análisis inferencial 2022

Belmonte Valeria

2024-08-06

getwd()

## [1] "C:/Users/USUARIO/Desktop/copia de seguridad/trabajo/trabajo/consultorias/consultorias 2022/proyecto IA/analisis ingresantes 2022/toda\_la\_facu\_2022"

## carga de paquetes

## Warning: package 'paletteer' was built under R version 4.3.3

## Warning: package 'treemap' was built under R version 4.3.3

## Borro los datos

##cargo los datos con caracteristicas de los estudiantes

La base **ingresantes2023** todos los ingresantes 2022, con los resultados de regualridad en ese año por carrera

## tibble [239 × 52] (S3: tbl\_df/tbl/data.frame)  
## $ Documento : num [1:239] 28329545 29074481 29979048 31481836 31482298 ...  
## $ año\_ing\_F : num [1:239] 2022 2022 2022 2022 2022 ...  
## $ Nombre : chr [1:239] "Tropea, Norberto Ariel" "Carrillo, Fernando Matías" "Martin, Cintia Anabel" "Ochoa, Mauricio Matías" ...  
## $ Carrera : chr [1:239] "TPVI" "TPVI" "TPVI" "TPVI" ...  
## $ edad : num [1:239] 42 40 38 37 37 35 28 27 25 25 ...  
## $ Nacionalidad : chr [1:239] "Argentino" "Argentino" "Argentino" "Argentino" ...  
## $ Genero : chr [1:239] "Masculino" "Masculino" "Masculino" "Masculino" ...  
## $ Estado\_Civil : chr [1:239] "Soltero" "Casado" "Soltero" "Soltero" ...  
## $ Cantidad\_Hijos : chr [1:239] NA "Uno" "No tiene" NA ...  
## $ Situacion\_Ocupacional\_Alumno : chr [1:239] "T" "T" "T" "NTyNBT" ...  
## $ Tipo\_Residencia\_Periodo\_Lectivo : chr [1:239] NA NA "Casa" NA ...  
## $ Con\_quien\_vive\_Periodo\_de\_Clase : chr [1:239] "Otros" NA "Otros" NA ...  
## $ Localidad\_Periodo\_de\_Clase : chr [1:239] "SANTA ROSA" "TOAY" NA "SANTA ROSA" ...  
## $ Cost\_est\_\_aporte\_familiar : chr [1:239] "N" "N" "N" "N" ...  
## $ Cost\_est\_p\_sociales : chr [1:239] "S" "S" "S" "N" ...  
## $ Cos\_est\_Trabajo : chr [1:239] "N" "N" "N" "N" ...  
## $ Cost\_est\_Becas : chr [1:239] "N" "N" "N" "N" ...  
## $ Cost\_est\_otros : chr [1:239] NA NA NA NA ...  
## $ Colegio\_Secundario : chr [1:239] NA "ESCUELA NORMAL SUPERIOR TTE. GRAL. JULIO A. ROCA (media)" "COLEGIO CIUDAD DE SANTA ROSA" NA ...  
## $ Localidad\_Col.Sec. : chr [1:239] NA "SANTA ROSA" "SANTA ROSA" NA ...  
## $ Titulo\_Secundario : chr [1:239] NA "BACHILLER PEDAGÓGICO" "Comunicación Artes y Diseño" NA ...  
## $ Año\_Egreso : num [1:239] 1998 1999 2002 2002 2011 ...  
## $ Orientacion\_Vocacional : chr [1:239] "Ninguna" NA NA NA ...  
## $ Tiene\_beca? : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_de\_la\_Universidad? : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_Nacional : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_Provincial : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_Municipal : chr [1:239] "N" "N" "N" "N" ...  
## $ Otra\_fuente\_de\_Beca : chr [1:239] "N" "N" "N" "N" ...  
## $ Tipo\_de\_Beca,\_Economica : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_de\_transporte : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_de\_comedor : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_de\_fotocopias : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_de\_efectivo : chr [1:239] "N" "N" "N" "N" ...  
## $ Beca\_Habitacional : chr [1:239] "N" "N" "N" "N" ...  
## $ Padre\_Vive : chr [1:239] "S" "S" "S" "S" ...  
## $ Estudios\_Padre : chr [1:239] "Estudios secundarios incompletos" "Estudios secundarios completos" "Estudios superiores completos" "Estudios primarios completos" ...  
## $ Madre\_vive : chr [1:239] "S" "S" "S" "S" ...  
## $ Estudios\_Madre : chr [1:239] "Estudios secundarios completos" "Estudios de Post grado" "Estudios primarios completos" "Estudios primarios completos" ...  
## $ Otros\_Est\_Sup : chr [1:239] NA "Universidad Nacional de Córdoba" "CREAR" "Universidad Nacional de La Plata" ...  
## $ Otros\_est\_sup\_Carrera : chr [1:239] NA "Comunicación Social" "Profesoor de Arte en Artes Visuales" "Prof. Música or. Música Popular" ...  
## $ Otros\_Estudios\_Superiores,\_Estado: chr [1:239] NA "Abandonó" "Abandonó" "Abandonó" ...  
## $ Practica\_Deporte? : chr [1:239] "S" "N" "N" "S" ...  
## $ deporte\_Uni : chr [1:239] "N" "N" "N" "N" ...  
## $ Cobertura\_medica : chr [1:239] "Por su propio trabajo" NA "Por su propio trabajo" NA ...  
## $ Localidad\_Procedencia : chr [1:239] "LINCOLN" "TOAY" NA "SANTA ROSA" ...  
## $ PROVINCIA : chr [1:239] "Buenos Aires" "La Pampa" "SD" "La Pampa" ...  
## $ reincriptos\_23 : num [1:239] 0 0 0 0 1 0 1 0 1 1 ...  
## $ Num\_Regularizadas : num [1:239] 0 0 0 1 4 0 4 0 6 7 ...  
## $ Porc\_Regularizadas : num [1:239] 0 0 0 11,1 44,4 0 44,4 0 66,7 77,8 ...  
## $ asitio\_alguna\_materia : num [1:239] 0 0 0 1 1 0 1 0 1 1 ...  
## $ Otros\_Est\_Sup\_indicator : num [1:239] 0 1 1 1 0 1 1 0 0 0 ...

## Resumen de los datos

## Warning in prettyNum(r, big.mark = big.mark, big.interval = big.interval, :  
## 'big.mark' and 'decimal.mark' are both ',', which could be confusing

| reincriptos\_23 | n |
| --- | --- |
| 0 | 125 |
| 1 | 114 |

En función de En la regla de “eventos por variable predictora” (EPV)  
tenemos = n(1)/ v. predictoras= 10; entonces; 114/5= v predictoras daría 11 variables predicotoras además n(0)=125/v predictoras = 10 -> 12 variables predictoras

por lo tanto no está muy desbalanceados

## Voy a hacer ua selección de modelo

m2=glm(formula = reincriptos\_23 ~   
 Carrera +   
 Estado\_Civil +  
 Genero +   
 Situacion\_Ocupacional\_Alumno +  
 Porc\_Regularizadas +  
 edad +   
 as.factor(asitio\_alguna\_materia) +  
 Cobertura\_medica + Estudios\_Madre +   
 Estudios\_Padre +  
 Año\_Egreso,  
 family = binomial, data = ingresantes\_2022)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

step(m2,direction="backward")

## Start: AIC=99,32  
## reincriptos\_23 ~ Carrera + Estado\_Civil + Genero + Situacion\_Ocupacional\_Alumno +   
## Porc\_Regularizadas + edad + as.factor(asitio\_alguna\_materia) +   
## Cobertura\_medica + Estudios\_Madre + Estudios\_Padre + Año\_Egreso

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Df Deviance AIC  
## <none> 25,316 99,316  
## - edad 1 31,154 103,154  
## - Estado\_Civil 3 35,292 103,292  
## - Estudios\_Padre 9 47,372 103,372  
## - Cobertura\_medica 4 38,388 104,388  
## - Situacion\_Ocupacional\_Alumno 4 39,225 105,225  
## - as.factor(asitio\_alguna\_materia) 1 33,227 105,227  
## - Genero 1 34,373 106,373  
## - Año\_Egreso 1 37,575 109,575  
## - Carrera 2 41,642 111,642  
## - Estudios\_Madre 9 57,546 113,546  
## - Porc\_Regularizadas 1 120,125 192,125

##   
## Call: glm(formula = reincriptos\_23 ~ Carrera + Estado\_Civil + Genero +   
## Situacion\_Ocupacional\_Alumno + Porc\_Regularizadas + edad +   
## as.factor(asitio\_alguna\_materia) + Cobertura\_medica + Estudios\_Madre +   
## Estudios\_Padre + Año\_Egreso, family = binomial, data = ingresantes\_2022)  
##   
## Coefficients:  
## (Intercept)   
## -7,108e+03   
## CarreraLANA   
## 5,097e+00   
## CarreraTPVI   
## -5,733e+01   
## Estado\_CivilDivorciado   
## -9,905e+00   
## Estado\_CivilSeparado   
## -9,956e+01   
## Estado\_CivilSoltero   
## -6,689e+01   
## GeneroMasculino   
## 1,267e+01   
## Situacion\_Ocupacional\_AlumnoNT yBT   
## -2,569e+01   
## Situacion\_Ocupacional\_AlumnoNTyBT   
## 3,036e+01   
## Situacion\_Ocupacional\_AlumnoNTyNBT   
## 1,071e+01   
## Situacion\_Ocupacional\_AlumnoT   
## 1,023e+01   
## Porc\_Regularizadas   
## 1,923e+00   
## edad   
## -7,496e-01   
## as.factor(asitio\_alguna\_materia)1   
## -2,625e+01   
## Cobertura\_medicaComo afiliado voluntario   
## 2,183e+01   
## Cobertura\_medicaOtorgada por la universidad   
## 1,408e+01   
## Cobertura\_medicaPor ser familiar a cargo   
## 5,837e+00   
## Cobertura\_medicaPor su propio trabajo   
## -2,578e+01   
## Estudios\_MadreEstudios primarios completos   
## 5,738e+01   
## Estudios\_MadreEstudios primarios incompletos   
## 6,452e+01   
## Estudios\_MadreEstudios secundarios completos   
## 3,972e+01   
## Estudios\_MadreEstudios secundarios incompletos   
## 3,773e+01   
## Estudios\_MadreEstudios superiores completos   
## 1,399e+01   
## Estudios\_MadreEstudios superiores incompletos   
## 1,377e+02   
## Estudios\_MadreEstudios universitarios completos   
## 3,865e+01   
## Estudios\_MadreEstudios universitarios incompletos   
## 5,669e+01   
## Estudios\_MadreNo hizo estudios   
## 1,002e+02   
## Estudios\_PadreEstudios primarios completos   
## -4,070e+01   
## Estudios\_PadreEstudios primarios incompletos   
## -4,836e+01   
## Estudios\_PadreEstudios secundarios completos   
## -4,710e+01   
## Estudios\_PadreEstudios secundarios incompletos   
## -2,774e+01   
## Estudios\_PadreEstudios superiores completos   
## -4,673e+01   
## Estudios\_PadreEstudios superiores incompletos   
## -1,995e+01   
## Estudios\_PadreEstudios universitarios completos   
## -6,290e+01   
## Estudios\_PadreEstudios universitarios incompletos   
## 8,661e-02   
## Estudios\_PadreNo hizo estudios   
## -8,428e+01   
## Año\_Egreso   
## 3,549e+00   
##   
## Degrees of Freedom: 145 Total (i.e. Null); 109 Residual  
## (93 observations deleted due to missingness)  
## Null Deviance: 201,7   
## Residual Deviance: 25,32 AIC: 99,32

summary(m2)

##   
## Call:  
## glm(formula = reincriptos\_23 ~ Carrera + Estado\_Civil + Genero +   
## Situacion\_Ocupacional\_Alumno + Porc\_Regularizadas + edad +   
## as.factor(asitio\_alguna\_materia) + Cobertura\_medica + Estudios\_Madre +   
## Estudios\_Padre + Año\_Egreso, family = binomial, data = ingresantes\_2022)  
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -7,108e+03 7,265e+03 -0,978  
## CarreraLANA 5,097e+00 4,630e+00 1,101  
## CarreraTPVI -5,733e+01 2,905e+01 -1,973  
## Estado\_CivilDivorciado -9,905e+00 2,996e+04 0,000  
## Estado\_CivilSeparado -9,956e+01 2,996e+04 -0,003  
## Estado\_CivilSoltero -6,689e+01 6,578e+03 -0,010  
## GeneroMasculino 1,267e+01 7,098e+00 1,786  
## Situacion\_Ocupacional\_AlumnoNT yBT -2,569e+01 5,862e+03 -0,004  
## Situacion\_Ocupacional\_AlumnoNTyBT 3,036e+01 1,513e+01 2,006  
## Situacion\_Ocupacional\_AlumnoNTyNBT 1,071e+01 6,635e+00 1,614  
## Situacion\_Ocupacional\_AlumnoT 1,023e+01 5,947e+00 1,721  
## Porc\_Regularizadas 1,923e+00 9,201e-01 2,090  
## edad -7,497e-01 3,606e-01 -2,079  
## as.factor(asitio\_alguna\_materia)1 -2,625e+01 1,683e+01 -1,559  
## Cobertura\_medicaComo afiliado voluntario 2,183e+01 1,115e+01 1,958  
## Cobertura\_medicaOtorgada por la universidad 1,408e+01 4,745e+07 0,000  
## Cobertura\_medicaPor ser familiar a cargo 5,837e+00 3,829e+00 1,524  
## Cobertura\_medicaPor su propio trabajo -2,578e+01 4,404e+03 -0,006  
## Estudios\_MadreEstudios primarios completos 5,738e+01 2,826e+01 2,031  
## Estudios\_MadreEstudios primarios incompletos 6,452e+01 1,603e+04 0,004  
## Estudios\_MadreEstudios secundarios completos 3,972e+01 1,984e+01 2,002  
## Estudios\_MadreEstudios secundarios incompletos 3,773e+01 1,963e+01 1,922  
## Estudios\_MadreEstudios superiores completos 1,399e+01 7,913e+00 1,767  
## Estudios\_MadreEstudios superiores incompletos 1,377e+02 2,981e+04 0,005  
## Estudios\_MadreEstudios universitarios completos 3,865e+01 1,939e+01 1,993  
## Estudios\_MadreEstudios universitarios incompletos 5,669e+01 2,818e+01 2,012  
## Estudios\_MadreNo hizo estudios 1,002e+02 4,227e+04 0,002  
## Estudios\_PadreEstudios primarios completos -4,070e+01 1,907e+01 -2,134  
## Estudios\_PadreEstudios primarios incompletos -4,836e+01 2,195e+01 -2,203  
## Estudios\_PadreEstudios secundarios completos -4,710e+01 2,197e+01 -2,143  
## Estudios\_PadreEstudios secundarios incompletos -2,774e+01 1,347e+01 -2,059  
## Estudios\_PadreEstudios superiores completos -4,673e+01 2,163e+01 -2,160  
## Estudios\_PadreEstudios superiores incompletos -1,995e+01 2,956e+04 -0,001  
## Estudios\_PadreEstudios universitarios completos -6,290e+01 3,183e+04 -0,002  
## Estudios\_PadreEstudios universitarios incompletos 8,661e-02 7,217e+00 0,012  
## Estudios\_PadreNo hizo estudios -8,428e+01 3,053e+04 -0,003  
## Año\_Egreso 3,549e+00 1,536e+00 2,311  
## Pr(>|z|)   
## (Intercept) 0,3279   
## CarreraLANA 0,2709   
## CarreraTPVI 0,0484 \*  
## Estado\_CivilDivorciado 0,9997   
## Estado\_CivilSeparado 0,9973   
## Estado\_CivilSoltero 0,9919   
## GeneroMasculino 0,0742 .  
## Situacion\_Ocupacional\_AlumnoNT yBT 0,9965   
## Situacion\_Ocupacional\_AlumnoNTyBT 0,0448 \*  
## Situacion\_Ocupacional\_AlumnoNTyNBT 0,1066   
## Situacion\_Ocupacional\_AlumnoT 0,0853 .  
## Porc\_Regularizadas 0,0366 \*  
## edad 0,0376 \*  
## as.factor(asitio\_alguna\_materia)1 0,1189   
## Cobertura\_medicaComo afiliado voluntario 0,0503 .  
## Cobertura\_medicaOtorgada por la universidad 1,0000   
## Cobertura\_medicaPor ser familiar a cargo 0,1274   
## Cobertura\_medicaPor su propio trabajo 0,9953   
## Estudios\_MadreEstudios primarios completos 0,0423 \*  
## Estudios\_MadreEstudios primarios incompletos 0,9968   
## Estudios\_MadreEstudios secundarios completos 0,0452 \*  
## Estudios\_MadreEstudios secundarios incompletos 0,0546 .  
## Estudios\_MadreEstudios superiores completos 0,0772 .  
## Estudios\_MadreEstudios superiores incompletos 0,9963   
## Estudios\_MadreEstudios universitarios completos 0,0463 \*  
## Estudios\_MadreEstudios universitarios incompletos 0,0442 \*  
## Estudios\_MadreNo hizo estudios 0,9981   
## Estudios\_PadreEstudios primarios completos 0,0328 \*  
## Estudios\_PadreEstudios primarios incompletos 0,0276 \*  
## Estudios\_PadreEstudios secundarios completos 0,0321 \*  
## Estudios\_PadreEstudios secundarios incompletos 0,0395 \*  
## Estudios\_PadreEstudios superiores completos 0,0307 \*  
## Estudios\_PadreEstudios superiores incompletos 0,9995   
## Estudios\_PadreEstudios universitarios completos 0,9984   
## Estudios\_PadreEstudios universitarios incompletos 0,9904   
## Estudios\_PadreNo hizo estudios 0,9978   
## Año\_Egreso 0,0208 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 201,714 on 145 degrees of freedom  
## Residual deviance: 25,316 on 109 degrees of freedom  
## (93 observations deleted due to missingness)  
## AIC: 99,316  
##   
## Number of Fisher Scoring iterations: 20

#elimine asisitio a alguna materia  
m3=glm(formula = reincriptos\_23 ~  
 Carrera +   
 Estado\_Civil +  
 Genero +   
 Situacion\_Ocupacional\_Alumno +  
 Porc\_Regularizadas +  
 edad +  
 Cobertura\_medica +  
 Estudios\_Madre +   
 Estudios\_Padre +  
 Año\_Egreso,  
 family = binomial, data = ingresantes\_2022)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(m3)

##   
## Call:  
## glm(formula = reincriptos\_23 ~ Carrera + Estado\_Civil + Genero +   
## Situacion\_Ocupacional\_Alumno + Porc\_Regularizadas + edad +   
## Cobertura\_medica + Estudios\_Madre + Estudios\_Padre + Año\_Egreso,   
## family = binomial, data = ingresantes\_2022)  
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -4,890e+03 2,783e+03 -1,757  
## CarreraLANA 1,287e+00 3,340e+00 0,385  
## CarreraTPVI -2,451e+01 1,232e+01 -1,989  
## Estado\_CivilDivorciado 6,365e+00 1,086e+04 0,001  
## Estado\_CivilSeparado -6,271e+01 1,086e+04 -0,006  
## Estado\_CivilSoltero -3,911e+01 1,491e+03 -0,026  
## GeneroMasculino 6,663e+00 3,827e+00 1,741  
## Situacion\_Ocupacional\_AlumnoNT yBT -1,715e+01 8,831e+02 -0,019  
## Situacion\_Ocupacional\_AlumnoNTyBT 1,219e+01 7,167e+00 1,701  
## Situacion\_Ocupacional\_AlumnoNTyNBT 4,566e+00 4,638e+00 0,984  
## Situacion\_Ocupacional\_AlumnoT 5,498e+00 4,298e+00 1,279  
## Porc\_Regularizadas 9,188e-01 4,086e-01 2,249  
## edad -4,261e-01 3,287e-01 -1,296  
## Cobertura\_medicaComo afiliado voluntario 1,728e+01 1,674e+01 1,033  
## Cobertura\_medicaOtorgada por la universidad 5,686e+00 4,745e+07 0,000  
## Cobertura\_medicaPor ser familiar a cargo 2,533e+00 2,037e+00 1,244  
## Cobertura\_medicaPor su propio trabajo -1,911e+01 1,490e+03 -0,013  
## Estudios\_MadreEstudios primarios completos 2,316e+01 1,063e+01 2,179  
## Estudios\_MadreEstudios primarios incompletos 2,369e+01 2,102e+03 0,011  
## Estudios\_MadreEstudios secundarios completos 1,360e+01 6,113e+00 2,225  
## Estudios\_MadreEstudios secundarios incompletos 1,208e+01 6,247e+00 1,934  
## Estudios\_MadreEstudios superiores completos 2,886e+00 3,335e+00 0,866  
## Estudios\_MadreEstudios superiores incompletos 8,201e+01 1,079e+04 0,008  
## Estudios\_MadreEstudios universitarios completos 1,417e+01 6,865e+00 2,064  
## Estudios\_MadreEstudios universitarios incompletos 2,295e+01 1,035e+01 2,216  
## Estudios\_MadreNo hizo estudios 4,986e+01 1,531e+04 0,003  
## Estudios\_PadreEstudios primarios completos -2,669e+01 1,333e+01 -2,002  
## Estudios\_PadreEstudios primarios incompletos -3,070e+01 1,490e+01 -2,061  
## Estudios\_PadreEstudios secundarios completos -3,037e+01 1,496e+01 -2,030  
## Estudios\_PadreEstudios secundarios incompletos -2,164e+01 1,112e+01 -1,946  
## Estudios\_PadreEstudios superiores completos -2,941e+01 1,425e+01 -2,065  
## Estudios\_PadreEstudios superiores incompletos -1,879e+00 1,086e+04 0,000  
## Estudios\_PadreEstudios universitarios completos -4,541e+01 2,830e+02 -0,160  
## Estudios\_PadreEstudios universitarios incompletos 2,157e+00 1,511e+01 0,143  
## Estudios\_PadreNo hizo estudios -4,266e+01 1,089e+04 -0,004  
## Año\_Egreso 2,439e+00 1,172e+00 2,081  
## Pr(>|z|)   
## (Intercept) 0,0790 .  
## CarreraLANA 0,7001   
## CarreraTPVI 0,0467 \*  
## Estado\_CivilDivorciado 0,9995   
## Estado\_CivilSeparado 0,9954   
## Estado\_CivilSoltero 0,9791   
## GeneroMasculino 0,0817 .  
## Situacion\_Ocupacional\_AlumnoNT yBT 0,9845   
## Situacion\_Ocupacional\_AlumnoNTyBT 0,0889 .  
## Situacion\_Ocupacional\_AlumnoNTyNBT 0,3249   
## Situacion\_Ocupacional\_AlumnoT 0,2008   
## Porc\_Regularizadas 0,0245 \*  
## edad 0,1948   
## Cobertura\_medicaComo afiliado voluntario 0,3018   
## Cobertura\_medicaOtorgada por la universidad 1,0000   
## Cobertura\_medicaPor ser familiar a cargo 0,2136   
## Cobertura\_medicaPor su propio trabajo 0,9898   
## Estudios\_MadreEstudios primarios completos 0,0294 \*  
## Estudios\_MadreEstudios primarios incompletos 0,9910   
## Estudios\_MadreEstudios secundarios completos 0,0261 \*  
## Estudios\_MadreEstudios secundarios incompletos 0,0531 .  
## Estudios\_MadreEstudios superiores completos 0,3867   
## Estudios\_MadreEstudios superiores incompletos 0,9939   
## Estudios\_MadreEstudios universitarios completos 0,0390 \*  
## Estudios\_MadreEstudios universitarios incompletos 0,0267 \*  
## Estudios\_MadreNo hizo estudios 0,9974   
## Estudios\_PadreEstudios primarios completos 0,0453 \*  
## Estudios\_PadreEstudios primarios incompletos 0,0393 \*  
## Estudios\_PadreEstudios secundarios completos 0,0424 \*  
## Estudios\_PadreEstudios secundarios incompletos 0,0516 .  
## Estudios\_PadreEstudios superiores completos 0,0389 \*  
## Estudios\_PadreEstudios superiores incompletos 0,9999   
## Estudios\_PadreEstudios universitarios completos 0,8725   
## Estudios\_PadreEstudios universitarios incompletos 0,8865   
## Estudios\_PadreNo hizo estudios 0,9969   
## Año\_Egreso 0,0374 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 201,714 on 145 degrees of freedom  
## Residual deviance: 33,227 on 110 degrees of freedom  
## (93 observations deleted due to missingness)  
## AIC: 105,23  
##   
## Number of Fisher Scoring iterations: 18

# Eliminar filas con NA en las variables usadas en el modelo  
ingresantes\_2022\_clean <- na.omit(ingresantes\_2022[, c("reincriptos\_23", "Carrera", "Genero", "Estado\_Civil","Cost\_est\_\_aporte\_familiar", "Situacion\_Ocupacional\_Alumno" , "Porc\_Regularizadas", "edad", "Estudios\_Madre", "Estudios\_Padre")])  
nrow(ingresantes\_2022\_clean)

## [1] 238

nrow(ingresantes\_2022)

## [1] 239

str(ingresantes\_2022\_clean)

## tibble [238 × 10] (S3: tbl\_df/tbl/data.frame)  
## $ reincriptos\_23 : num [1:238] 0 0 0 0 1 0 1 0 1 1 ...  
## $ Carrera : chr [1:238] "TPVI" "TPVI" "TPVI" "TPVI" ...  
## $ Genero : chr [1:238] "Masculino" "Masculino" "Masculino" "Masculino" ...  
## $ Estado\_Civil : chr [1:238] "Soltero" "Casado" "Soltero" "Soltero" ...  
## $ Cost\_est\_\_aporte\_familiar : chr [1:238] "N" "N" "N" "N" ...  
## $ Situacion\_Ocupacional\_Alumno: chr [1:238] "T" "T" "T" "NTyNBT" ...  
## $ Porc\_Regularizadas : num [1:238] 0 0 0 11,1 44,4 0 44,4 0 66,7 77,8 ...  
## $ edad : num [1:238] 42 40 38 37 37 35 28 27 25 25 ...  
## $ Estudios\_Madre : chr [1:238] "Estudios secundarios completos" "Estudios de Post grado" "Estudios primarios completos" "Estudios primarios completos" ...  
## $ Estudios\_Padre : chr [1:238] "Estudios secundarios incompletos" "Estudios secundarios completos" "Estudios superiores completos" "Estudios primarios completos" ...  
## - attr(\*, "na.action")= 'omit' Named int 228  
## ..- attr(\*, "names")= chr "228"

Hay un estudiante de IA que no informa el estado civil Cuantos no informan el año de egreso: 41,donde 27 son de IA, 14 de LANA No se puede incluir en el modelo Cuantos no informan cobertura medica: 59 , 19 IA, 28 de LANA, y 12 de TPVI

una variable que conserva 224 estudiantes es “Tiene\_beca?” que podría tambien probar si puede agregarse.

## Corro el modelo con las varaibles que tienen información de todos los estudiantes

# Ajustar el modelo sin valores faltantes  
m4 <- glm(reincriptos\_23 ~   
 Carrera +  
 Genero +  
 Estado\_Civil+  
 Cost\_est\_\_aporte\_familiar+  
 Situacion\_Ocupacional\_Alumno+  
 Porc\_Regularizadas +   
 edad +  
 Estudios\_Madre +   
 Estudios\_Padre ,  
 data = ingresantes\_2022\_clean, family = binomial)

drop1(m4, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Carrera + Genero + Estado\_Civil + Cost\_est\_\_aporte\_familiar +   
## Situacion\_Ocupacional\_Alumno + Porc\_Regularizadas + edad +   
## Estudios\_Madre + Estudios\_Padre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 105,79 173,79   
## Carrera 2 107,85 171,85 2,057 0,35757   
## Genero 1 110,25 176,25 4,464 0,03462 \*   
## Estado\_Civil 3 106,37 168,37 0,579 0,90131   
## Cost\_est\_\_aporte\_familiar 1 107,93 173,93 2,142 0,14329   
## Situacion\_Ocupacional\_Alumno 4 115,50 175,50 9,708 0,04565 \*   
## Porc\_Regularizadas 1 226,92 292,92 121,128 < 2e-16 \*\*\*  
## edad 1 106,01 172,01 0,217 0,64136   
## Estudios\_Madre 10 124,59 172,59 18,802 0,04285 \*   
## Estudios\_Padre 10 118,68 166,68 12,891 0,22985   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4)

## [1] 173,789

#elimino Estado\_Civil+  
m4\_1 <- glm(reincriptos\_23 ~   
 Carrera +  
 Genero +  
   
 Cost\_est\_\_aporte\_familiar+  
 Situacion\_Ocupacional\_Alumno+  
 Porc\_Regularizadas +   
 edad +  
 Estudios\_Madre +   
 Estudios\_Padre ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m4\_1, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Carrera + Genero + Cost\_est\_\_aporte\_familiar +   
## Situacion\_Ocupacional\_Alumno + Porc\_Regularizadas + edad +   
## Estudios\_Madre + Estudios\_Padre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 106,37 168,37   
## Carrera 2 108,17 166,17 1,805 0,40564   
## Genero 1 110,77 170,77 4,403 0,03587 \*   
## Cost\_est\_\_aporte\_familiar 1 108,42 168,42 2,056 0,15163   
## Situacion\_Ocupacional\_Alumno 4 116,08 170,09 9,717 0,04547 \*   
## Porc\_Regularizadas 1 228,09 288,09 121,722 < 2e-16 \*\*\*  
## edad 1 107,08 167,07 0,708 0,40019   
## Estudios\_Madre 10 125,62 167,62 19,257 0,03711 \*   
## Estudios\_Padre 10 118,91 160,91 12,540 0,25056   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4\_1)

## [1] 168,3677

#elimino Estado\_Civil edad   
m4\_2 <- glm(reincriptos\_23 ~   
 Carrera +  
 Genero +  
   
 Cost\_est\_\_aporte\_familiar+  
 Situacion\_Ocupacional\_Alumno+  
 Porc\_Regularizadas +   
   
 Estudios\_Madre +   
 Estudios\_Padre ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m4\_2, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Carrera + Genero + Cost\_est\_\_aporte\_familiar +   
## Situacion\_Ocupacional\_Alumno + Porc\_Regularizadas + Estudios\_Madre +   
## Estudios\_Padre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 107,08 167,07   
## Carrera 2 109,67 165,67 2,596 0,27311   
## Genero 1 111,28 169,28 4,200 0,04042 \*   
## Cost\_est\_\_aporte\_familiar 1 110,52 168,52 3,442 0,06357 .   
## Situacion\_Ocupacional\_Alumno 4 116,41 168,41 9,339 0,05317 .   
## Porc\_Regularizadas 1 231,08 289,08 124,007 < 2e-16 \*\*\*  
## Estudios\_Madre 10 125,65 165,65 18,572 0,04604 \*   
## Estudios\_Padre 10 119,89 159,89 12,819 0,23398   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4\_2)

## [1] 167,0754

#elimino Estado\_Civil edad carrera  
m4\_3 <- glm(reincriptos\_23 ~   
   
 Genero +  
   
 Cost\_est\_\_aporte\_familiar+  
 Situacion\_Ocupacional\_Alumno+  
 Porc\_Regularizadas +   
   
 Estudios\_Madre +   
 Estudios\_Padre ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m4\_3, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Genero + Cost\_est\_\_aporte\_familiar + Situacion\_Ocupacional\_Alumno +   
## Porc\_Regularizadas + Estudios\_Madre + Estudios\_Padre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 109,67 165,67   
## Genero 1 112,97 166,97 3,295 0,06951 .   
## Cost\_est\_\_aporte\_familiar 1 113,74 167,74 4,070 0,04364 \*   
## Situacion\_Ocupacional\_Alumno 4 122,16 170,16 12,490 0,01406 \*   
## Porc\_Regularizadas 1 234,51 288,51 124,844 < 2e-16 \*\*\*  
## Estudios\_Madre 10 127,39 163,40 17,724 0,05981 .   
## Estudios\_Padre 10 122,28 158,28 12,613 0,24612   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4\_3)

## [1] 165,6712

#elimino Estado\_Civil edad carrera Estudios\_Padre  
m4\_4 <- glm(reincriptos\_23 ~   
   
 Genero +  
   
 Cost\_est\_\_aporte\_familiar+  
 Situacion\_Ocupacional\_Alumno+  
 Porc\_Regularizadas +   
   
 Estudios\_Madre   
 ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m4\_4, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Genero + Cost\_est\_\_aporte\_familiar + Situacion\_Ocupacional\_Alumno +   
## Porc\_Regularizadas + Estudios\_Madre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 122,28 158,28   
## Genero 1 124,13 158,13 1,843 0,17461   
## Cost\_est\_\_aporte\_familiar 1 126,97 160,97 4,688 0,03038 \*   
## Situacion\_Ocupacional\_Alumno 4 130,85 158,85 8,567 0,07289 .   
## Porc\_Regularizadas 1 244,67 278,67 122,386 < 2e-16 \*\*\*  
## Estudios\_Madre 10 138,80 154,80 16,519 0,08570 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4\_4)

## [1] 158,2843

#elimino Estado\_Civil edad carrera Estudios\_Padre Genero   
m4\_5 <- glm(reincriptos\_23 ~   
 Cost\_est\_\_aporte\_familiar+  
 Situacion\_Ocupacional\_Alumno+  
 Porc\_Regularizadas +   
 Estudios\_Madre   
 ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m4\_5, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Cost\_est\_\_aporte\_familiar + Situacion\_Ocupacional\_Alumno +   
## Porc\_Regularizadas + Estudios\_Madre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 124,13 158,13   
## Cost\_est\_\_aporte\_familiar 1 128,42 160,42 4,296 0,03820 \*   
## Situacion\_Ocupacional\_Alumno 4 131,31 157,31 7,179 0,12673   
## Porc\_Regularizadas 1 245,06 277,06 120,932 < 2e-16 \*\*\*  
## Estudios\_Madre 10 140,37 154,37 16,240 0,09298 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4\_5)

## [1] 158,1272

#elimino Estado\_Civil edad carrera Estudios\_Padre Genero Situacion\_Ocupacional\_Alumno  
m4\_6 <- glm(reincriptos\_23 ~   
 Cost\_est\_\_aporte\_familiar+  
   
 Porc\_Regularizadas +   
 Estudios\_Madre   
 ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m4\_6, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Cost\_est\_\_aporte\_familiar + Porc\_Regularizadas +   
## Estudios\_Madre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 131,31 157,31   
## Cost\_est\_\_aporte\_familiar 1 138,90 162,90 7,590 0,00587 \*\*   
## Porc\_Regularizadas 1 258,74 282,74 127,430 < 2e-16 \*\*\*  
## Estudios\_Madre 10 146,43 152,43 15,119 0,12778   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4\_6)

## [1] 157,3062

#elimino Estado\_Civil edad carrera Estudios\_Padre Genero Situacion\_Ocupacional\_Alumno Estudios\_Madre   
m4\_7 <- glm(reincriptos\_23 ~   
 Cost\_est\_\_aporte\_familiar+  
 Porc\_Regularizadas   
 ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m4\_7, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Cost\_est\_\_aporte\_familiar + Porc\_Regularizadas  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 146,43 152,43   
## Cost\_est\_\_aporte\_familiar 1 164,00 168,00 17,576 2,761e-05 \*\*\*  
## Porc\_Regularizadas 1 286,32 290,32 139,891 < 2,2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

AIC(m4\_7)

## [1] 152,4251

El modelo quedó definido en función si costea el estudio con aporte familiar y por el parcentaje de materias Regularizadas

### Chequeo la adecuación del modelo

debo cargar al libreria DHarma

## Warning: package 'DHARMa' was built under R version 4.3.3

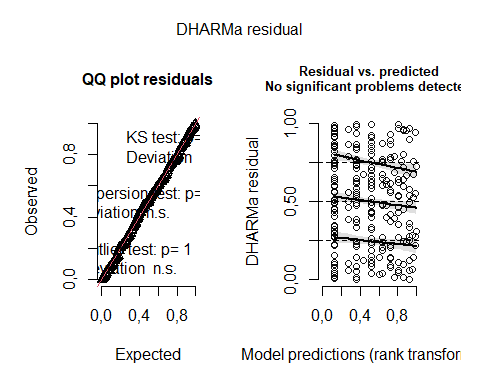
## This is DHARMa 0.4.6. For overview type '?DHARMa'. For recent changes, type news(package = 'DHARMa')

Luego de cargar el programa utilizar la siguiente sentencia: sí

getwd()

## [1] "C:/Users/USUARIO/Desktop/copia de seguridad/trabajo/trabajo/consultorias/consultorias 2022/proyecto IA/analisis ingresantes 2022/toda\_la\_facu\_2022"

sim <-simulateResiduals(fittedModel=m4\_7, plot = T)



#plotResiduals(sim, bd$exclusion)

El modelo se chequera a través de la librería dharma Luego de cargar el programa utilizar la siguiente sentencia: sí sim <- simulateResiduals(fittedModel = m1, plot = T)

# Ajustar el modelo sin valores faltantes  
m5 <- glm(reincriptos\_23 ~   
 Carrera +  
 Genero +  
 Cost\_est\_\_aporte\_familiar+  
 Porc\_Regularizadas   
 ,  
 data = ingresantes\_2022\_clean, family = binomial)

drop1(m4, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Carrera + Genero + Estado\_Civil + Cost\_est\_\_aporte\_familiar +   
## Situacion\_Ocupacional\_Alumno + Porc\_Regularizadas + edad +   
## Estudios\_Madre + Estudios\_Padre  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 105,79 173,79   
## Carrera 2 107,85 171,85 2,057 0,35757   
## Genero 1 110,25 176,25 4,464 0,03462 \*   
## Estado\_Civil 3 106,37 168,37 0,579 0,90131   
## Cost\_est\_\_aporte\_familiar 1 107,93 173,93 2,142 0,14329   
## Situacion\_Ocupacional\_Alumno 4 115,50 175,50 9,708 0,04565 \*   
## Porc\_Regularizadas 1 226,92 292,92 121,128 < 2e-16 \*\*\*  
## edad 1 106,01 172,01 0,217 0,64136   
## Estudios\_Madre 10 124,59 172,59 18,802 0,04285 \*   
## Estudios\_Padre 10 118,68 166,68 12,891 0,22985   
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

drop1(m5, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Carrera + Genero + Cost\_est\_\_aporte\_familiar +   
## Porc\_Regularizadas  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 139,22 151,22   
## Carrera 2 145,86 153,86 6,642 0,0361182 \*   
## Genero 1 141,19 151,19 1,977 0,1597441   
## Cost\_est\_\_aporte\_familiar 1 150,50 160,50 11,286 0,0007809 \*\*\*  
## Porc\_Regularizadas 1 267,30 277,30 128,087 < 2,2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1

# Ajustar el modelo sin valores faltantes  
m6 <- glm(reincriptos\_23 ~   
 Carrera +  
 Cost\_est\_\_aporte\_familiar+  
 Porc\_Regularizadas   
 ,  
 data = ingresantes\_2022\_clean, family = binomial)  
drop1(m6, test="Chisq")

## Single term deletions  
##   
## Model:  
## reincriptos\_23 ~ Carrera + Cost\_est\_\_aporte\_familiar + Porc\_Regularizadas  
## Df Deviance AIC LRT Pr(>Chi)   
## <none> 141,19 151,19   
## Carrera 2 146,43 152,43 5,232 0,073098 .   
## Cost\_est\_\_aporte\_familiar 1 151,64 159,64 10,449 0,001227 \*\*   
## Porc\_Regularizadas 1 267,58 275,58 126,390 < 2,2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1