.998
## expminreq7    -0.0007803  0.2946457 -0.003  0.998
## expminreq8    -0.0017766  0.3023176 -0.006  0.995
## expminreqsome  0.0034677  0.2561720  0.014  0.989
## eoe           -0.0009425  0.0172196 -0.055  0.956
## manager       -0.0032643  0.0277232 -0.118  0.906
## supervisor    -0.0108401  0.0372959 -0.291  0.771
## secretary     -0.0118168  0.0291814 -0.405  0.686
## offsupport    -0.0113119  0.0326889 -0.346  0.729
## salesrep      -0.0017763  0.0261932 -0.068  0.946
## retailsales   NA          NA          NA          NA
## req           -0.0037710  0.0250377 -0.151  0.880
## expreq        -0.0021317  0.2565086 -0.008  0.993
## comreq        -0.0011993  0.0243622 -0.049  0.961
## educreq       -0.0023296  0.0253834 -0.092  0.927
## compreq       -0.0038826  0.0212582 -0.183  0.855
## orgreq        -0.0023689  0.0306003 -0.077  0.938
## manuf         0.0032519  0.0314666  0.103  0.918
## transcom      0.0078364  0.0463584  0.169  0.866
## bankreal      0.0026233  0.0313559  0.084  0.933
## trade         0.0073828  0.0252456  0.292  0.770
## busservice    0.0019322  0.0230003  0.084  0.933
## othservice    0.0015816  0.0269363  0.059  0.953
## missind       NA          NA          NA          NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5018 on 4825 degrees of freedom
## Multiple R-squared:  0.001906, Adjusted R-squared: -0.007196
## F-statistic: 0.2094 on 44 and 4825 DF, p-value: 1

```