Pruebas estadísticas con rstatix :: CHEAT SHEET

rstatix ejecuta las pruebas habituales de R base con una gramática compatible con pipes y devuelve data frames ordenados, listos para reportar y graficar.

1 Elige la prueba

El **diseño** elige la fila; los **supuestos** (sección 3), la columna. En gris: tamaño del efecto u opción clave.

	SE CUMPLEN	NO CUMPLEN / n BAJO
2 grupos independientes	`t_test()`	`wilcox_test()`
	cohens_d()	wilcox_effsize()
2 grupos pareados + paired = TRUE, id =	`t_test()`	`wilcox_test()`
	cohens_d()	wilcox_effsize()
3+ grupos independientes	`anova_test()`	`kruskal_test()`
	columna ges	kruskal_effsize()
pares post hoc	`tukey_hsd()`	`dunn_test()`
3+ repetidas + wid = , within =	`anova_test()`	`friedman_test()`
	columna ges	friedman_effsize()
2 variables numéricas	`cor_test()`	`cor_test()`
	method = "pearson"	*method = "spearman"*
2 variables categóricas	`chisq_test()`	`fisher_test()`
	cramer_v()	*frecuencias esperadas < 5*
categóricas pareadas	`mcnemar_test()`	

CONSEJO `t_test()` es una prueba de **Welch** por defecto (var.equal = FALSE). Con 3+ grupos y varianzas distintas usa `welch_anova_test()` + `games_howell_test()`.

PRUEBAS MÚLTIPLES p.adjust.method =

`"holm"` por defecto en pruebas por pares

`"bonferroni"` la más simple y conservadora

`"BH"` tasa de falsos descubrimientos

`"none"` solo para una prueba planificada

2 Una sola gramática

```
library(rstatix)
df <- ToothGrowth   # len, supp, dose
df$dose <- factor(df$dose)
```

```
df %>% t_test(len ~ supp)
```

- `df` → los datos van primero → pipes
- `t_test` → función de prueba
- `len` → desenlace
- `supp` → grupo

.y.	group1	group2	n1	n2	statistic	df	p
len	OJ	VC	30	30	1.92	55.3	**0.061**

Un tibble, no un texto impreso: fíltralo, únelo, expórtalo o dale formato como a cualquier tabla.

POR SUBGRUPO → AJUSTA → ASTERISCOS

```
df %>% group_by(dose) %>%
  t_test(len ~ supp) %>%
  adjust_pvalue(method = "bonferroni") %>%
  add_significance()
```

dose	p	p.adj	p.adj.signif
0.5	0.006	0.019	*
1	0.001	0.003	**
2	0.964	1.000	ns

Asteriscos: **** < 0.0001 · *** < 0.001 · ** < 0.01 · * < 0.05 · ns

¿QUÉ COMPARACIONES?

`len ~ dose` todos los pares (ajuste de Holm)

`ref.group = "0.5"` cada nivel frente a una referencia

`comparisons = list(c("0.5", "2"))` solo los pares planificados

`effect.size = TRUE` añade la d de Cohen: 0.49 ("small")

`detailed = TRUE` añade estimación, IC 95%, método: OJ − VC = 3.7 (−0.17 a 7.57)

REPORTA EN TEXTO

```
res <- df %>% t_test(len ~ supp)
get_test_label(res, detailed = TRUE,
               type = "text")
```

→ **T test, t(55.31) = 1.92, p = 0.061, n = 60**

El type por defecto devuelve un subtítulo para gráficos.

3 Revisa los supuestos

NORMALIDAD (SHAPIRO–WILK)

```
df %>% group_by(supp, dose) %>%
  shapiro_test(len)
```

6 celdas, p mínimo = 0.17 → **no se rechaza la normalidad.**

Evalúa los grupos que comparas. OJ sin separar por dosis da p = 0.024: mezcla tres dosis.

```
fit <- lm(len ~ supp * dose, data = df)
shapiro_test(residuals(fit))  # p = 0.67
```

En diseños de ANOVA, revisa los residuos del modelo.

```
ggpubr::ggqqplot(df, "len",
  facet.by = "dose")   # línea = normal
```

VARIANZAS IGUALES (LEVENE)

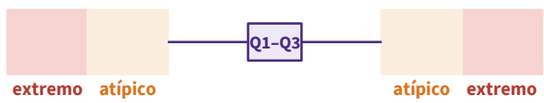
```
df %>% levene_test(len ~ supp * dose)
```

F(5, 54) = 1.71, p = 0.15 → **sin evidencia de varianzas distintas.**

VALORES ATÍPICOS

```
df %>% group_by(dose) %>%
  identify_outliers(len)
```

dose	len	supp	is.outlier	is.extreme
0.5	21.5	OJ	TRUE	FALSE



Atípico: a más de 1.5 × RIC de Q1 o Q3. Extremo: a más de 3 × RIC. Investiga antes de eliminar.

CONSEJO Shapiro–Wilk detecta desvíos triviales en muestras **grandes** y pasa por alto desvíos importantes en muestras **pequeñas**: léelo junto al gráfico QQ.

DATOS PAREADOS: EVALÚA LAS DIFERENCIAS

```
d <- with(sleep, extra[group == 2] -
          extra[group == 1])
shapiro_test(d)   # p = 0.033
```

→ No normal: usa el Wilcoxon pareado (p = 0.009).
Varios desenlaces a la vez: `mshapiro_test()`, `box_m()`

4 Describe y grafica

ESTADÍSTICOS DESCRIPTIVOS

```
df %>% group_by(supp) %>%
  get_summary_stats(len, type = "mean_sd")
```

supp	variable	n	mean	sd
OJ	len	30	20.66	6.61
VC	len	30	16.96	8.27

`type =`

`"full"`	`"common"`	`"robust"`
`"five_number"`	`"mean_sd"`	`"mean_se"`
`"mean_ci"`	`"median_iqr"`	`"median_mad"`
`"quantile"`		

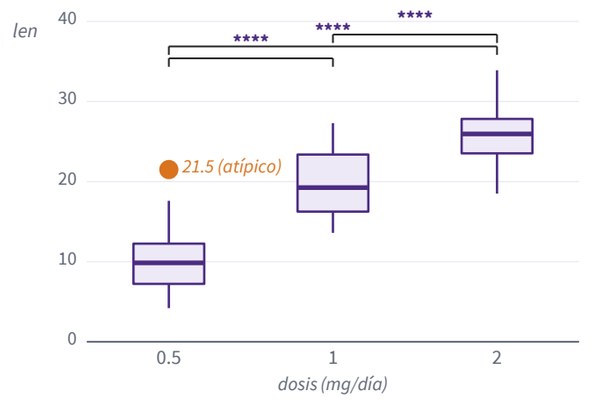
O elige columnas: `show = c("mean", "sd", "median", "iqr")`

FRECUENCIAS

```
df %>% freq_table(supp, dose)   # n, prop
```

DE LA PRUEBA AL GRÁFICO

```
st <- df %>% t_test(len ~ dose) %>%
  add_xy_position(x = "dose")
ggpubr::ggboxplot(df, "dose", "len") +
  ggpubr::stat_pvalue_manual(st,
    label = "p.adj.signif")
```



OJO Las pruebas pareadas emparejan las filas por su **orden**. Añade `id = "ID"` para emparejar por sujeto. Con las filas de sleep desordenadas una vez: p = 0.12 sin id, p = 0.003 con id.

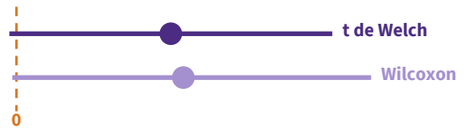
Pruebas estadísticas con rstatix :: **CHEAT SHEET**

Compara grupos, identifica cuáles difieren, mide el efecto y pasa a diseños factoriales, tablas de frecuencias y correlaciones.

5 Dos grupos

```
df %>% t_test(len ~ supp, detailed = TRUE)
df %>% wilcox_test(len ~ supp,
  detailed = TRUE)
```

prueba	OJ − VC	IC 95%	p
t de Welch	3.7	−0.17 a 7.57	0.061
Wilcoxon	4.0	−0.10 a 8.50	0.064



Ambos intervalos incluyen el 0: **sin diferencia clara**, pero los datos son compatibles con una diferencia considerable. Reporta el IC, no solo el valor p.

Estimación de Wilcoxon = desplazamiento de Hodges–Lehmann, no diferencia de medianas.

TAMAÑO DEL EFECTO

```
df %>% cohens_d(len ~ supp)
df %>% wilcox_effsize(len ~ supp)
```

→ **d = 0.49 · r = 0.24, ambos pequeños**



Etiquetas de Cohen (1992), en la columna magnitude. O usa effect.size = TRUE en la prueba.

PAREADOS (sleep: 10 personas, dos fármacos)

```
sleep %>% t_test(extra ~ group,
  paired = TRUE, id = "ID", detailed = TRUE)
sleep %>% cohens_d(extra ~ group,
  paired = TRUE, id = "ID")
```

Diferencia de medias −1.58 (−2.46 a −0.70), t(9) = −4.06, p = 0.003; **d = −1.28, grande.**

CONSEJO `alternative = "greater"` da una prueba unilateral (OJ > VC: p = 0.030). Elige la dirección **antes** de ver los datos, nunca para rescatar un valor p.

6 Tres o más grupos

GLOBAL ¿Algún grupo difiere? → **POST HOC** ¿Qué pares difieren?

ANOVA DE UNA VÍA

```
df %>% anova_test(len ~ dose)
```

Efecto	DFn	DFd	F	p	ges
dose	2	57	67.4	< 0.001	**0.70**

ges = eta² generalizado, la proporción de varianza debida a la dosis. `effect.size = "pes"` da eta² parcial.

```
df %>% tukey_hsd(len ~ dose)
```

par	diferencia	IC 95%	p.adj
1 − 0.5	9.1	5.9 a 12.4	< 0.0001
2 − 0.5	15.5	12.3 a 18.7	< 0.0001
2 − 1	6.4	3.1 a 9.6	< 0.0001

Cada paso 0.5 → 1 → 2 mg/día aumenta la longitud.

VARIANZAS DISTINTAS

```
df %>% welch_anova_test(len ~ dose)
df %>% games_howell_test(len ~ dose)
```

Welch: F(2, 37.7) = 68.4; luego pares de Games–Howell.

BASADAS EN RANGOS

```
df %>% kruskal_test(len ~ dose)
df %>% kruskal_effsize(len ~ dose)
df %>% dunn_test(len ~ dose)   # Holm
```

H(2) = 40.7, p < 0.001; eta²[H] = 0.68, grande.

par	0.5 vs 1	0.5 vs 2	1 vs 2
p.adj	0.0008	< 0.0001	0.005
signif	***	****	**

CONSEJO `pairwise_t_test(pool.sd = TRUE)` combina la DE de todos los grupos, como el ANOVA.

7 Factoriales y repetidas

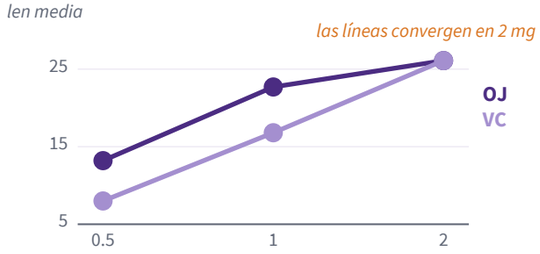
ANOVA DE DOS VÍAS

```
df %>% anova_test(len ~ supp * dose)
```

Efecto	DFn	DFd	F	p	ges
supp	1	54	15.6	< 0.001	0.22
dose	2	54	92.0	< 0.001	0.77
supp:dose	2	54	4.1	**0.022**	0.13

Sumas de cuadrados tipo II por defecto; type = 3 para tipo III.

len media



¿INTERACCIÓN? EVALÚA EFECTOS SIMPLES

```
df %>% group_by(dose) %>%
  emmeans_test(len ~ supp,
    p.adjust.method = "bonferroni")
```

OJ > VC con 0.5 mg (p = 0.002) y 1 mg (p < 0.001), **no con 2 mg** (p = 0.96).

MEDIDAS REPETIDAS

```
res <- sleep %>% anova_test(dv = extra,
  wid = ID, within = group)
get_anova_table(res)   # F(1, 9) = 16.5
```

Esfericidad: Mauchly y la corrección de Greenhouse–Geisser se aplican **automáticamente** si hace falta (3+ niveles). Diseño mixto: añade `between =`. Por rangos:

```
sleep %>% friedman_test(extra ~ group | ID)
```

MUCHOS DESENLACES A LA VEZ

```
iris %>% tidyr::pivot_longer(-Species,
    names_to = "measure") %>%
  group_by(measure) %>%
  kruskal_test(value ~ Species) %>%
  adjust_pvalue(method = "holm")
```

No llames "variable" a la columna: `get_summary_stats()` falla.

8 Tablas y correlación

```
bw <- MASS::birthwt
xt <- xtabs(~ smoke + low, data = bw)
```

bajo peso al nacer

	no	sí	% bajo
no fumadora	86	29	**25.2%**
fumadora	44	30	**40.5%**

```
chisq_test(xt)
```

→ **$\chi^2(1) = 4.24$, p = 0.040 (con corrección de Yates)**

```
expected_freq(chisq_test(xt))
```

→ **todas ≥ 23: χ^2 es válido**

```
fisher_test(xt, detailed = TRUE)
```

→ **OR 2.01 (1.03–3.96), p = 0.036**

```
cramer_v(xt)
```

→ **0.16**

¿Qué celdas lo explican? `chisq_descriptives()` da los residuos estandarizados (aquí ±2.22; más allá de ±2 destaca).

Fisher si las frecuencias esperadas son pequeñas (< 5); mcnemar_test() para sí/no pareado.

CORRELACIÓN

```
mtcars %>% cor_test(mpg, wt)
```

→ **r = −0.87 (−0.93 a −0.74), p < 0.001** · Spearman ρ = −0.89

```
mtcars %>% select(mpg, wt, hp, qsec) %>%
  cor_mat() %>% cor_mark_significant()
```

	mpg	wt	hp
wt	−0.87 ****		
hp	−0.78 ****	0.66 ****	
qsec	0.42 *	−0.17	−0.71 ****

Violeta: negativa · naranja: positiva. Valores p: cor_get_pval().

CHECKLIST Reporta la prueba, estadístico, gl, valor p exacto, **tamaño del efecto con su IC** y n. Un valor p por sí solo no indica el tamaño.