

Modelos Lineares Generalizados

unificação metodológica

Alexandre Adalardo de Oliveira

PlanECO 2025

Modelos Lineares Generalizados

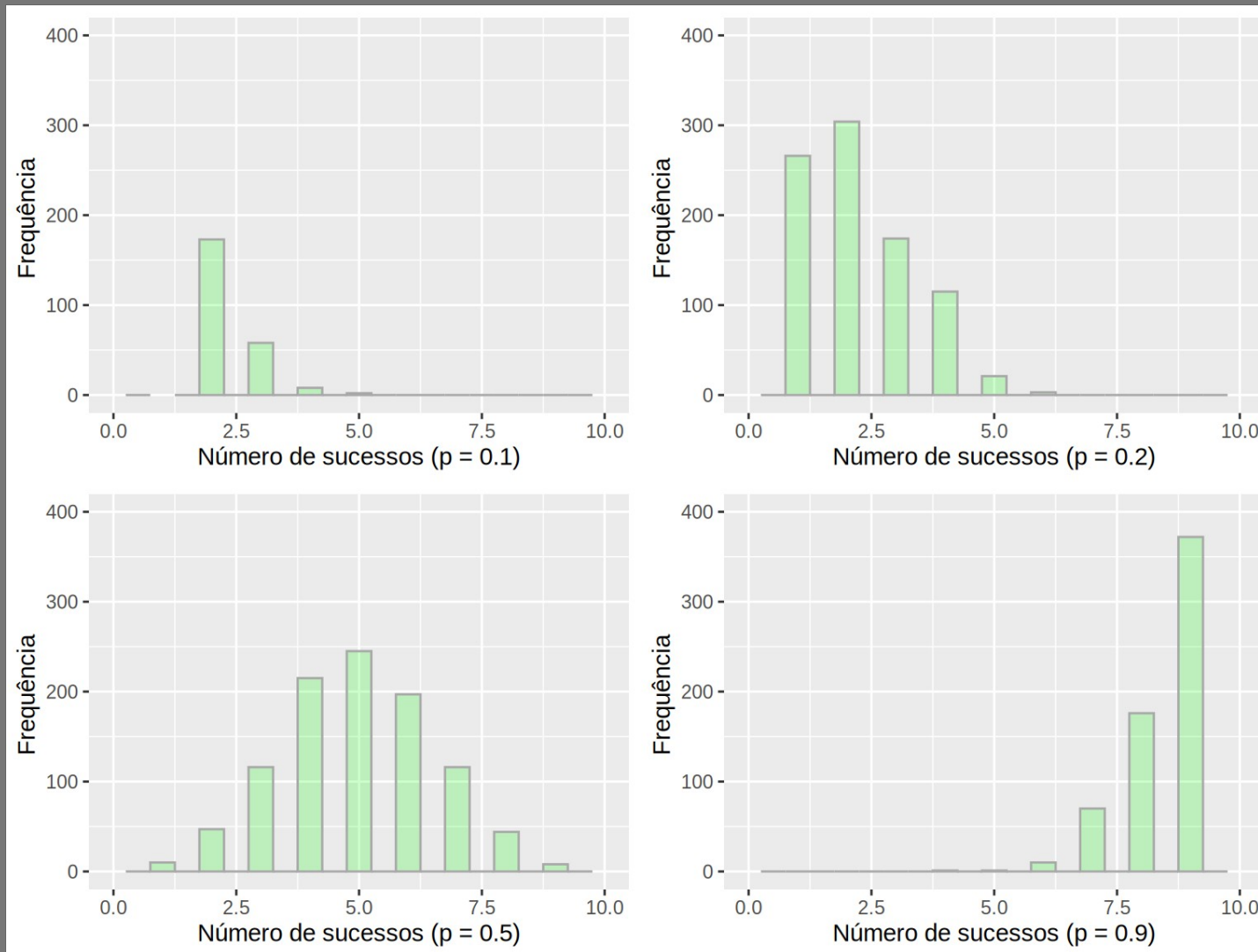
GLM Binomial

- Binária:
 - sim x não
 - vivo x morto
 - germinou x não germinou
- Proporção
 - sucessos/tentativas

limite [0-1]

variância depende da média

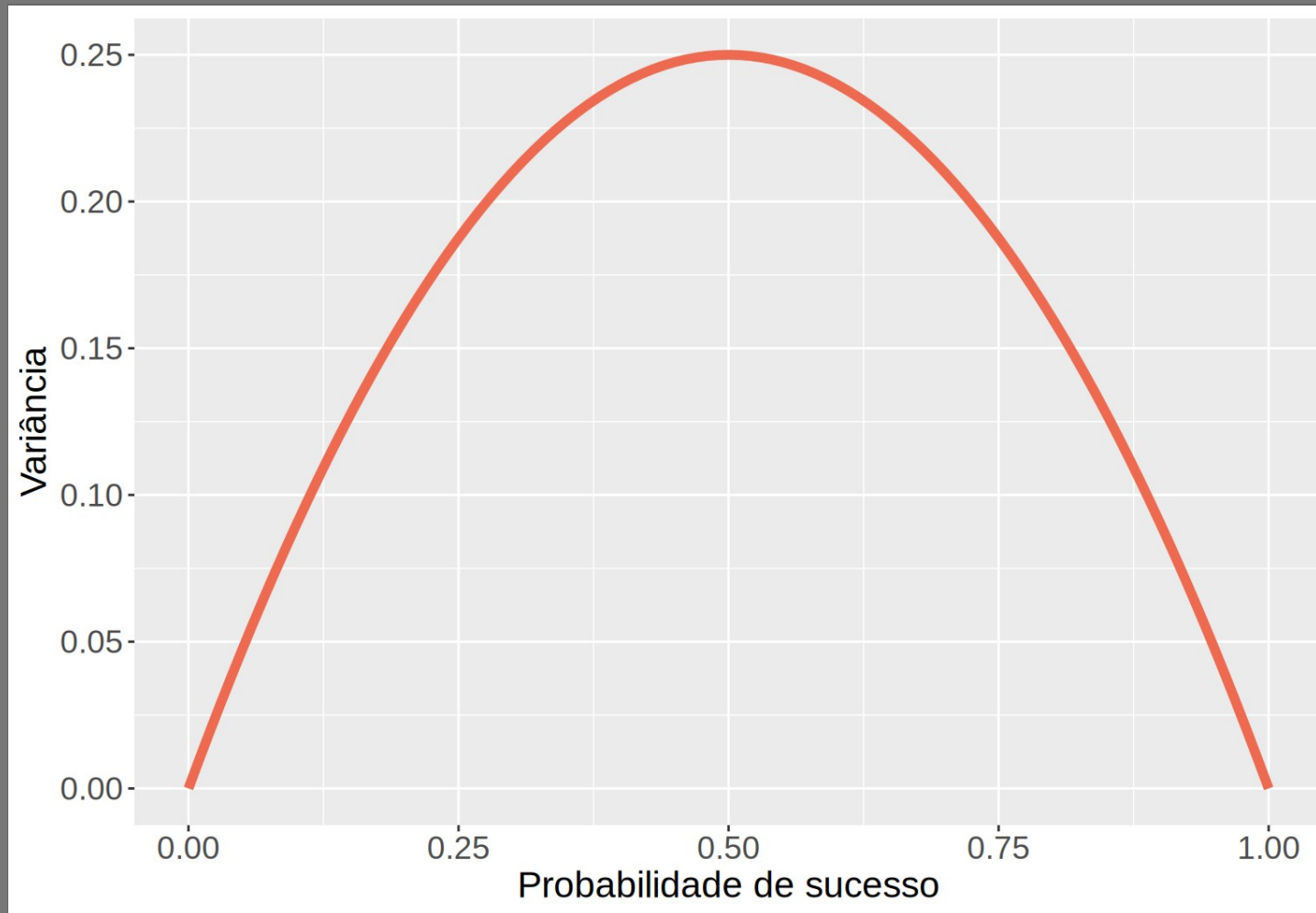
GLM Binomial





GLM Binomial: variância

$$\sigma^2 = np(1 - p)$$





GLM Binomial

- **probabilidade: sucessos/tentativas**
- chance: sucessos/falhas
n = tentativas
s = sucessos
f = falhas

Probabilidade de sucesso

$$p = \frac{s}{n}$$

Probabilidade de falha

$$q = \frac{f}{n}$$
$$q = 1 - p$$

GLM Binomial

- probabilidade : sucessos/tentativas
- **chance: sucessos/falhas**

n = tentativas

s = sucessos

f = falhas

CHANCE

$$\text{odds} = \frac{s}{f} = \frac{p}{1 - p}$$

GLM Binomial: logit

FUNÇÃO DE LIGAÇÃO

Logaritmo da Chance (log(Odds))

$$\eta = \log\left(\frac{p}{1-p}\right)$$

$$\log\left(\frac{p}{1-p}\right) = a + \sum (b_i x_i)$$

Razão de Chance

odds ratio

$$\text{odds ratio} = \frac{\text{odds}_{\text{trat}}}{\text{odds}_{\text{control}}}$$

PREDITORA CATEGÓRICA

$$\text{intercept} = \log(\text{odds}_{\text{control}})$$

$$\text{coef}_{\text{level}_i} = \log(\text{odds}_{\text{level}_i}) - \log(\text{odds}_{\text{control}})$$

COEFICIENTE NA ESCALA ODDS RATIO

$$\exp(\log(\text{odds}_{\text{trat}}) - \log(\text{odds}_{\text{control}})) = \frac{\text{odds}_{\text{trat}}}{\text{odds}_{\text{control}}}$$

Razão de Chance (odds ratio)

$$\text{odds ratio} = \frac{\text{odds}_{\text{trat}}}{\text{odds}_{\text{control}}}$$

PREDITORA CONTÍNUA

$$\text{coef}_X = \log(\text{odds}_{x+1}) - \log(\text{odds}_x)$$

COEFICIENTE NA ESCALA ODDS RATIO

$$\exp(\log(\text{odds}_{x+1}) - \log(\text{odds}_x)) = \frac{\text{odds}_{x+1}}{\text{odds}_x}$$

GLM: binária

Exemplo: pássaros na ilha

incidence	area	isolation
1	7.93	3.32
0	1.93	7.55
1	2.04	5.88
0	4.78	5.93
0	1.54	5.31
1	7.37	4.93
1	8.60	2.88
0	2.42	8.77
1	6.40	6.09
1	7.20	6.98
0	2.65	7.75
1	4.13	4.30

incidence	area	isolation
0	4.18	8.52
1	7.10	3.32
0	2.39	9.29

Modelo Binomial: pássaros

```
glmave <- glm(incidence~ area + isolation + area:isolation,  
              family=binomial, data = ave)  
summary(glmave)
```

```
Call:  
glm(formula = incidence ~ area + isolation + area:isolation,  
     family = binomial, data = ave)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.0313	7.1747	0.562	0.574
area	1.3807	2.1373	0.646	0.518
isolation	-0.9422	1.1689	-0.806	0.420
area:isolation	-0.1291	0.3389	-0.381	0.703

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 68.029 on 49 degrees of freedom
Residual deviance: 28.252 on 46 degrees of freedom
AIC: 36.252

Number of Fisher Scoring iterations: 7



Modelo Binomial: pássaros

```
glmave01<- glm(incidence~ area + isolation,  
               family=binomial, data = ave)  
anova(glmave01, glmave, test="Chisq")
```

Analysis of Deviance Table

Model 1: incidence ~ area + isolation

Model 2: incidence ~ area + isolation + area:isolation

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	47	28.402			
2	46	28.252	1	0.15043	0.6981

Modelo Binomial: pássaros

```
summary(glmave01)
```

Call:

```
glm(formula = incidence ~ area + isolation, family = binomial,  
     data = ave)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	6.6417	2.9218	2.273	0.02302	*
area	0.5807	0.2478	2.344	0.01909	*
isolation	-1.3719	0.4769	-2.877	0.00401	**

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 68.029 on 49 degrees of freedom
Residual deviance: 28.402 on 47 degrees of freedom
AIC: 34.402

Number of Fisher Scoring iterations: 6



Modelo Binomial: pássaros

```
glmave02<- glm(incidence~ area,  
               family=binomial, data = ave)
```

```
anova(glmave02, glmave01, test="Chisq")
```

Analysis of Deviance Table

Model 1: incidence ~ area

Model 2: incidence ~ area + isolation

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	48	50.172			
2	47	28.402	1	21.77	3.073e-06 ***

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

Modelo Binomial: pássaros

```
glmave03<- glm(incidence~ isolation,  
               family=binomial, data = ave)  
anova(glmave03, glmave01, test= "Chisq")
```

Analysis of Deviance Table

Model 1: incidence ~ isolation

Model 2: incidence ~ area + isolation

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	48	36.640			
2	47	28.402	1	8.2375	0.004103 **

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

Modelo Binomial: pássaros

```
summary(glmave01)
```

Call:

```
glm(formula = incidence ~ area + isolation, family = binomial,  
     data = ave)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	6.6417	2.9218	2.273	0.02302	*
area	0.5807	0.2478	2.344	0.01909	*
isolation	-1.3719	0.4769	-2.877	0.00401	**

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 68.029 on 49 degrees of freedom
Residual deviance: 28.402 on 47 degrees of freedom
AIC: 34.402

Number of Fisher Scoring iterations: 6



Modelo Binomial: coeficientes

Preditor Linear

```
coef(glmave01)
```

(Intercept)	area	isolation
6.6416611	0.5807241	-1.3719411

Razão de Chance

```
exp(coef(glmave01))
```

(Intercept)	area	isolation
766.3669575	1.7873322	0.2536142

Modelo Binomial: predito

area	isolation	logit
7.928	3.317	6.6949132
1.925	7.554	-2.6040881
2.045	5.883	-0.2418876
4.781	5.932	1.2797485
1.536	5.308	0.2513900
7.369	4.934	4.1518597
8.599	2.882	7.6813735
2.422	8.771	-3.9851205
6.403	6.092	2.0021724
7.199	6.977	1.2502609
2.654	7.748	-2.4468968
4.128	4.297	3.1436593

area	isolation	logit
4.175	8.516	-2.6172662
7.098	3.318	6.2115403
2.392	9.292	-4.7173235

Modelo Binomial: probabilidade

$$\text{logit}^{-1} = \frac{e^{\eta}}{1 + e^{\eta}}$$

area	isolation	logit	probabilidade
7.928	3.317	6.695	0.999
1.925	7.554	-2.604	0.069
2.045	5.883	-0.242	0.440
4.781	5.932	1.280	0.782
1.536	5.308	0.251	0.563
7.369	4.934	4.152	0.985
8.599	2.882	7.681	1.000
2.422	8.771	-3.985	0.018
6.403	6.092	2.002	0.881
7.199	6.977	1.250	0.777

Gráfico do modelo



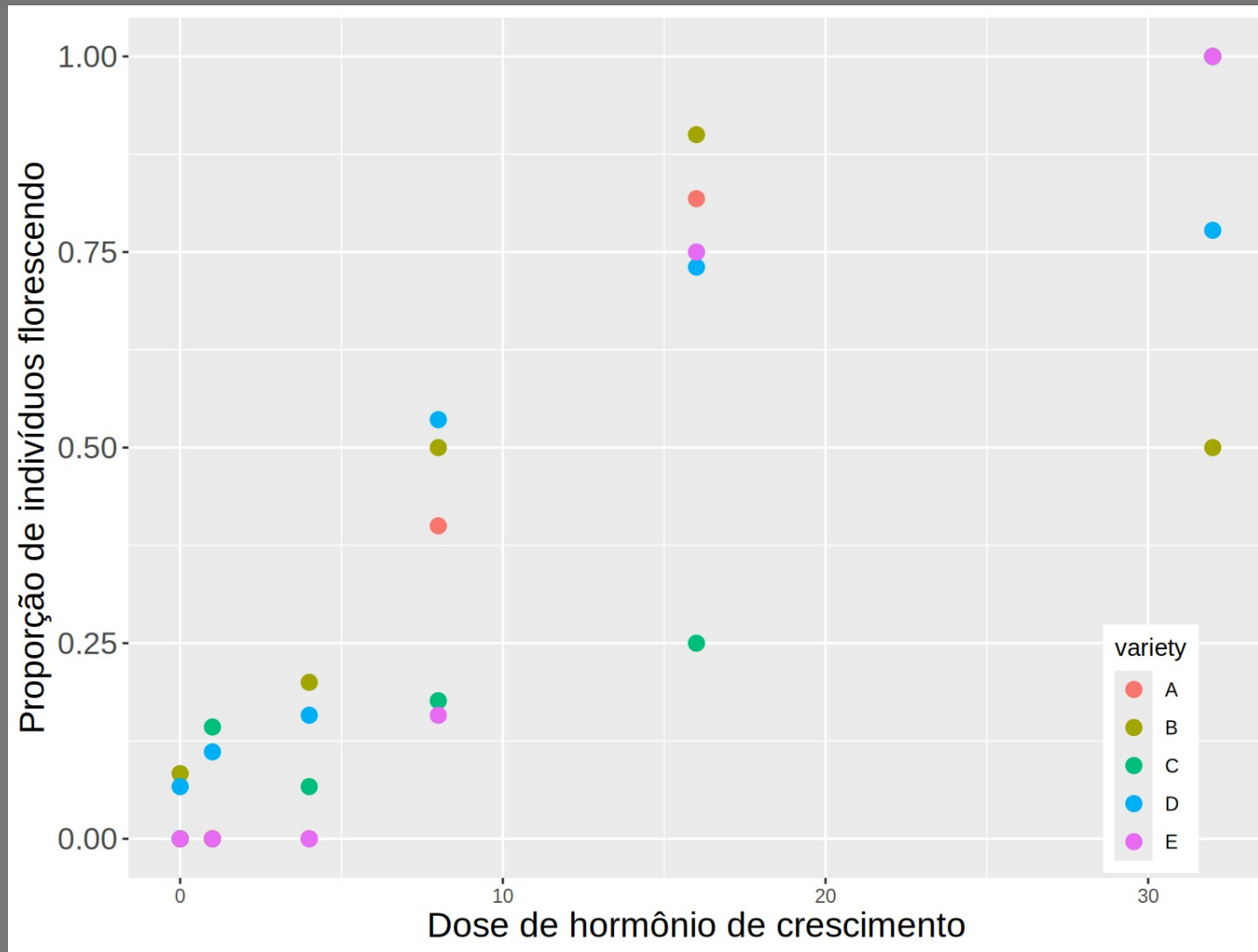


GLM: proporção

GLM binomial: florescer

flowered	number	dose	variety
0	12	1	A
0	17	4	A
0	17	1	B
3	15	4	B
2	14	1	C
1	15	4	C
2	18	1	D
3	19	4	D

Gráfico dos dados





Variável resposta

- combinação de sucesso e falhas

```
yb <- cbind(sucesso = flor$flowered,  
            falha= flor$number-flor$flowered)  
kable(yb[1:10,])
```

sucesso	falha
0	12
0	17
4	6
9	2
10	0
0	17
3	12
6	6
9	1

sucesso	falha
9	9

GLM Binomial: ajustando

(sucesso, falha) ~ dose + variety + dose:variety

```
bin01 <- glm(yb ~ dose + variety + dose:variety, data=flor,  
             family=binomial(link = "logit"))
```

GLM Binomial: resumo

```
Call:
glm(formula = yb ~ dose + variety + dose:variety, family =
  binomial(link = "logit"),
  data = flor)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.59165	1.03215	-4.449	8.64e-06	***
dose	0.41262	0.10033	4.113	3.91e-05	***
varietyB	3.06197	1.09317	2.801	0.005094	**
varietyC	1.23248	1.18812	1.037	0.299576	
varietyD	3.17506	1.07516	2.953	0.003146	**
varietyE	-0.71466	1.54849	-0.462	0.644426	
dose:varietyB	-0.34282	0.10239	-3.348	0.000813	***
dose:varietyC	-0.23039	0.10698	-2.154	0.031274	*
dose:varietyD	-0.30481	0.10257	-2.972	0.002961	**
dose:varietyE	-0.00649	0.13292	-0.049	0.961057	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 303.350 on 29 degrees of freedom
Residual deviance: 51.083 on 20 degrees of freedom

AIC: 123.55

GLM Binomial: simplificando

```
bin02 <- glm(yb ~ dose + variety,  
             data=flor, family=binomial)  
anova(bin02, bin01, test = "Chisq")
```

Analysis of Deviance Table

Model 1: yb ~ dose + variety

Model 2: yb ~ dose + variety + dose:variety

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	24	96.769			
2	20	51.083	4	45.686	2.863e-09 ***

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

GLM Binomial: coeficientes

Preditor Linear

(Intercept)	dose	varietyB	varietyC	varietyD
-4.591651493	0.412616333	3.061966652	1.232483782	3.175056472
varietyE	dose:varietyB	dose:varietyC	dose:varietyD	dose:varietyE
-0.714658038	-0.342819661	-0.230386519	-0.304814286	-0.006490092

Razão de Chance

(Intercept)	dose	varietyB	varietyC	varietyD
0.0101361	1.5107653	21.3695423	3.4297377	23.9281711
varietyE	dose:varietyB	dose:varietyC	dose:varietyD	dose:varietyE
0.4893594	0.7097662	0.7942266	0.7372603	0.9935309

Predito pelo modelo: linear

Logaritmo da Chance

dose	variety	logit
1	A	-4.1790352
4	A	-2.9411862
1	B	-1.4598882
4	B	-1.2504982
1	C	-3.1769379
4	C	-2.6302485
1	D	-1.3087930
4	D	-0.9853868
1	E	-4.9001833
4	E	-3.6818046



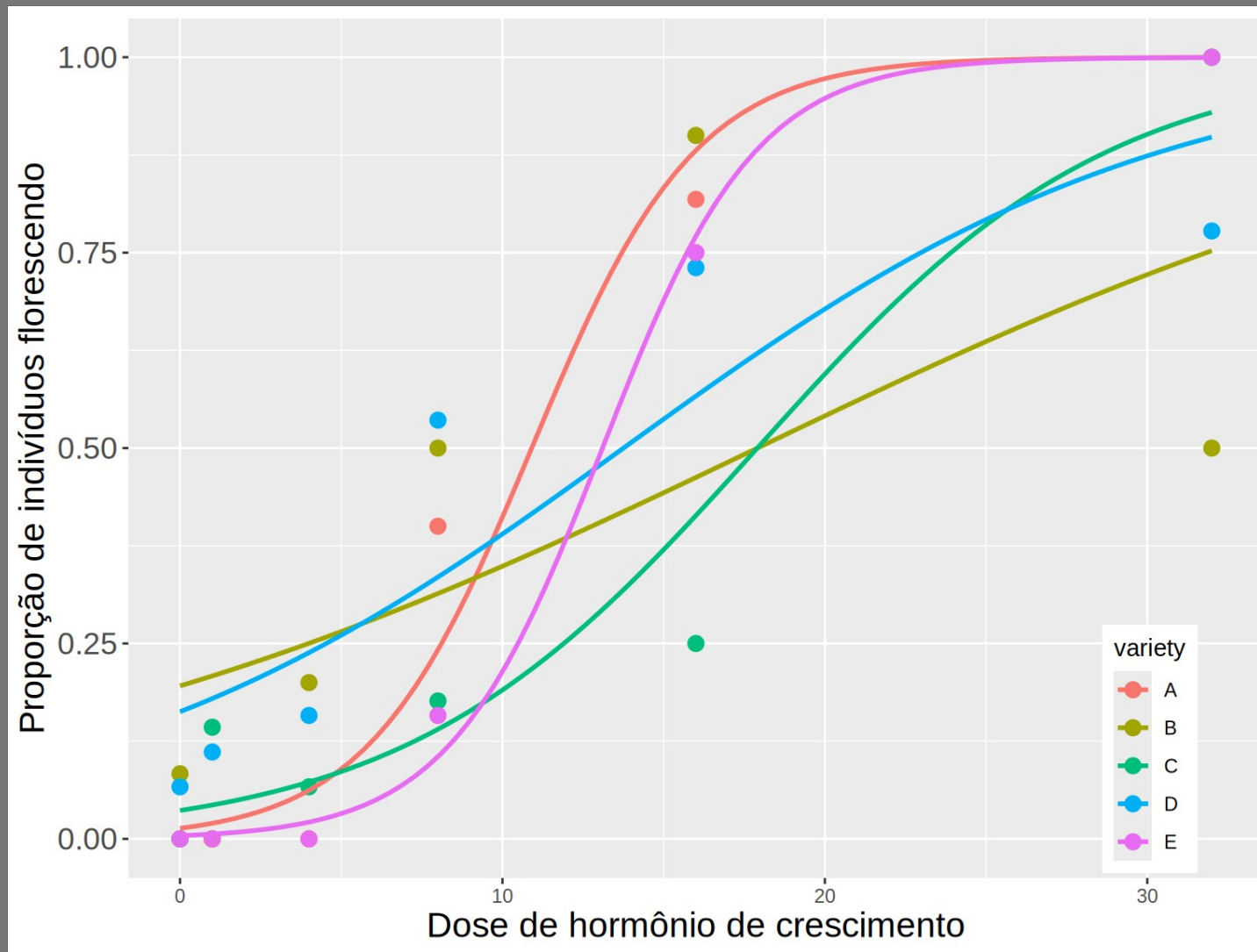
Predito: probabilidade

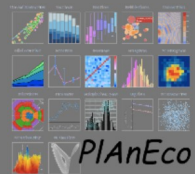
$$\text{logit}^{-1} = \frac{\exp(\eta)}{1 + \exp(\eta)}$$

dose	variety	logit	prob
1	A	-4.1790352	0.0150823
4	A	-2.9411862	0.0501547
1	B	-1.4598882	0.1884844
4	B	-1.2504982	0.2226139
1	C	-3.1769379	0.0400429
4	C	-2.6302485	0.0672169
1	D	-1.3087930	0.2126889
4	D	-0.9853868	0.2718242
1	E	-4.9001833	0.0073902
4	E	-3.6818046	0.0245592



GLM Binomial: florescer





GLM: binomial

- ajuste a variável resposta **(sucesso, falha)** ou **proporção**
- não transforme (sucesso,falha) em proporção a menos que o número de tentativas seja igual para todas as unidades amostrais
- use a família de ligação **binomial(logit)**
- avalie o sobre-dispersão do erro pela razão **Residual deviance / degrees of freedom (apenas proporção)**
- razão > 1, use família **quasibinom**
- busque o modelo mínimo adequado com **anova**
 - binomial use **test = "Chisq"**
 - quasibinom use **test = "F"**
- retenha o modelo mínimo adequado
- retorne os coeficientes e preditos do modelo para escala original (antilogit)

GLM Binomial: FIM