

UNIVERSIDAD NACIONAL
DE ROSARIO
FACULTAD DE CIENCIAS
ECONÓMICAS Y ESTADÍSTICA



UNR

TRABAJO PRÁCTICO DE MODELOS LINEALES GENERALIZADOS

Integrantes:

Candela, Ornella

Mac Kay, Agustina

Ovando, Francisco

Licenciatura en Estadística

Año 2025

Introducción

Análisis descriptivo

Tabla 1: Frecuencias absolutas y relativas de diabetes

Diabetes	Frecuencia	Porcentaje (%)
No	261	66.8
Sí	130	33.2
Total	391	100

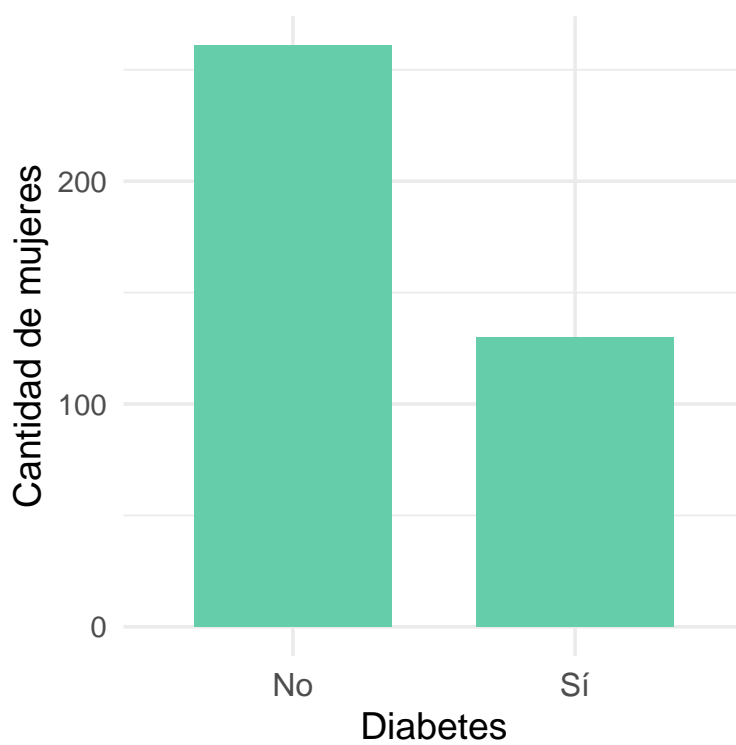


Figura 1: Distribución de diabetes

Tabla 2: Frecuencias absolutas y relativas de obesidad

Obesidad	Frecuencia	Porcentaje (%)
Peso Normal o sobrepeso	129	33
Obesidad	262	67
Total	391	100

Tabla 3: Frecuencias absolutas y relativas de glucosa

Glucosa	Frecuencia	Porcentaje (%)
Normal	116	29.7
Elevada	275	70.3
Total	391	100

Tabla 4: Frecuencias absolutas y relativas de presión

Presión	Frecuencia	Porcentaje (%)
Normal	313	80.1
Elevada	78	19.9
Total	391	100

Tabla 5: Frecuencias absolutas y relativas de DPF

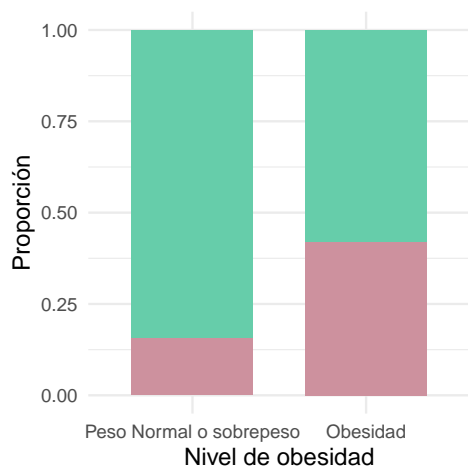
DPF	Frecuencia	Porcentaje (%)
Riesgo bajo	223	57
Riesgo alto	168	43
Total	391	100

Tabla 6: Frecuencias absolutas y relativas de embarazo

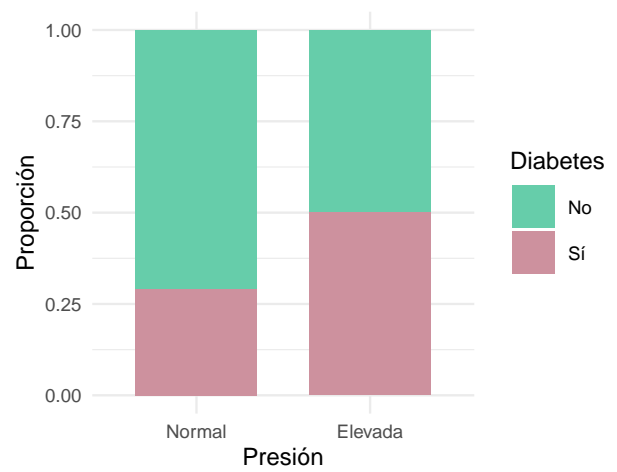
Embarazos	Frecuencia	Porcentaje (%)
Ninguno	56	14.3
1 a 3	202	51.7
Más de 3	133	34
Total	391	100

Tabla 7: Estadísticas descriptivas de la variable edad

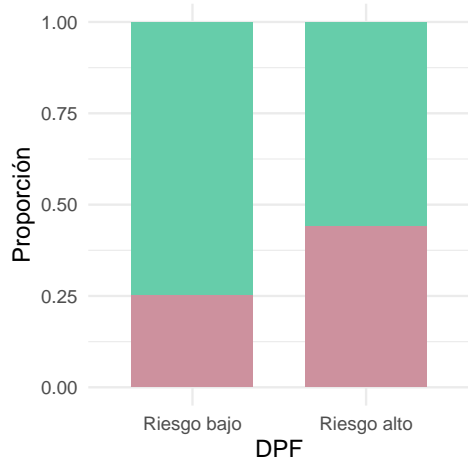
Variable	Media	Mediana	DE	Mínimo	Máximo
Edad	30.73657	27	9.892482	21	63



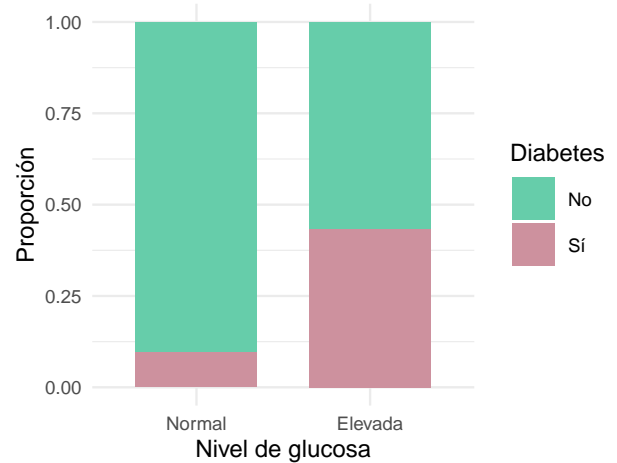
Proporción de personas con diabetes según nivel de obesidad



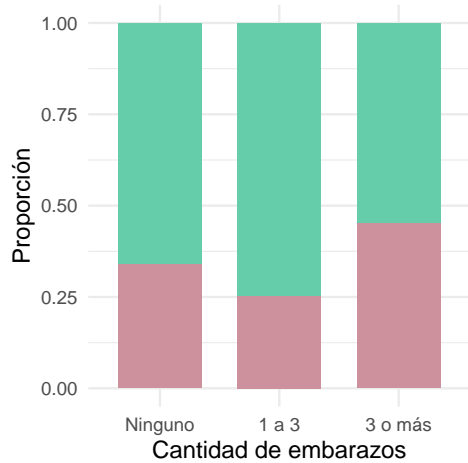
Proporción de personas con diabetes según presión



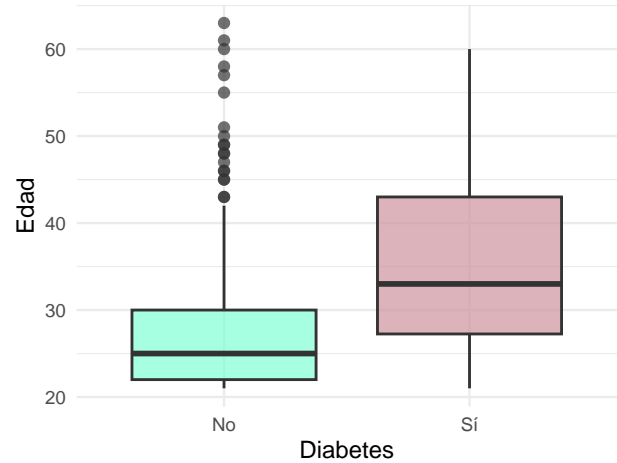
Proporción de personas con diabetes según DPF



Proporción de personas con diabetes según nivel de glucosa



Proporción de personas con diabetes según cantidad de embarazos



Distribución de la edad según presencia de diabetes

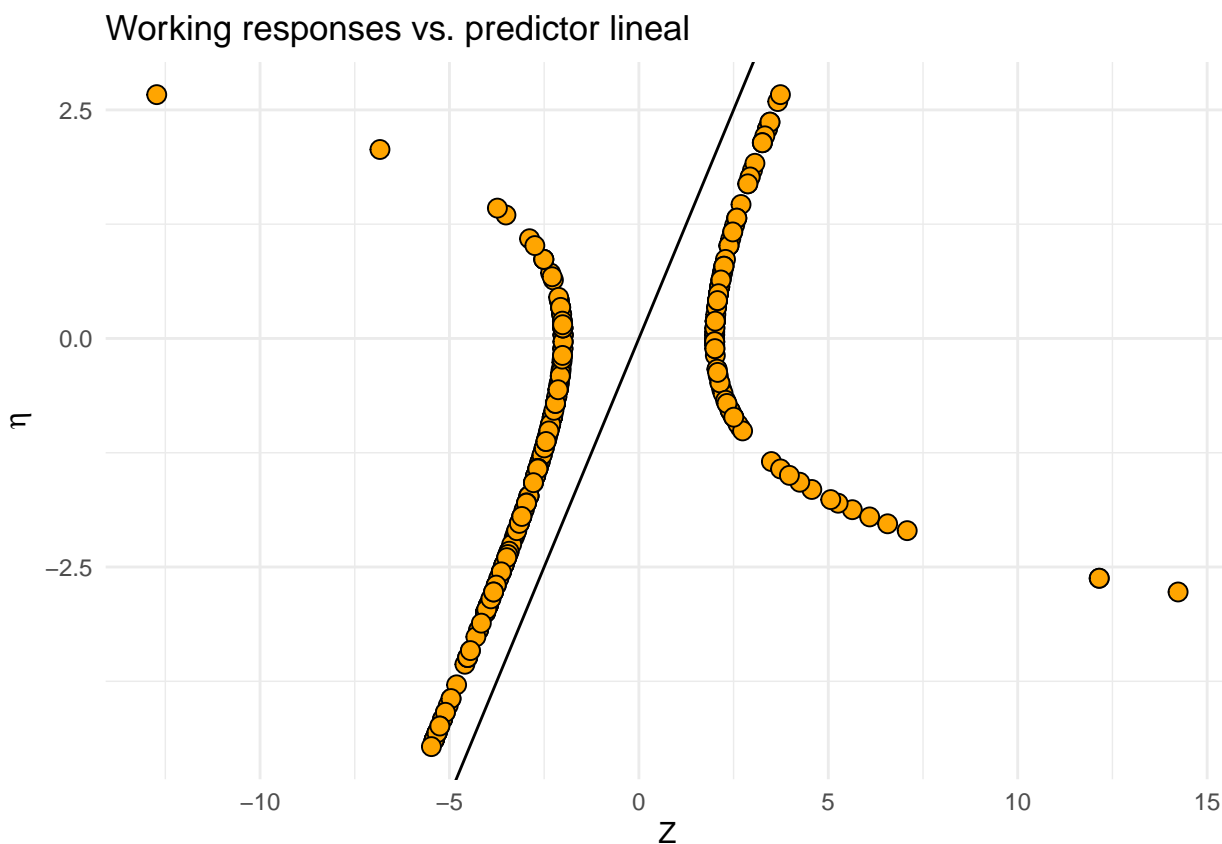
Selección del modelo

```
## Start:  AIC=327.49
## diabetes ~ edad + glucosa + dpf + obesidad
##
##           Df Deviance    AIC
## + dpf:obesidad      1   314.47 326.47
## <none>                317.49 327.49
## + edad:glucosa       1   317.35 329.35
## + glucosa:obesidad   1   317.45 329.45
## + edad:obesidad      1   317.46 329.46
## + glucosa:dpf        1   317.47 329.47
## + edad:dpf           1   317.48 329.48
## - dpf                1   330.10 338.10
## - obesidad           1   338.02 346.02
## - edad               1   347.22 355.22
## - glucosa            1   347.52 355.52
##
## Step:  AIC=326.47
## diabetes ~ edad + glucosa + dpf + obesidad + dpf:obesidad
##
##           Df Deviance    AIC
## <none>                314.47 326.47
## - dpf:obesidad      1   317.49 327.49
## + edad:obesidad     1   314.37 328.37
## + edad:glucosa       1   314.37 328.37
## + glucosa:dpf        1   314.41 328.41
## + glucosa:obesidad   1   314.45 328.45
## + edad:dpf           1   314.46 328.46
## - glucosa            1   344.84 354.84
## - edad               1   346.04 356.04
```

Bondad de ajuste

```
## [1] 0.6365891
```

Verificación el enlace elegido



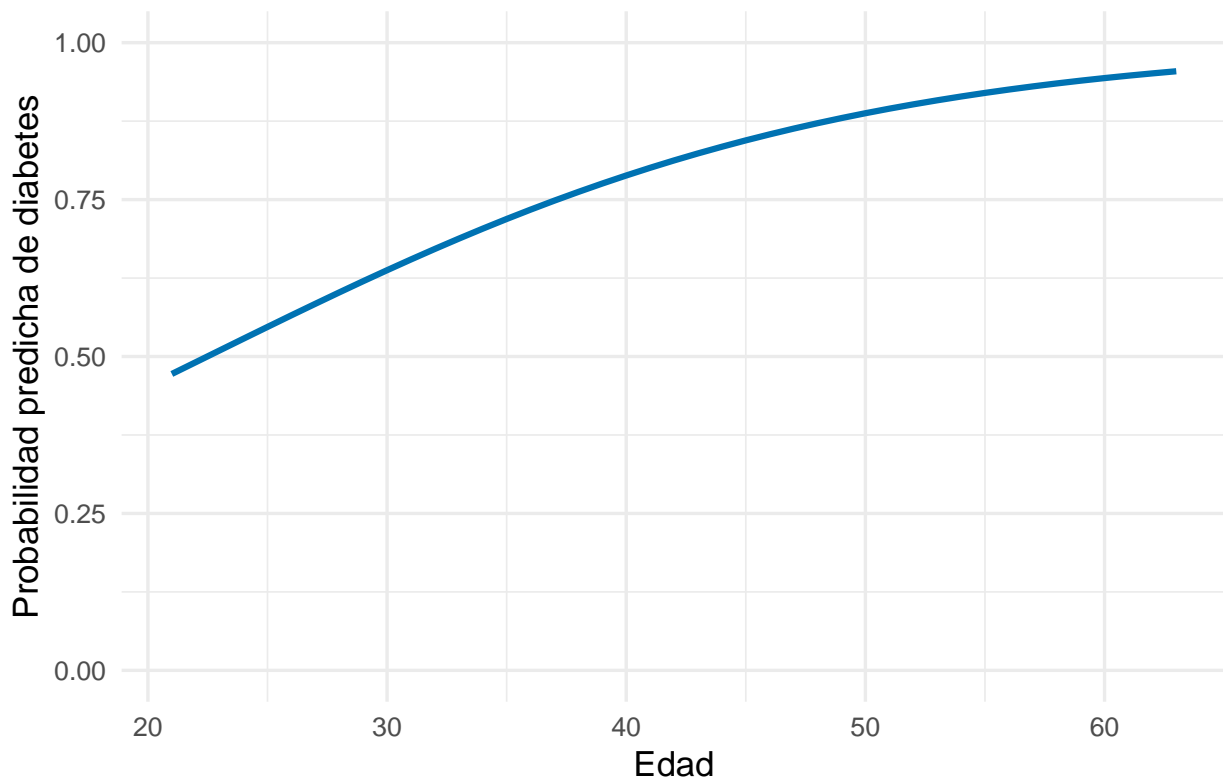
```
## Analysis of Deviance Table
##
## Model 1: diabetes ~ edad + glucosa + dpf + obesidad
## Model 2: diabetes ~ glucosa + obesidad + edad + dpf + pred.2.logit
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       327      317.49
## 2       326      317.28  1   0.20728   0.6489

##
## Call:
## glm(formula = diabetes ~ glucosa + obesidad + edad + dpf + pred.2.logit,
##      family = binomial(link = "logit"), data = datos_train)
##
## Coefficients:
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)  -5.77245    0.91927  -6.279 0.00000000034 ***
## glucosa1      1.78401    0.48396   3.686   0.000228 ***
## obesidad1     1.38501    0.39021   3.549   0.000386 ***
```

```
## edad          0.07408    0.01475    5.021 0.00000051330 ***
## dpf1          0.94772    0.28436    3.333    0.000860 ***
## pred.2.logit -0.03576    0.07933   -0.451    0.652138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 428.40  on 331  degrees of freedom
## Residual deviance: 317.28  on 326  degrees of freedom
## AIC: 329.28
##
## Number of Fisher Scoring iterations: 6
```

Curvas ajustadas

Probabilidad de diabetes según edad (dpf, obesidad y gluc



Análisis de linealidad en la edad

```
## 20% 40% 60% 80%
```

```
## 23 25 29 38

##
## Call:
## glm(formula = diabetes ~ glucosa + obesidad + dpf + presion +
##      edad_2 + edad_3 + edad_4 + edad_5, family = binomial(link = "logit"),
##      data = datos_train)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.1272     0.6492  -7.898 2.83e-15 ***
## glucosa1      1.9719     0.4124   4.782 1.73e-06 ***
## obesidad1     1.3182     0.3589   3.673 0.000240 ***
## dpf1          1.0491     0.2868   3.658 0.000254 ***
## presion1      0.5076     0.3395   1.495 0.134868
## edad_2        1.0651     0.5202   2.047 0.040614 *
## edad_3        1.2084     0.4867   2.483 0.013045 *
## edad_4        1.6494     0.4874   3.384 0.000714 ***
## edad_5        2.5325     0.4852   5.219 1.80e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 428.40  on 331  degrees of freedom
## Residual deviance: 307.12  on 323  degrees of freedom
## AIC: 325.12
##
## Number of Fisher Scoring iterations: 5

##
## Call:
## glm(formula = diabetes ~ glucosa + obesidad + dpf + presion +
##      edad_ORDEN, family = binomial(link = "logit"), data = datos_train)
##
## Coefficients:
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)  -5.4969     0.6519  -8.432    < 2e-16 ***
## glucosa1      1.9229     0.4080   4.713 0.0000024363 ***
## obesidad1     1.3270     0.3493   3.799   0.000145 ***
## dpf1          1.0507     0.2855   3.680   0.000233 ***
## presion1      0.4791     0.3323   1.441   0.149451
## edad_ORDEN    0.5649     0.1043   5.416 0.0000000609 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 428.40  on 331  degrees of freedom
## Residual deviance: 308.96  on 326  degrees of freedom
## AIC: 320.96
##
## Number of Fisher Scoring iterations: 5

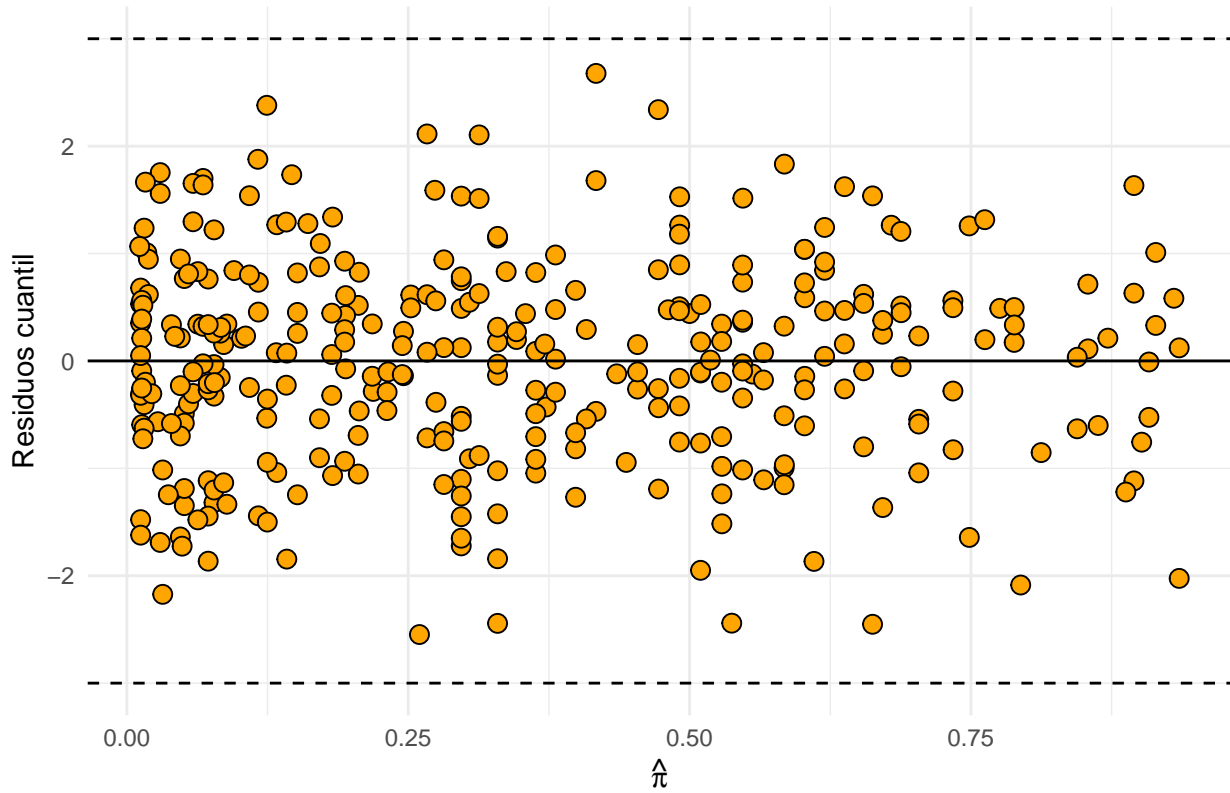
## Analysis of Deviance Table
##
## Model 1: diabetes ~ glucosa + obesidad + dpf + presion + edad_ORDEN
## Model 2: diabetes ~ glucosa + obesidad + dpf + presion + edad_2 + edad_3 +
##          edad_4 + edad_5
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         326       308.96
## 2         323       307.12  3    1.8439    0.6054
```

Análisis de residuos

Evaluación de la componente sistemática

Observe como se ve el gráfico de los residuos vs. las medias estimadas bajo el enlace logit:

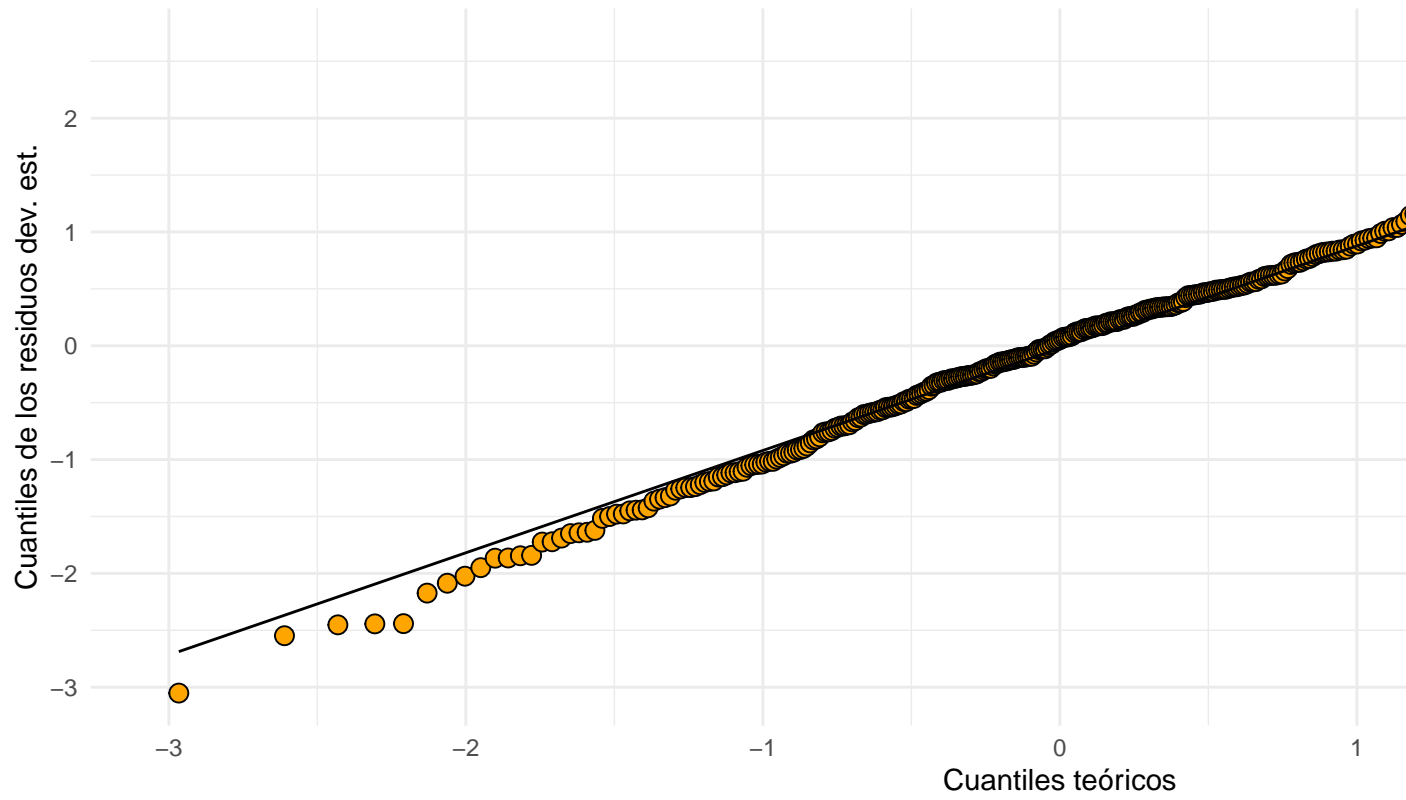
Gráfico de residuos cuantil vs. las probabilidades estimadas



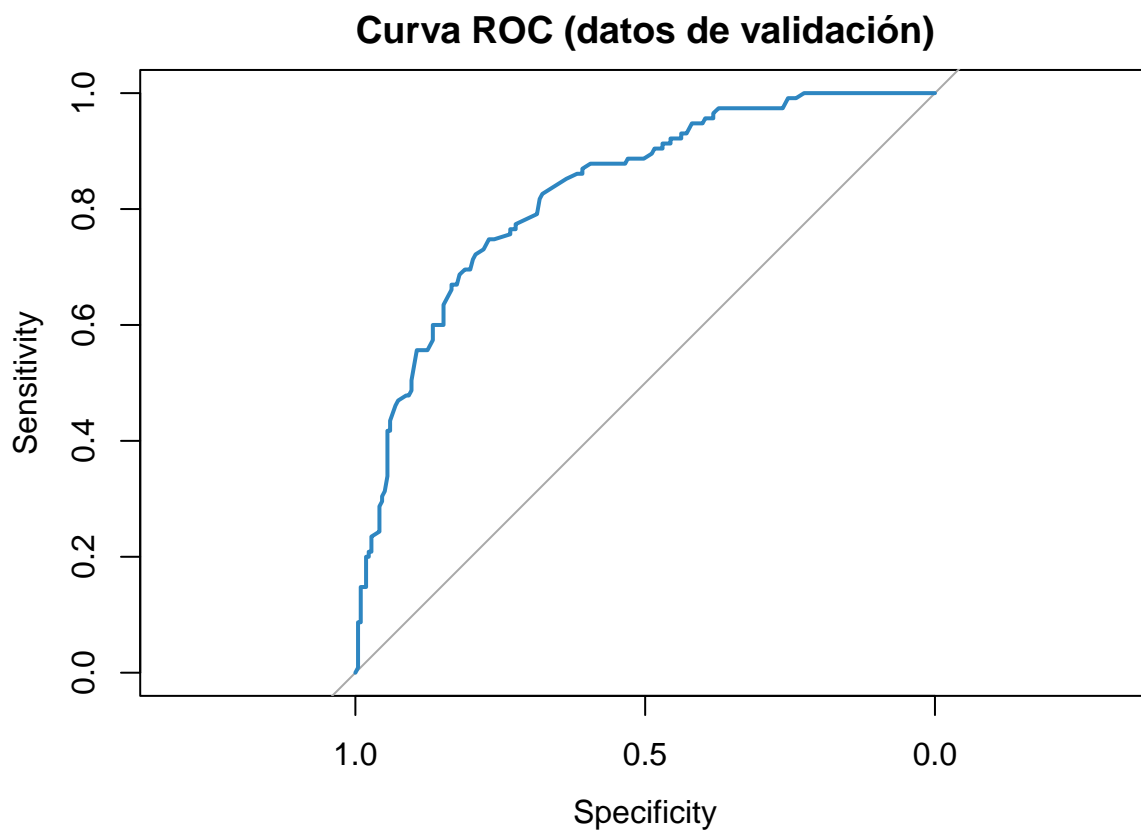
Evaluación de la componente aleatoria

Mediante un gráfico QQ puede evaluarse si la distribución supuesta es adecuada.

Gráfico probabilístico normal con residuos deviance est.



Capacidad predictiva del modelo



```
## threshold
## 1 0.3767094
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##           Reference
```

```
## Prediction 0  1
```

```
##           0 31  7
```

```
##           1 10 11
```

```
##
```

```
##           Accuracy : 0.7119
```

```
##           95% CI : (0.5792, 0.8224)
```

```
## No Information Rate : 0.6949
```

```
## P-Value [Acc > NIR] : 0.4511
```

```
##
```

```
##           Kappa : 0.3508
```

```
##
```

```
## McNemar's Test P-Value : 0.6276
```

```
##
##          Sensitivity : 0.6111
##          Specificity : 0.7561
##          Pos Pred Value : 0.5238
##          Neg Pred Value : 0.8158
##          Prevalence : 0.3051
##          Detection Rate : 0.1864
##          Detection Prevalence : 0.3559
##          Balanced Accuracy : 0.6836
##
##          'Positive' Class : 1
##
```

Resultados

Conclusión