

## Actividad 2

```
# ---- Paquetes necesarios ----  
if(!require("readxl")){install.packages("readxl")}
```

Cargando paquete requerido: readxl

```
if(!require("pwr")){install.packages("pwr")}
```

Cargando paquete requerido: pwr

Warning: package 'pwr' was built under R version 4.4.3

```
if(!require("car")){install.packages("car")}
```

Cargando paquete requerido: car

Cargando paquete requerido: carData

```
if(!require("MASS")){install.packages("MASS")}
```

Cargando paquete requerido: MASS

Warning: package 'MASS' was built under R version 4.4.3

```

library(readxl)
library(pwr)
library(car)
library(MASS)

# ---- Importar y explorar datos ----
# Carga de datos desde Excel
insectos = read_xlsx("C:/Users/dvalh/Desktop/UA/MatesUA/Prácticas de empresa + privado/Activi
summary(insectos) # Visualización básica de los datos

```

Tratamiento	Replica	Conteo
Length:40	Min. : 1.0	Min. : 0.00
Class :character	1st Qu.: 3.0	1st Qu.: 11.75
Mode :character	Median : 5.5	Median : 26.50
	Mean : 5.5	Mean : 44.70
	3rd Qu.: 8.0	3rd Qu.: 55.25
	Max. :10.0	Max. :292.00

```

# ---- Cálculo del tamaño muestral para ANOVA ----
pwr.anova.test(k = 4, n = NULL, f = 0.4, sig.level = 0.05, power = 0.8)

```

Balanced one-way analysis of variance power calculation

```

      k = 4
      n = 18.04262
      f = 0.4
sig.level = 0.05
power = 0.8

```

NOTE: n is number in each group

```

# Estima que n 15 por grupo sería suficiente

# ---- ANOVA tradicional ----
insectos$Tratamiento = as.factor(insectos$Tratamiento)
modelo_aovtr = aov(Conteo ~ Tratamiento, data = insectos)
summary(modelo_aovtr)

```

Df Sum Sq Mean Sq F value Pr(>F)

Tratamiento	3	18256	6085	2.199	0.105
Residuals	36	99609	2767		

```
# ---- Diagnóstico de supuestos ----
# Normalidad de residuos
shapiro.test(resid(modelo_aovtr)) # (p < 0.05 → No hay normalidad)
```

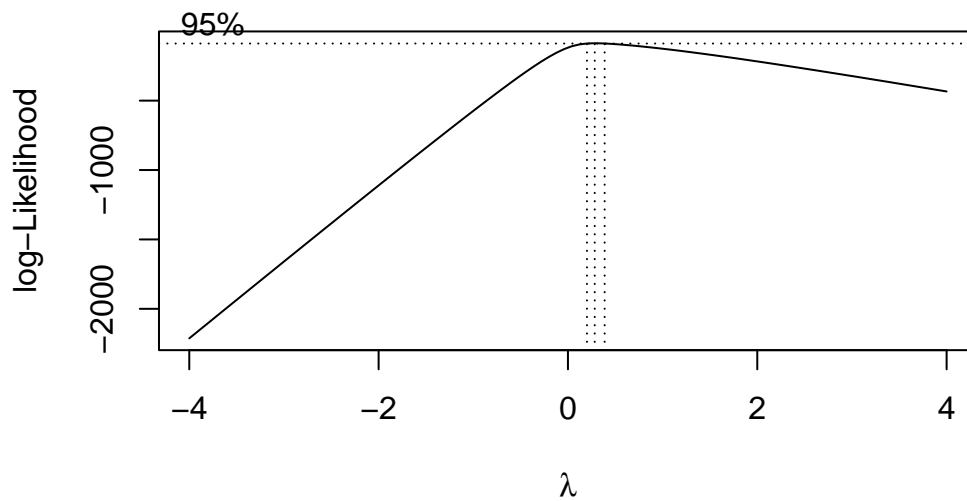
Shapiro-Wilk normality test

```
data: resid(modelo_aovtr)
W = 0.75571, p-value = 9.121e-07
```

```
# Homogeneidad de varianzas
leveneTest(Conteo ~ Tratamiento, data = insectos) # (p > 0.05 → Hay homogeneidad)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  1.0685 0.3746
      36
```

```
# ---- Transformación Box-Cox ----
# Añade una pequeña constante para evitar ceros
insectos$Conteo_mod = insectos$Conteo+0.00001
boxcox_result <- boxcox(insectos$Conteo_mod ~ insectos$Tratamiento, lambda = seq(-4, 4, 0.1))
```



```
lambda_optimo <- boxcox_result$x[which.max(boxcox_result$y)]
print(lambda_optimo)
```

```
[1] 0.2828283
```

```
# Aplica la transformación óptima
if(abs(lambda_optimo) < 0.01) {
  insectos$conteo_trans <- log(insectos$Conteo_mod + 1) # +1 para evitar log(0)
} else {
  insectos$conteo_trans <- ((insectos$Conteo_mod + 1)^lambda_optimo - 1)/lambda_optimo
}

# ANOVA sobre datos transformados
modelo_aov_box <- aov(conteo_trans ~ Tratamiento, data = insectos)
summary(modelo_aov_box)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Tratamiento	3	93.75	31.248	4.427	0.0095 **
Residuals	36	254.14	7.059		

---

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

```
# Comprobación de supuestos en datos transformados
shapiro.test(resid(modelo_aov_box)) # Ahora sí hay normalidad
```

Shapiro-Wilk normality test

```
data:  resid(modelo_aov_box)
W = 0.96634, p-value = 0.274
```

```
leveneTest(conteo_trans ~ Tratamiento, data = insectos) # Hay homogeneidad
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 3  0.1369 0.9373
      36
```

```
# ---- Modelos lineales generalizados (GLM) ----
if(!require("DHARMa")){install.packages("DHARMa")}
```

Cargando paquete requerido: DHARMa

Warning: package 'DHARMa' was built under R version 4.4.3

This is DHARMa 0.4.7. For overview type '?DHARMa'. For recent changes, type news(package = 'DHARMa')

```
library(DHARMa)

# Fija el nivel de referencia (Control)
insectos$Tratamiento <- factor(insectos$Tratamiento,
                              levels = c("Control", "A", "B", "A+B"))

# Modelo GLM Gaussiano
modelo_glm1 <- glm(Conteo ~ Tratamiento, family = gaussian(link = "identity"), data = insectos)
summary(modelo_glm1)
```

```
Call:
glm(formula = Conteo ~ Tratamiento, family = gaussian(link = "identity"),
    data = insectos)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	66.40	16.63	3.992	0.000309	***
TratamientoA	-33.50	23.52	-1.424	0.163036	
TratamientoB	-2.50	23.52	-0.106	0.915955	
TratamientoA+B	-50.80	23.52	-2.159	0.037550	*

---

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

(Dispersion parameter for gaussian family taken to be 2766.906)

Null deviance: 117864 on 39 degrees of freedom  
Residual deviance: 99609 on 36 degrees of freedom  
AIC: 436.32

Number of Fisher Scoring iterations: 2

```
Anova(modelo_glm1)
```

Analysis of Deviance Table (Type II tests)

Response: Conteo

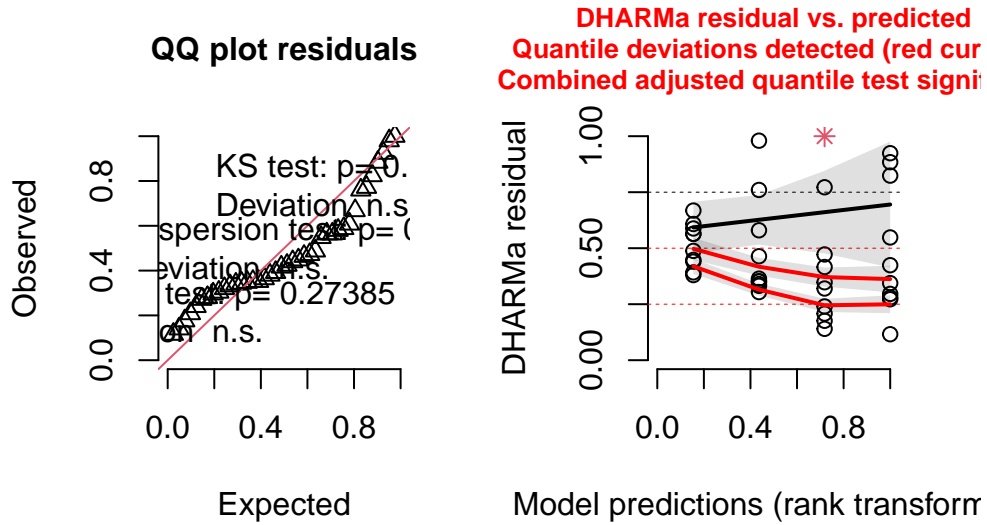
	LR	Chisq	Df	Pr(>Chisq)
Tratamiento	6.5979	3	0.08588	.

---

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

```
residuos_simulados1 <- simulateResiduals(fittedModel = modelo_glm1)
plot(residuos_simulados1)
```

## DHARMA residual



```
# Modelo GLM Poisson
modelo_glm2 <- glm(Conteo ~ Tratamiento, family = poisson(link = "log"), data = insectos)
summary(modelo_glm2)
```

Call:

```
glm(formula = Conteo ~ Tratamiento, family = poisson(link = "log"),
     data = insectos)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.19570	0.03881	108.116	<2e-16 ***
TratamientoA	-0.70222	0.06742	-10.416	<2e-16 ***
TratamientoB	-0.03838	0.05542	-0.693	0.489
TratamientoA+B	-1.44843	0.08897	-16.279	<2e-16 ***

---

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1904.3 on 39 degrees of freedom  
 Residual deviance: 1452.2 on 36 degrees of freedom

AIC: 1657.6

Number of Fisher Scoring iterations: 5

```
Anova(modelo_glm2)
```

Analysis of Deviance Table (Type II tests)

Response: Conteo

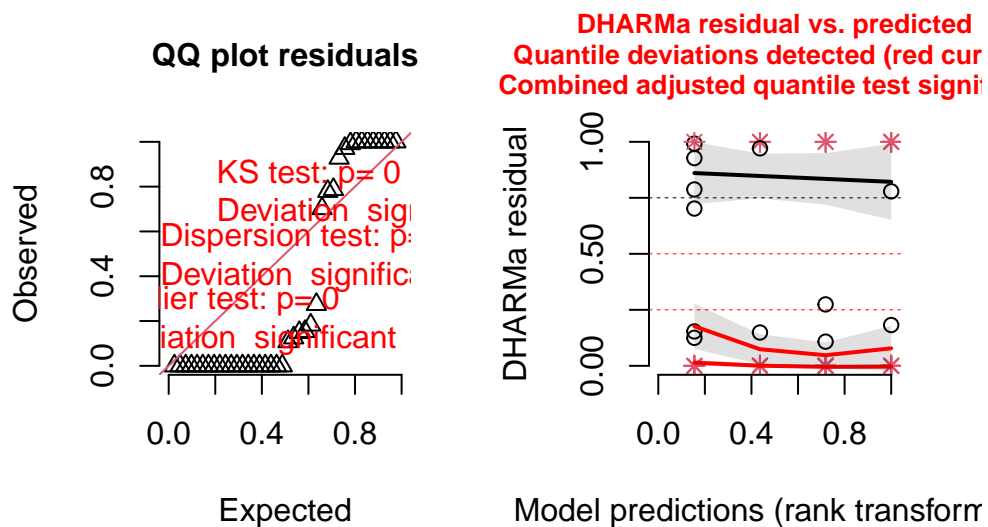
	LR	Chisq	Df	Pr(>Chisq)
Tratamiento	452.09	3	< 2.2e-16	***

---

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

```
residuos_simulados2 <- simulateResiduals(fittedModel = modelo_glm2)  
plot(residuos_simulados2)
```

DHARMA residual



```
# Modelo Binomial Negativo  
modelo_glmnb <- glm.nb(Conteo ~ Tratamiento, data = insectos)  
summary(modelo_glmnb)
```



Call:

```
glm.nb(formula = Conteo ~ Tratamiento, data = insectos, init.theta = 1.166866553,  
       link = log)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.19570	0.29531	14.208	< 2e-16 ***
TratamientoA	-0.70222	0.41946	-1.674	0.094106 .
TratamientoB	-0.03838	0.41770	-0.092	0.926794
TratamientoA+B	-1.44843	0.42346	-3.420	0.000625 ***

---

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

(Dispersion parameter for Negative Binomial(1.1669) family taken to be 1)

Null deviance: 60.235 on 39 degrees of freedom  
Residual deviance: 46.558 on 36 degrees of freedom  
AIC: 382.64

Number of Fisher Scoring iterations: 1

Theta: 1.167  
Std. Err.: 0.258

2 x log-likelihood: -372.636

```
Anova(modelo_glmnb)
```

Analysis of Deviance Table (Type II tests)

Response: Conteo

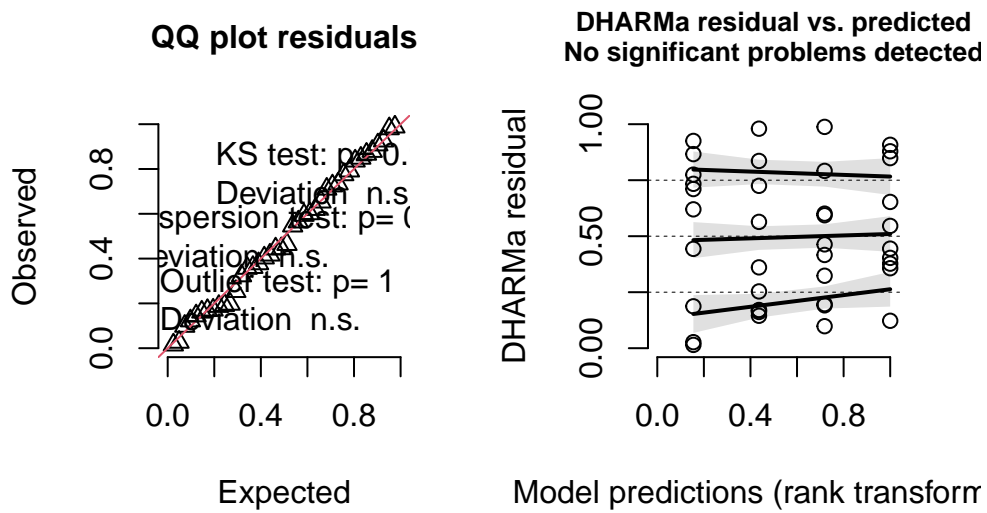
	LR	Chisq	Df	Pr(>Chisq)
Tratamiento	13.677	3	0.00338	**

---

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

```
residuos_simulados3 <- simulateResiduals(fittedModel = modelo_glmnb)  
plot(residuos_simulados3)
```

## DHARMA residual



```
# Este modelo suele ajustarse mejor en recuentos con sobredispersión
```

```
# ---- Modelo beta para proporciones ----
```

```
if(!require("betareg")){install.packages("betareg")}
```

Cargando paquete requerido: betareg

Warning: package 'betareg' was built under R version 4.4.3

```
if(!require("statmod")){install.packages("statmod")}
```

Cargando paquete requerido: statmod

Warning: package 'statmod' was built under R version 4.4.3

```
library(betareg)
```

```
library(statmod)
```

```
# Escala el conteo al rango (0, 1)
```

```
insectos$conteo_esc <- (insectos$Conteo - min(insectos$Conteo)) / (max(insectos$Conteo) - min(insectos$Conteo))
```

```
# Ajusta valores extremos
epsilon <- 1e-4
insectos$conteo_esc[insectos$conteo_esc == 0] <- epsilon
insectos$conteo_esc[insectos$conteo_esc == 1] <- 1 - epsilon

modelo_glmbeta <- betareg(conteo_esc ~ Tratamiento, data = insectos)
summary(modelo_glmbeta)
```

Call:

```
betareg(formula = conteo_esc ~ Tratamiento, data = insectos)
```

Quantile residuals:

	Min	1Q	Median	3Q	Max
	-1.3696	-0.6081	-0.3076	0.1772	4.6053

Coefficients (mean model with logit link):

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.8666	0.3530	-2.455	0.0141 *
TratamientoA	-0.5364	0.4876	-1.100	0.2713
TratamientoB	0.2625	0.4862	0.540	0.5892
TratamientoA+B	-1.0254	0.4893	-2.096	0.0361 *

Phi coefficients (precision model with identity link):

	Estimate	Std. Error	z value	Pr(> z )
(phi)	2.1421	0.4648	4.609	4.05e-06 ***

---

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

Type of estimator: ML (maximum likelihood)

Log-likelihood: 32.14 on 5 Df

Pseudo R-squared: 0.2074

Number of iterations: 14 (BFGS) + 2 (Fisher scoring)

```
Anova(modelo_glmbeta)
```

Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include arithmetic operators in their names;  
the printed representation of the hypothesis will be omitted

Analysis of Deviance Table (Type II tests)

```

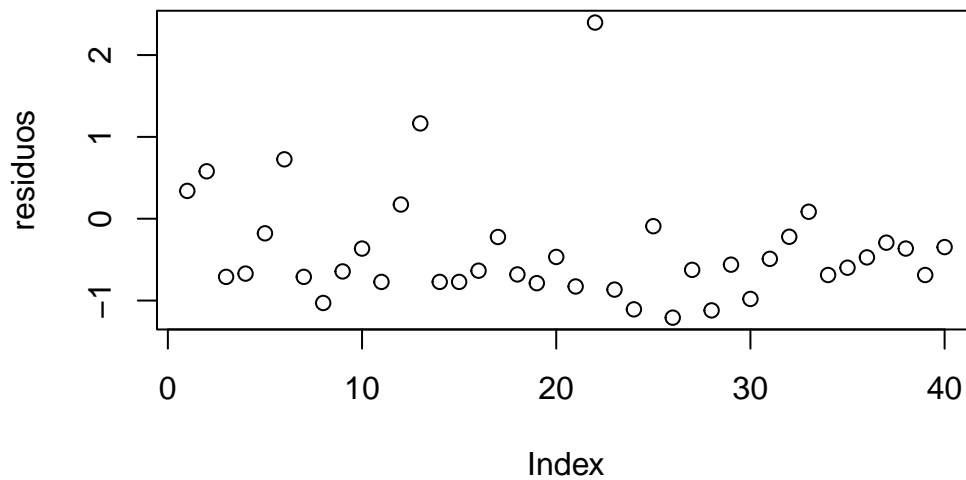
Response: conteo_esc
      Df  Chisq Pr(>Chisq)
Tratamiento  3 8.0406    0.04518 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

residuos <- residuals(modelo_glm-beta, type = "pearson")
plot(residuos)

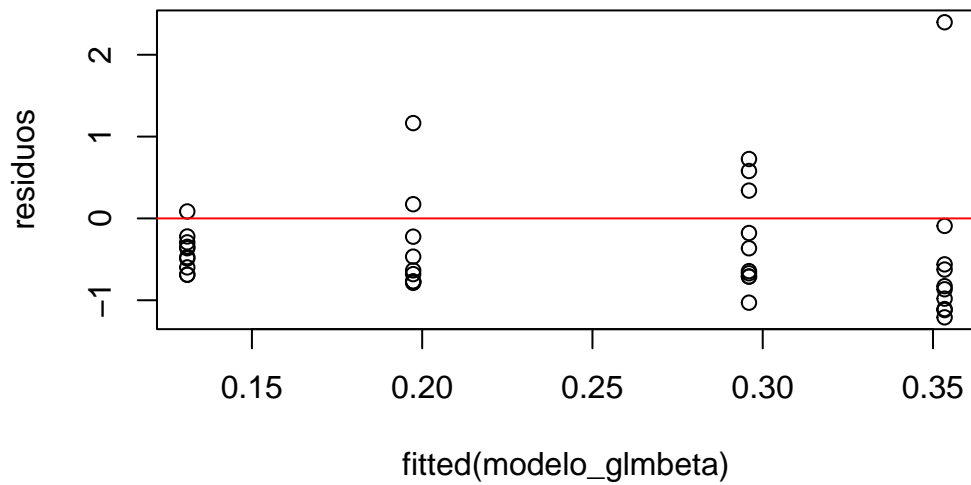
```



```

plot(fitted(modelo_glm-beta), residuos)
abline(h = 0, col = "red")

```



```
# ---- GLM Gamma ----
modelo_glm4 <- glm(Conteo_mod ~ Tratamiento, family = Gamma(link = "log"), data = insectos)
summary(modelo_glm4)
```

Call:

```
glm(formula = Conteo_mod ~ Tratamiento, family = Gamma(link = "log"),
     data = insectos)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.19570	0.33449	12.543	1.05e-14	***
TratamientoA	-0.70222	0.47305	-1.484	0.14639	
TratamientoB	-0.03838	0.47305	-0.081	0.93579	
TratamientoA+B	-1.44843	0.47305	-3.062	0.00414	**

---

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

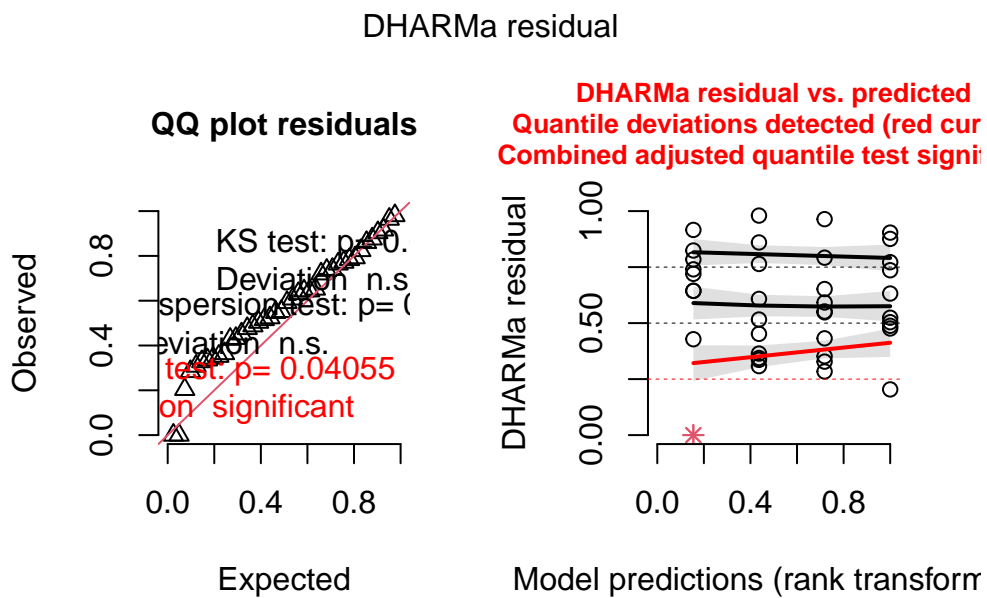
(Dispersion parameter for Gamma family taken to be 1.118864)

Null deviance: 95.597 on 39 degrees of freedom  
Residual deviance: 83.474 on 36 degrees of freedom

AIC: 374.08

Number of Fisher Scoring iterations: 19

```
residuos_simulados4 <- simulateResiduals(fittedModel = modelo_glm4)
plot(residuos_simulados4)
```



```
# ---- Detección de distribución: Cullen & Frey ----
if(!require("fitdistrplus")){install.packages("fitdistrplus")}
```

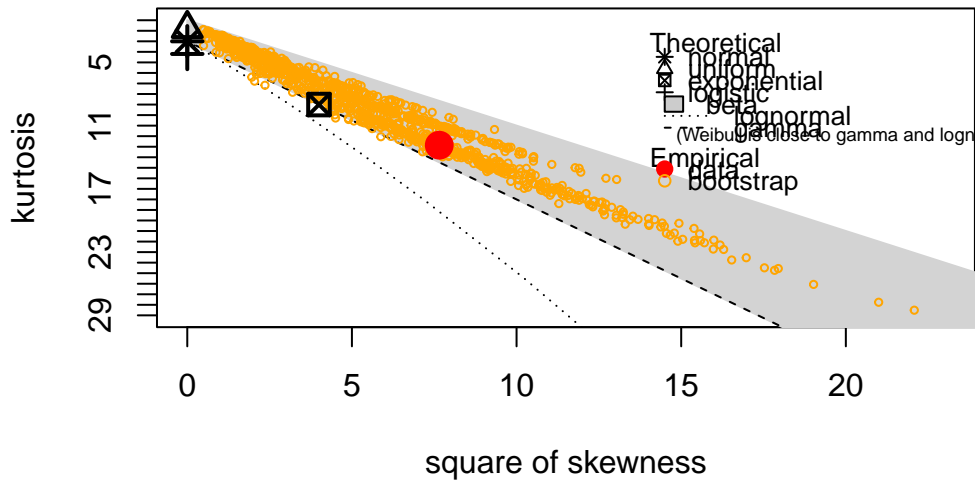
Cargando paquete requerido: fitdistrplus

Warning: package 'fitdistrplus' was built under R version 4.4.3

Cargando paquete requerido: survival

```
library(fitdistrplus)
descdist(insectos$Conteo, boot = 1000)
```

## Cullen and Frey graph



summary statistics

-----

```
min: 0    max: 292
median: 26.5
mean: 44.7
estimated sd: 54.97421
estimated skewness: 2.766358
estimated kurtosis: 12.8552
```

```
# ---- Comparaciones post-hoc (Tukey) para binomial negativa ----
if(!require("multcomp")){install.packages("multcomp")}
```

Cargando paquete requerido: multcomp

Cargando paquete requerido: mvtnorm

Cargando paquete requerido: TH.data

Adjuntando el paquete: 'TH.data'

The following object is masked from 'package:MASS':

geyser

```
library(multcomp)
posthoc <- glht(modelo_glmb, linfct = mcp(Tratamiento = "Tukey"))
summary(posthoc)
```

## Simultaneous Tests for General Linear Hypotheses

### Multiple Comparisons of Means: Tukey Contrasts

```
Fit: glm.nb(formula = Conteo ~ Tratamiento, data = insectos, init.theta = 1.166866553,
  link = log)
```

#### Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z )
A - Control == 0	-0.70222	0.41946	-1.674	0.33738
B - Control == 0	-0.03838	0.41770	-0.092	0.99972
A+B - Control == 0	-1.44843	0.42346	-3.420	0.00361 **
B - A == 0	0.66385	0.41953	1.582	0.38870
A+B - A == 0	-0.74620	0.42526	-1.755	0.29543
A+B - B == 0	-1.41005	0.42353	-3.329	0.00480 **

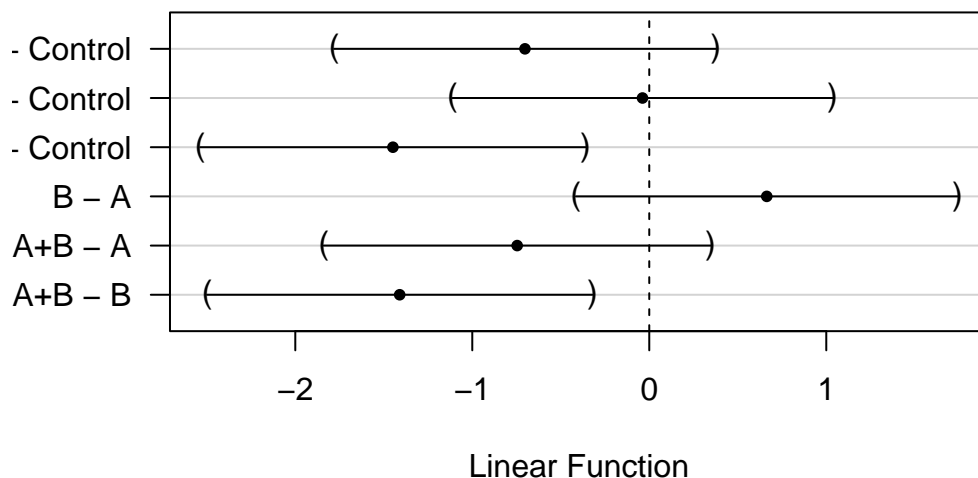
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)

```
plot(posthoc)
```



## 95% family-wise confidence level



```
# ---- Comparaciones con emmeans (Tukey) ----
if (!require("emmeans")) install.packages("emmeans")
```

Cargando paquete requerido: emmeans

```
library(emmeans)

emmeans_trat <- emmeans(modelo_glmnb, ~ Tratamiento)
emmeans_trat
```

Tratamiento	emmean	SE	df	asympt.LCL	asympt.UCL
Control	4.20	0.295	Inf	3.62	4.77
A	3.49	0.298	Inf	2.91	4.08
B	4.16	0.295	Inf	3.58	4.74
A+B	2.75	0.303	Inf	2.15	3.34

Results are given on the log (not the response) scale.  
Confidence level used: 0.95

```
comparaciones <- pairs(emmeans_trat, adjust = "tukey")
comparaciones
```

contrast	estimate	SE	df	z.ratio	p.value
Control - A	0.7022	0.419	Inf	1.674	0.3374
Control - B	0.0384	0.418	Inf	0.092	0.9997
Control - (A+B)	1.4484	0.423	Inf	3.420	0.0035
A - B	-0.6638	0.420	Inf	-1.582	0.3887
A - (A+B)	0.7462	0.425	Inf	1.755	0.2955
B - (A+B)	1.4100	0.424	Inf	3.329	0.0048

Results are given on the log (not the response) scale.

P value adjustment: tukey method for comparing a family of 4 estimates

```
if (!require("multcompView")) install.packages("multcompView")
```

Cargando paquete requerido: multcompView

Warning: package 'multcompView' was built under R version 4.4.3

```
library(multcompView)
cld(emmeans_trat, Letters = letters, adjust = "tukey") #Muestra las letras para las medias
```

Note: adjust = "tukey" was changed to "sidak"

because "tukey" is only appropriate for one set of pairwise comparisons

Tratamiento	emmean	SE	df	asympt.LCL	asympt.UCL	.group
A+B	2.75	0.303	Inf	1.99	3.50	a
A	3.49	0.298	Inf	2.75	4.24	ab
B	4.16	0.295	Inf	3.42	4.89	b
Control	4.20	0.295	Inf	3.46	4.93	b

Results are given on the log (not the response) scale.

Confidence level used: 0.95

Conf-level adjustment: sidak method for 4 estimates

P value adjustment: tukey method for comparing a family of 4 estimates

significance level used: alpha = 0.05

NOTE: If two or more means share the same grouping symbol,  
then we cannot show them to be different.

But we also did not show them to be the same.

```
# ---- Gráfico de boxplot con letras ----
if(!require("ggplot2")){install.packages("ggplot2")}
```

Cargando paquete requerido: ggplot2

Warning: package 'ggplot2' was built under R version 4.4.3

```
pos_letras <- data.frame(
  Tratamiento = factor(c("A+B", "A", "B", "Control"), levels = c("Control", "A", "B", "A+B")),
  ypos = c(280, 290, 300, 310), # Ajusta para evitar solapamientos
  Letters = c("a", "ab", "b", "b")
)

ggplot(insectos, aes(x = Tratamiento, y = Conteo, fill = Tratamiento)) +
  geom_boxplot(color = "black") +
  scale_y_continuous(limits = c(0, 320)) +
  scale_x_discrete(labels = c(
    "Control", "A", "B", "A+B")
  ) +
  scale_fill_manual(values = c(
    "#A6CEE3", "#1F78B4", "#B2DF8A", "#33A02C"
  )) +
  labs(
    title = "Boxplot CONTEO",
    x = expression("Tratamiento"),
    y = "Conteo (uds.)"
  ) +
  geom_text(data = pos_letras, aes(x = Tratamiento, y = ypos-120, label = Letters), size = 5) +
  theme_minimal() +
  theme(legend.position = "none")
```

