Ejemplo de un Diseño factorial

María Guzmán Martínez, mguzman@uagro.mx

27/11/2023

Índice


```r
library("ggplot2")    # ggplot
library("ggforce")    # geom_sina

library("tidyverse")  #%>%
library("dplyr")      # mutate, select,...
library("knitr")


library("nortest")    # Pruebas: ad.test, cvm.test, pearson.test, sf.test, lillie.test

library("agricolae")  # HSD.test, #skewness
```

1. Ejemplo

La avispa *Trichograma minutum* es un enemigo natural de varias plagas agrícolas. Por esta razón se desea conocer las condiciones optimas para producir esta avispa bajo condiciones de laboratorio. Se probó el efecto combinado de temperatura y las horas luz sobre la eclosión de huevecillos. Los niveles de temperatura que se probaron fueron $20°C$ y $25°C$ y los niveles de horas luz fueron 10 horas y 14 horas. Se utilizó un diseño experimental 2^2 con 4 repeticiones. La unidad experimental utilizada fue un conjunto de 100 huevecillos, y la variable respuesta fue el número de huevecillos que eclosionaron.

Las preguntas planteadas del experimento son las siguientes:

- ¿Existen diferencias entre las combinaciones de temperatura y horas luz?

- ¿Bajo qué condiciones de horas luz y temperatura se produce una mayor eclosión de huevecillos?

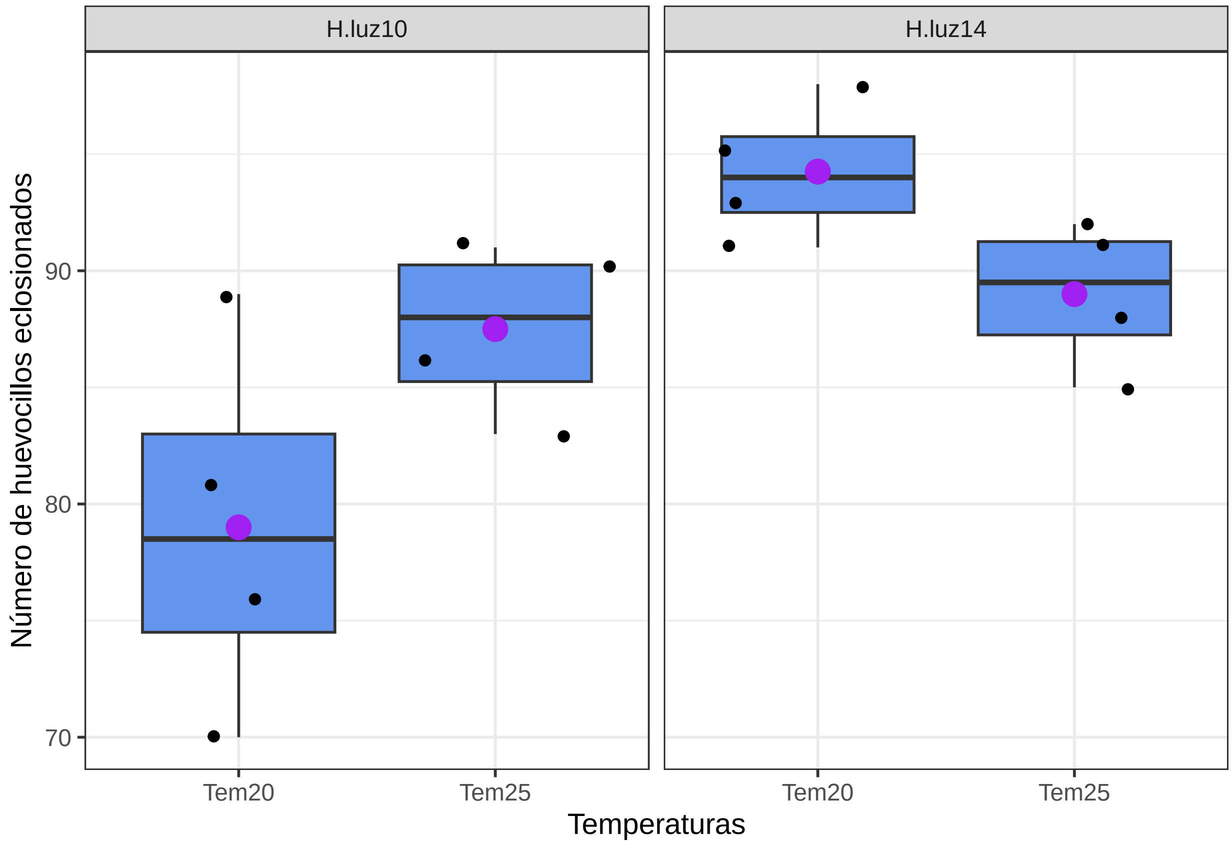
Los tratamientos a probar son:

- T_1: 10 hr con 20°C
- T_2: 10 hr con 25°C
- T_3: 14 hr con 20°C
- T_4: 14 hr con 25°C

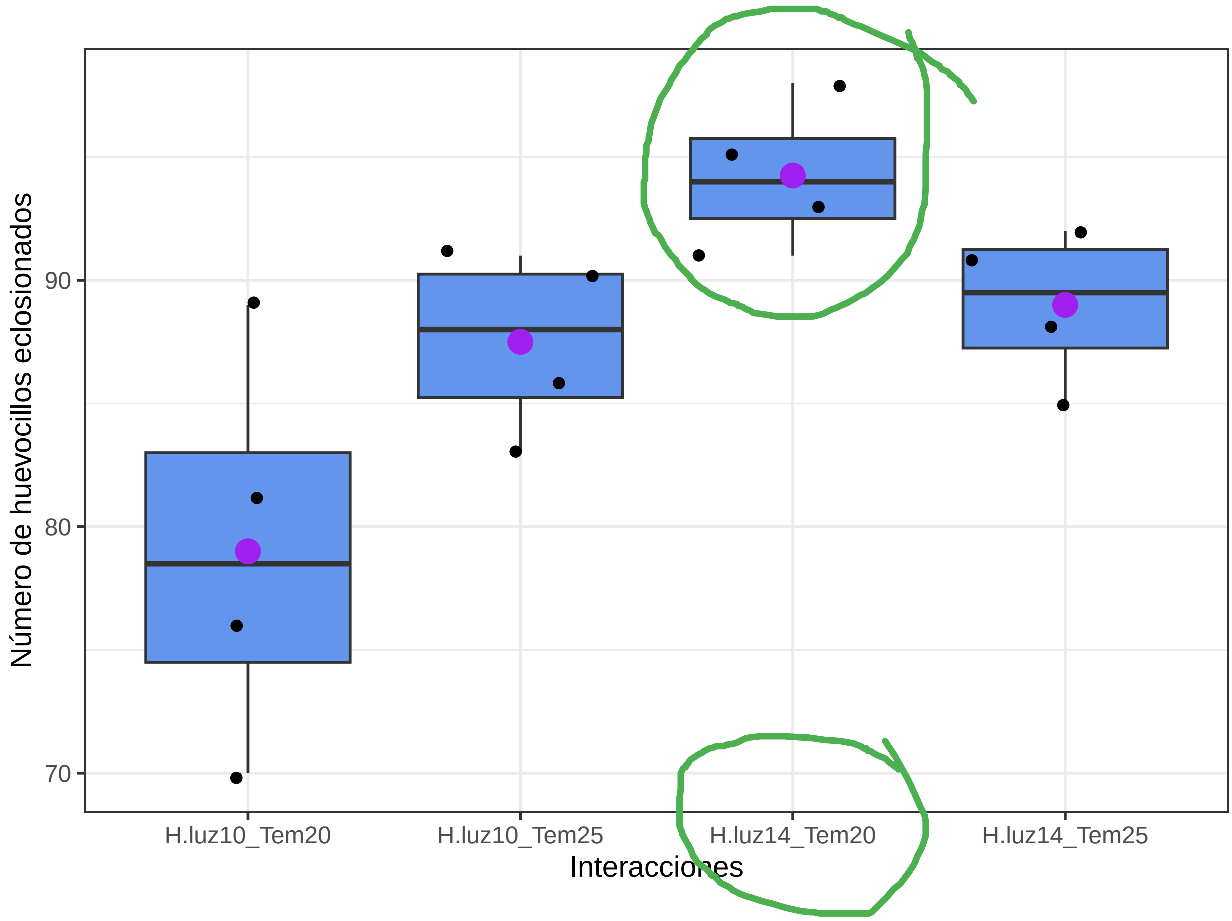
```
HE<-c(70,76,81,89,93,98,91,95,86,83,91,90,88,92,91,85)

H.luz<-c("H.luz10","H.luz10","H.luz10","H.luz10",
         "H.luz14","H.luz14","H.luz14","H.luz14",
         "H.luz10","H.luz10","H.luz10","H.luz10",
         "H.luz14","H.luz14","H.luz14","H.luz14")

Temperatura<-c("Tem20","Tem20","Tem20","Tem20","Tem20","Tem20","Tem20","Tem20",
               "Tem25","Tem25","Tem25","Tem25","Tem25","Tem25","Tem25","Tem25")

Interaccion<-paste(H.luz,Temperatura,sep = "_")

datos<-data.frame(H.luz,Temperatura,Interaccion,HE)

View(datos)
```

Gráficamente podemos observar que existen diferencias entre los tratamientos.

```
ggplot(datos, aes(x=Temperatura, y=HE)) +
  geom_boxplot(fill="cornflowerblue") + # Blox-plot
  geom_sina()+      # Grafica los puntos
  facet_wrap(~H.luz)+
  stat_summary(fun.y=mean, geom="point", shape=20, size=6, color="purple", fill="purple") + # Agrega el
  labs(x = "Temperaturas", y = "Número de huevocillos eclosionados")+
  theme_bw()
```

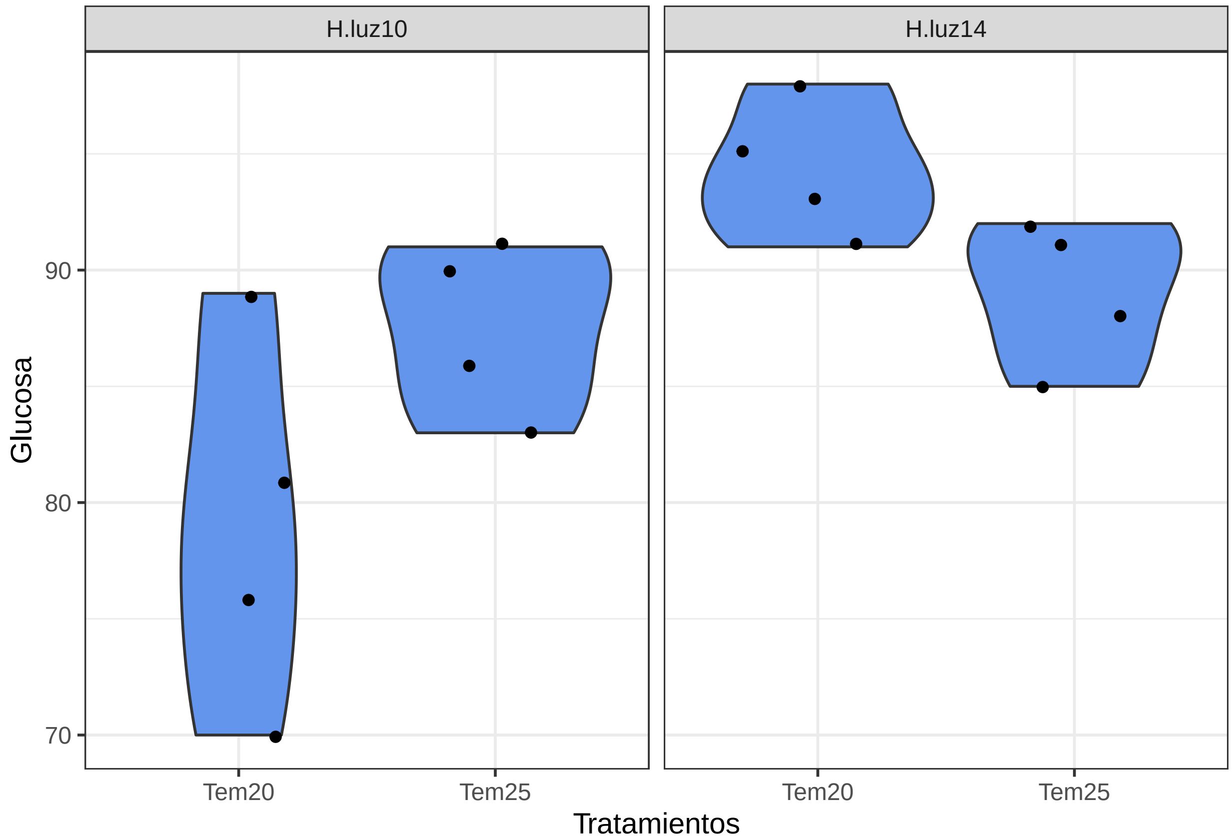



```r
ggplot(datos, aes(x=Interaccion, y=HE)) +
  geom_boxplot(fill="cornflowerblue") + # Blox-plot
  geom_sina()+       # Grafica los puntos
  stat_summary(fun.y=mean, geom="point", shape=20, size=6, color="purple", fill="purple") + # Agrega el
  labs(x = "Interacciones", y = "Número de huevocillos eclosionados")+
  theme_bw()
```

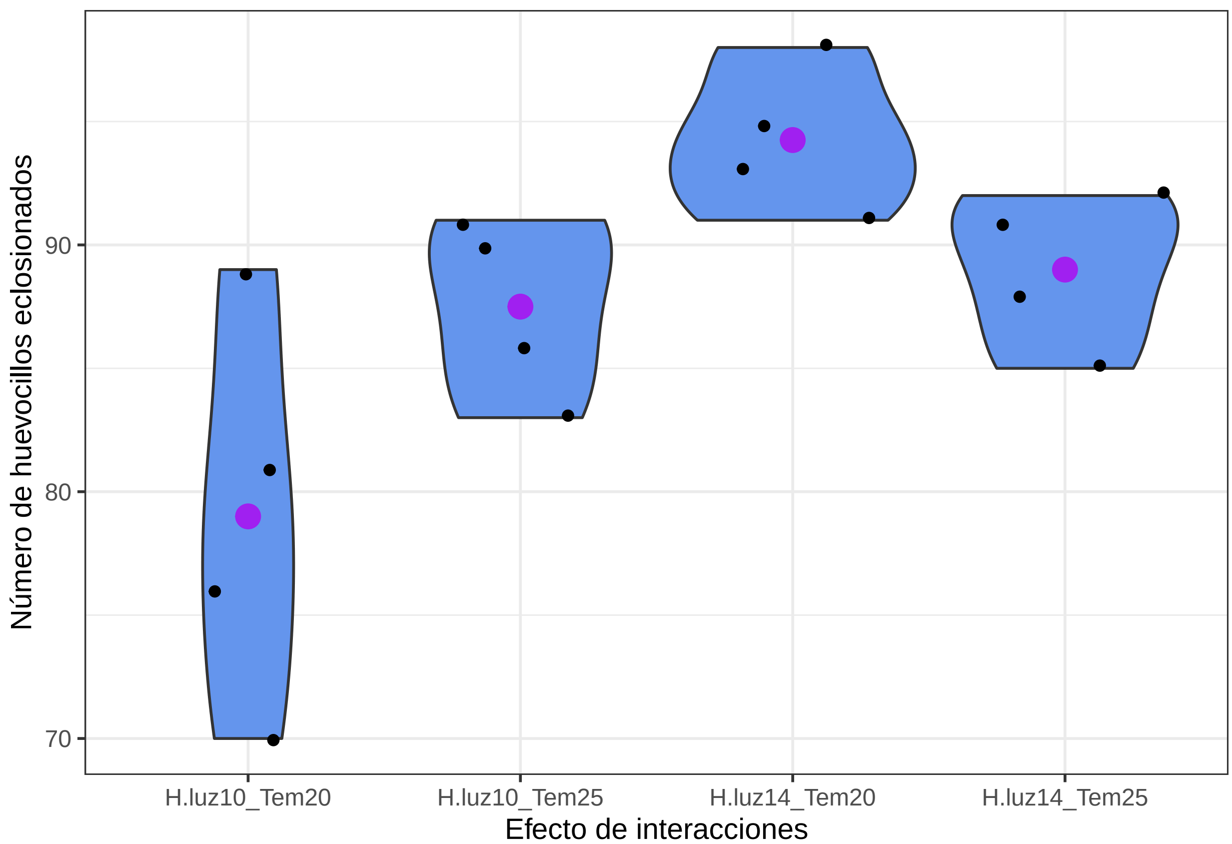


Veamos un gráfico de dispersión de los datos es el siguiente:

```
ggplot(datos, aes(x=Temperatura, y=HE)) +
  geom_violin(fill="cornflowerblue") +  # Muestra la forma de la dispersión
  geom_sina()+      # Coloca los puntos
  facet_wrap(~ H.luz)+  # Separa por sitio

  labs(x = "Tratamientos", y = "Glucosa")+
  theme_bw()
```




```r
ggplot(datos, aes(x=Interaccion, y=HE)) +
  geom_violin(fill="cornflowerblue") + # Blox-plot
  geom_sina()+       # Grafica los puntos
  stat_summary(fun.y=mean, geom="point", shape=20, size=6, color="purple", fill="purple") + # Agrega el
  labs(x = "Efecto de interacciones", y = "Número de huevocillos eclosionados")+
  theme_bw()
```



1.1. Anova

De los resultados, hay efecto de las horas luz en la eclosión de los huevos (p-valor=0.0053), también hay efecto de la interacción de los efectos (p-valor=0.0164), solo en la temperatura no existe efecto sobre la eclosión de los huevos.

Con un nivel de significancia del $\alpha = 0.05$, las hipótesis que se están probando son las siguientes:

Contraste de hipótesis del efecto principal del factor A (horas luz)

$$\boxed{\begin{array}{c} H_0^A : \alpha_1 = \alpha_2 = 0 \\ \text{vs} \\ H_1^A : \text{algún } \alpha_i \neq 0 \end{array}}$$

En este caso se rechaza H_0, ya que $p - valor = 0.0053 < \alpha$, es decir, existen diferencias entre los dos niveles de las horas luz en la eclosión de los huevecillos de *Thichogramma minutum*.

Contraste de hipótesis del efecto principal del factor B (Temperatura)

$$\boxed{\begin{array}{c} H_0^B : \beta_1 = \beta_2 = 0 \\ \text{vs} \\ H_1^B : \text{algún } \beta_j \neq 0 \end{array}}$$

En este caso no se rechaza H_0, ya que $p - valor = 0.5223 > 0.05$

Contraste de hipótesis de la interacción BA

$$H_0^{AB}: (\alpha\beta)_{11} = (\alpha\beta)_{12} = (\alpha\beta)_{21} = (\alpha\beta)_{22} = 0$$

vs

$$H_1^{AB}: \text{algún } (\alpha\beta)_{ij} \neq 0$$

con $i = 1, 2$ y $j = 1, 2$

0.05

En este caso se rechaza H_0, ya que $p - valor = 0.0164 <$ ~~0.5~~ Es decir, existe efecto de las interacciónes en la variable respuesta, número de huevecillos eclosionados.

```
mod_aov<-aov(HE~H.luz+Temperatura+Interaccion, datos)
summary(mod_aov)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
H.luz        1 280.56  280.56  11.540 0.0053 **
Temperatura  1  10.56   10.56   0.434 0.5223
Interaccion  1 189.06  189.06   7.776 0.0164 *
Residuals   12 291.75   24.31
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observación:

Significado	p-value
***	$[0, 0.001]$
**	$(0.001, 0.01]$
*	$(0.01, 0.05]$
.	$(0.05, 0.1]$
	$(0.1, 1]$

La función `anova()` muestra la tabla de la suma de cuadrados

```
res_anova<-anova(mod_aov)
res_anova
```

```
Analysis of Variance Table

Response: HE
            Df  Sum Sq Mean Sq F value   Pr(>F)
H.luz        1 280.562 280.562 11.5398 0.005299 **
Temperatura  1  10.562  10.562  0.4344 0.522262
Interaccion  1 189.062 189.062  7.7763 0.016387 *
Residuals   12 291.750  24.312
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
kable(res_anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
H.luz	1	280.5625	280.5625	11.5398458	0.0052987
Temperatura	1	10.5625	10.5625	0.4344473	0.5222624
Interaccion	1	189.0625	189.0625	7.7763496	0.0163867
Residuals	12	291.7500	24.3125	–	–

1.2. Supuestos

Juego de hipótesis:

H_0: Los errores tienen distribución Normalidad.

H_A: Los errores no tienen distribución Normalidad.

Se rechaza H_0 si $p - valor < \alpha$

Pruebas para verificar el supuesto de normalidad en los errores, e_{ij}.

```
(prueba_shapiro<-shapiro.test(mod_aov$residuals)) # Shapiro-Wilk
```

```
    Shapiro-Wilk normality test

data:  mod_aov$residuals
W = 0.96719, p-value = 0.7913
```

```
(prueba_ad<-ad.test(mod_aov$residuals))      # Anderson-Darling
```

```
    Anderson-Darling normality test

data:  mod_aov$residuals
A = 0.28318, p-value = 0.5856
```

```
(prueba_cvm<-cvm.test(mod_aov$residuals))      # Cramer-von Mises
```

```
    Cramer-von Mises normality test

data:  mod_aov$residuals
W = 0.038698, p-value = 0.6837
```

```
(prueba_lillie<-lillie.test(mod_aov$residuals)) # Kolmogorov-Smirnov
```

```
    Lilliefors (Kolmogorov-Smirnov) normality test

data:  mod_aov$residuals
D = 0.13508, p-value = 0.6067
```

```
(prueba_pearson<-pearson.test(mod_aov$residuals)) # Pearson chi-square
```

```
    Pearson chi-square normality test

  data:  mod_aov$residuals
  P = 5, p-value = 0.2873
```

```
(prueba_sf<-sf.test(mod_aov$residuals))      # Shapiro-Francia
```

```
    Shapiro-Francia normality test

  data:  mod_aov$residuals
  W = 0.95174, p-value = 0.4398
```

Resumen:

Prueba estadística	Valor del estadístico de prueba	p-valor
Shapiro-Wilk	0.967	0.791
Anderson-Darling	0.283	0.586
Cramer-von Mises	0.039	0.684
Kolmogorov-Smirnov	0.135	0.607
Pearson chi-square	5	0.287
Shapiro-Francia	0.952	0.44

De acuerdo con los resultados anteriores, se cumple el supuesto de normalidad en los errores.

Para distribuciones simétricas con tamaños de muestra pequeños, los investigadores pueden elegir alguna de las siguientes pruebas de normalidad (Yazici & Yolacan, 2007):

- Kolmogorov-Smirnov,
- Kolmogorov-Smirnov modificada
- Anderson-Darling .

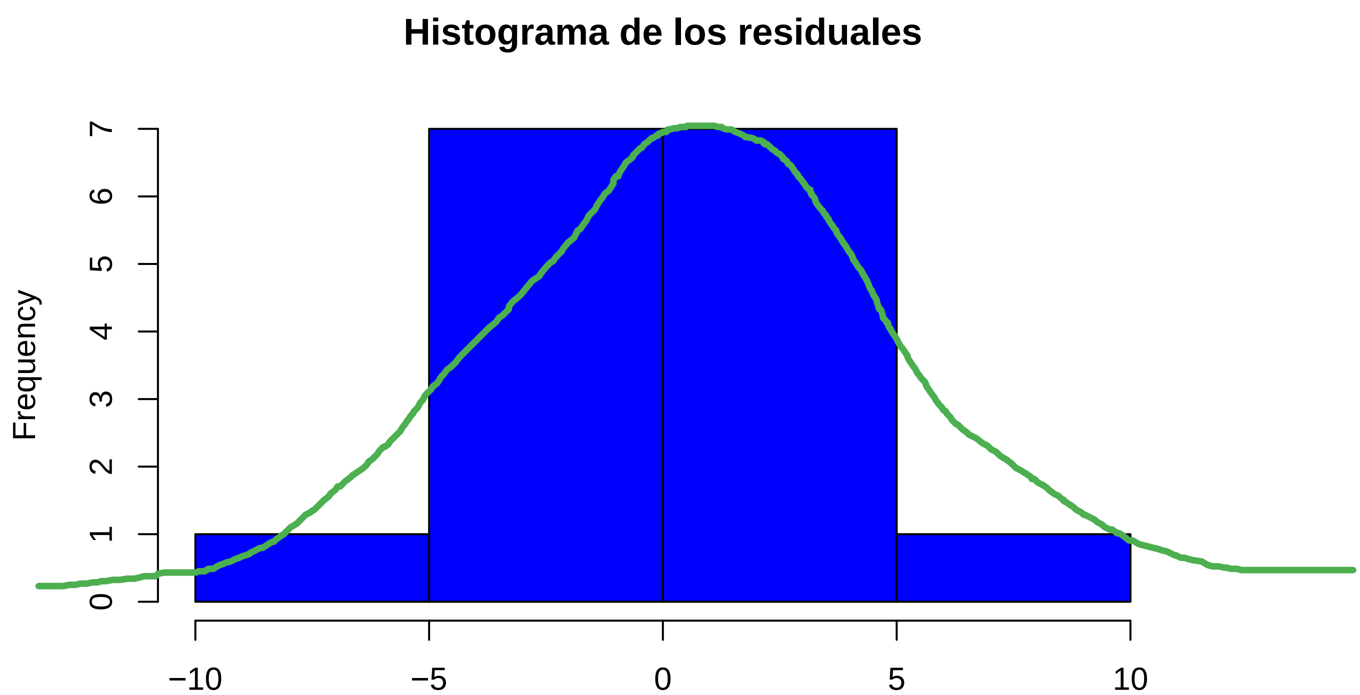
La función **skewness**, mide la falta de simetría en un conjunto de datos.

```
skewness(mod_aov$residuals) # mide la falta de simetría
```

```
  [1] 0.1801979
```

Gráficamente, se puede observar que se cumple el supuesto de normalidad.

```
hist(mod_aov$residuals, main = "Histograma de los residuales", xlab = " ", col="blue")
```



Juego de hipótesis:

H_0: Los errores tienen varianza constante

H_A: Los errores no tienen varianza constante.

```
bartlett.test(mod_aov$residuals, datos$H.luz)    # Prueba CONFIABLE SI HAY NORMALIDAD
```

```
	Bartlett test of homogeneity of variances

data:  mod_aov$residuals and datos$H.luz
Bartlett's K-squared = 3.0528, df = 1, p-value = 0.0806
```

```
bartlett.test(mod_aov$residuals, datos$Temperatura)    # Prueba CONFIABLE SI HAY NORMALIDAD
```

```
	Bartlett test of homogeneity of variances

data:  mod_aov$residuals and datos$Temperatura
Bartlett's K-squared = 1.9978, df = 1, p-value = 0.1575
```

De acuerdo con la prueba de Bartlett, se cumple el supuesto de normalidad en los errores, ya que se tiene un $p-valor = 0.0806$ en los niveles de las horas luz, y un $p-valor = 0.1575$ para los niveles de las temperaturas.

Prueba de aleatoriedad:

H_0: Los datos se produjeron de forma aleatoria.

H_a: Los datos no se produjeron de forma aleatoria.

Se rechaza H_0 si el $p - valor < \alpha$

```
library(remotes)
#install_github("debinqiu/snpar")

library(snpar)

runs.test(mod_aov$residuals)  #  prueba de rachas
```

1.3. Comparaciones múltiples

1.3.1. Factor A

```
comp_tukey_Fac.A<-HSD.test(mod_aov, "H.luz", group=TRUE)
comp_tukey_Fac.A
```

```
$statistics
  MSerror Df    Mean       CV      MSD
  24.3125 12 87.4375 5.639195 5.371613

$parameters
   test name.t ntr StudentizedRange alpha
  Tukey  H.luz   2         3.081307  0.05

$means
            HE      std r        se Min Max   Q25  Q50   Q75
H.luz10 83.250 7.363035 8 1.743291  70  91 79.75 84.5 89.25
H.luz14 91.625 3.997767 8 1.743291  85  98 90.25 91.5 93.50

$comparison
NULL

$groups
            HE groups
H.luz14 91.625      a
H.luz10 83.250      b

attr(,"class")
[1] "group"
```

```
knitr::kable(comp_tukey_Fac.A$means,
             digits = 3,
             caption = "Estadísticas")
```

Cuadro 4: Estadísticas

	HE	std	r	se	Min	Max	Q25	Q50	Q75
H.luz10	83.250	7.363	8	1.743	70	91	79.75	84.5	89.25
H.luz14	91.625	3.998	8	1.743	85	98	90.25	91.5	93.50

```
kable(comp_tukey_Fac.A$groups,
      digits = 3,
      caption = "Comparaciones mútiples")
```

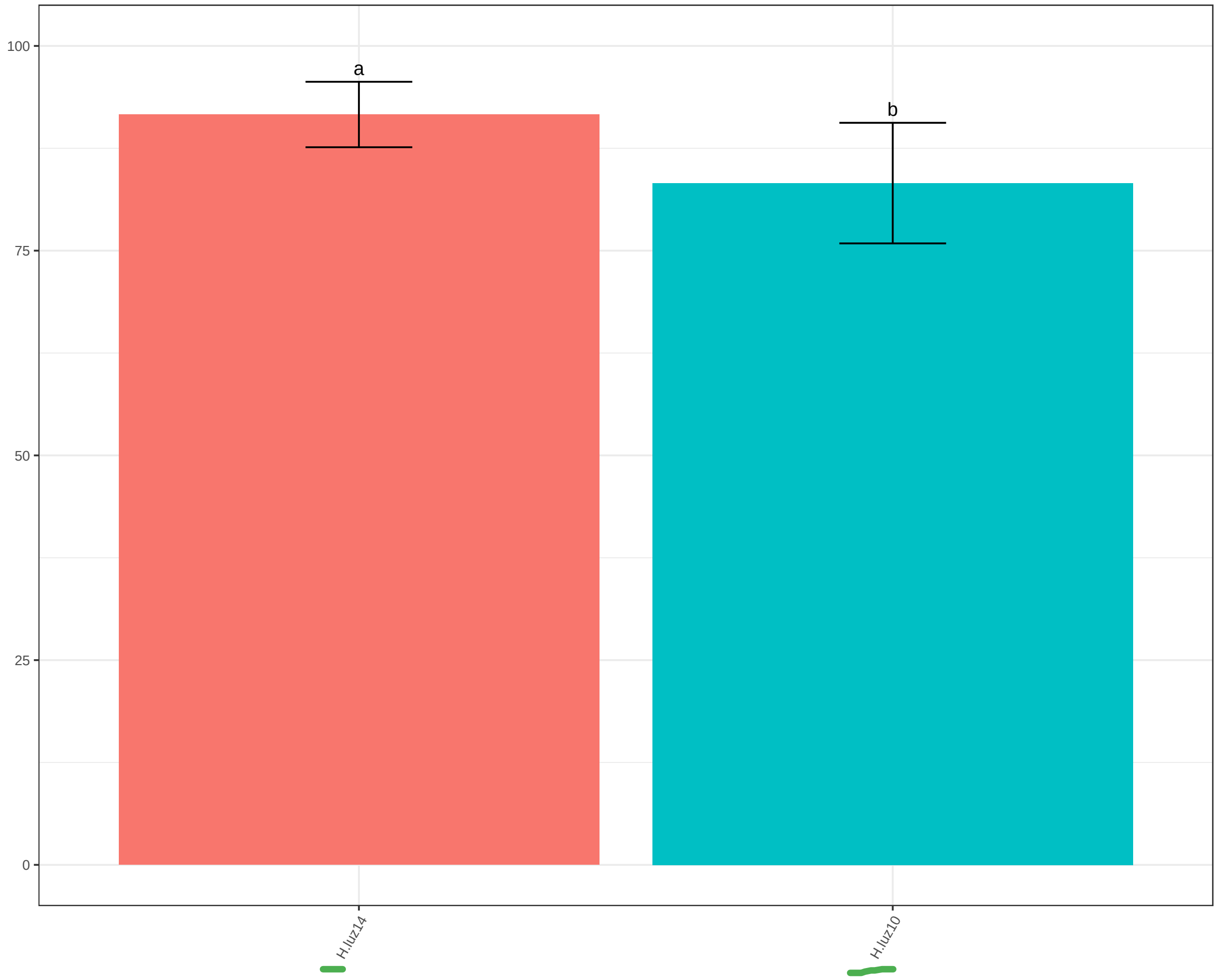
Cuadro 5: Comparaciones mútiples

	HE	groups
H.luz14	91.625	a
H.luz10	83.250	b

```
#--- Código para ordenar la aparición de los niveles en el gráfico
orden_trat<-c("H.luz14","H.luz10")

res_compa.dosis<-data.frame( Trat=rownames(comp_tukey_Fac.A$groups),
                             Promedio=comp_tukey_Fac.A$groups$HE,
                             Grupo=comp_tukey_Fac.A$groups$groups,
                             E.Estandar=sort(comp_tukey_Fac.A$means$std)
                           )

res_compa.dosis<-res_compa.dosis %>%
                 mutate(Trat=factor(Trat,levels =orden_trat))

res_compa.dosis
```

```
      Trat Promedio Grupo E.Estandar
  1 H.luz14   91.625     a   3.997767
  2 H.luz10   83.250     b   7.363035
```

```
ggplot(res_compa.dosis, aes(x = Trat, y = Promedio)) +
  geom_bar(stat = "identity", aes(fill = Trat), show.legend = FALSE) +
  geom_errorbar(aes(ymin = Promedio - E.Estandar, ymax = Promedio + E.Estandar), width = 0.2) +
  labs(x = " ", y = "") +
  geom_text(aes(label = Grupo, y = Promedio + E.Estandar), vjust = -0.5) +
  ylim(0,100)+
  theme_gray()+
  theme_bw()+
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle =60,h=1))
```



1.4. Interacción AB

```
comp_tukey_Fac.AB<-HSD.test(mod_aov, "Interaccion", group=TRUE)
comp_tukey_Fac.AB
```

```
$statistics
  MSerror Df    Mean       CV      MSD
  24.3125 12 87.4375 5.639195 10.35132

$parameters
   test      name.t ntr StudentizedRange alpha
  Tukey Interaccion   4          4.19866  0.05

$means
                 HE      std r       se Min Max   Q25  Q50   Q75
H.luz10_Tem20 79.00 8.041559 4 2.465385  70  89 74.50 78.5 83.00
H.luz10_Tem25 87.50 3.696846 4 2.465385  83  91 85.25 88.0 90.25
H.luz14_Tem20 94.25 2.986079 4 2.465385  91  98 92.50 94.0 95.75
H.luz14_Tem25 89.00 3.162278 4 2.465385  85  92 87.25 89.5 91.25
```

```
$comparison
NULL

$groups
                   HE groups
H.luz14_Tem20 94.25      a
H.luz14_Tem25 89.00     ab
H.luz10_Tem25 87.50     ab
H.luz10_Tem20 79.00      b

attr(,"class")
[1] "group"
```

```
kable(comp_tukey_Fac.AB$means, digits = 3,caption = "Estadísticas")
```

Cuadro 6: Estadísticas

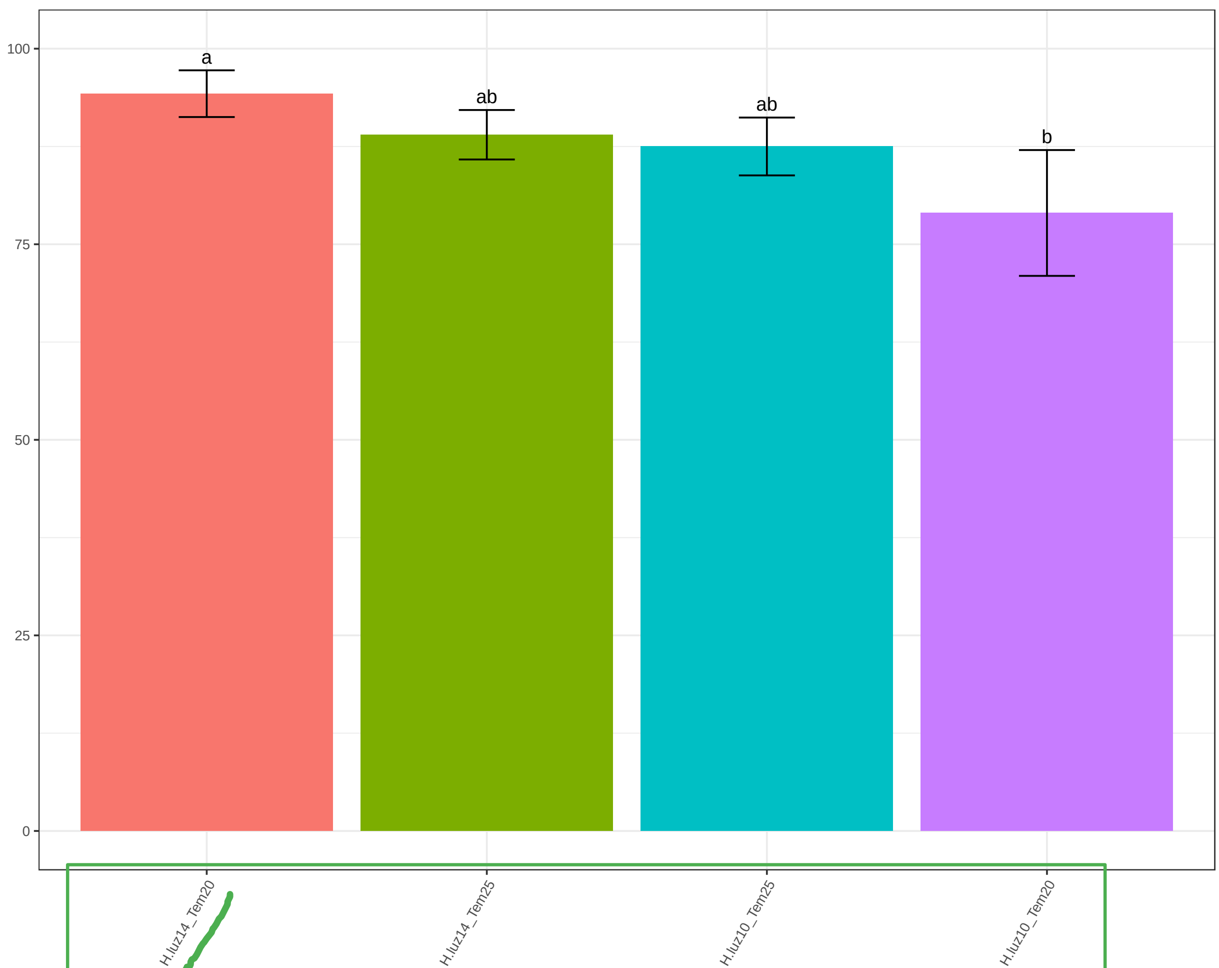
	HE	std	r	se	Min	Max	Q25	Q50	Q75
H.luz10_Tem20	79.00	8.042	4	2.465	70	89	74.50	78.5	83.00
H.luz10_Tem25	87.50	3.697	4	2.465	83	91	85.25	88.0	90.25
H.luz14_Tem20	94.25	2.986	4	2.465	91	98	92.50	94.0	95.75
H.luz14_Tem25	89.00	3.162	4	2.465	85	92	87.25	89.5	91.25

```
#--- Código para ordenar la aparición de los niveles en el gráfico
orden_trat<-c("H.luz14_Tem20","H.luz14_Tem25","H.luz10_Tem25","H.luz10_Tem20")


res_compa.dosis<-data.frame( Trat=rownames(comp_tukey_Fac.AB$groups),
                             Promedio=comp_tukey_Fac.AB$groups$HE,
                             Grupo=comp_tukey_Fac.AB$groups$groups,
                             E.Estandar=sort(comp_tukey_Fac.AB$means$std)
                          )

res_compa.dosis<-res_compa.dosis %>%
                 mutate(Trat=factor(Trat,levels =orden_trat))

res_compa.dosis
```

```
           Trat Promedio Grupo E.Estandar
1 H.luz14_Tem20    94.25     a   2.986079
2 H.luz14_Tem25    89.00    ab   3.162278
3 H.luz10_Tem25    87.50    ab   3.696846
4 H.luz10_Tem20    79.00     b   8.041559
```

```
ggplot(res_compa.dosis, aes(x = Trat, y = Promedio)) +
  geom_bar(stat = "identity", aes(fill = Trat), show.legend = FALSE) +
  geom_errorbar(aes(ymin = Promedio - E.Estandar, ymax = Promedio + E.Estandar), width = 0.2) +
  labs(x = " ", y = "") +
  geom_text(aes(label = Grupo, y = Promedio + E.Estandar), vjust = -0.5) +
  ylim(0,100)+
```

```r
  theme_gray()+
  theme_bw()+
  theme(text = element_text(size = 10),
        axis.text.x = element_text(angle =60,h=1))
```




```r
kable(comp_tukey_Fac.AB$groups, digits = 3,caption = "Comparaciones mútiples")
```

Cuadro 7: Comparaciones mútiples

	HE	groups
H.luz14_Tem20	94.25	a
H.luz14_Tem25	89.00	ab
H.luz10_Tem25	87.50	ab
H.luz10_Tem20	79.00	b

Por lo tanto, las mejores combinaciones de horas luz y temperatura para una buena eclosión de huevecillos de *Trichogramma minutum* son 14 horas luz y 20°c, 14 horas luz y 25° C y 10 horas luz y 25° C.

Bibliografía

Yazici, B., & Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, *77*(2), 175–183.