

Tablas reproducibles, presentables y con formato numérico local con {gtsummary}

Eva Retamal Riquelme.

Unidad de Neurología Adulto, Hospital Clínico La Florida. Santiago, Chile.

Abstract

El uso de tablas para resumir información es muy útil, especialmente para analizar valores uno a uno, y al comparar varias categorías. Con el paso del tiempo, frecuentemente es necesario generar nuevamente estas tablas. Hacer esto manualmente es un proceso tedioso y propenso a errores.

Por otra parte, la forma de escribir los números es distinta en diferentes idiomas. Si el software empleado utiliza un formato numérico diferente al idioma en que será presentado, es necesario cambiar este formato una vez concluido el análisis.

El paquete {gtsummary} permite elaborar tablas reproducibles que resumen conjuntos de datos. Estas tablas son altamente personalizables, permitiendo elegir el formato numérico del resultado. Esto disminuye la posibilidad de errores, y facilita la reproducibilidad de este tipo de análisis.

Usando el conjunto de datos *pinguinos*, del paquete {datos}, se ejemplifica cómo realizar tablas personalizadas, reproducibles, que pueden ser presentadas directamente e incorporan el formato numérico español.

Palabras clave: tablas - gtsummary - reproducibilidad - español

Demostración

Se muestran tablas realizadas con el paquete `gtsummary` (Sjoberg et al. 2021) a partir del conjunto de datos `pinguinos` (Quiroga et al. 2021), traducción al español del conjunto de datos `penguins` del paquete `palmerpenguins` (Horst, Hill, and Gorman 2020). Contiene datos de pingüinos de 3 islas del archipiélago de Palmer, en la Antártica. Para cada observación incluye: la especie del pingüino, la isla, sexo, año e información sobre su tamaño (*largo de aleta*, *masa corporal*, *largo* y *alto de pico*).

El análisis inicialmente será únicamente considerando la isla **Dream**. La función `tbl_summary` permite crear una tabla con medidas de resumen. Por defecto, aparece mediana y rango intercuartil para variables numéricas, y recuento (n) y porcentaje para variables categóricas (figura 1).

Figura 1

Characteristic	N = 124 ¹
<code>largo_pico_mm</code>	44.7 (39.2, 49.8)
<code>alto_pico_mm</code>	18.40 (17.50, 19.00)
<code>largo_aleta_mm</code>	193 (188, 198)
<code>masa_corporal_g</code>	3,688 (3,400, 3,956)
sexo	
hembra	61 (50%)
macho	62 (50%)
Unknown	1
¹ Median (IQR); n (%)	

Personalizando las tablas

Para excluir variables que no sean de interés, basta con filtrar y seleccionar los datos de interés antes de ejecutar `tbl_summary`. El argumento `by` permite **separar** los datos en **categorías**.

Para agregar test de hipótesis se emplea la función `add_p` sobre resultado de `tbl_summary`. Por defecto, para variables numéricas entrega el valor p de test de hipótesis no paramétricos (figura 2).

Figura 2

Characteristic	hembra, N = 61 ¹	macho, N = 62 ¹	p-value ²
especie			>0.9
Adelia	27 (44%)	28 (45%)	
Barbijo	34 (56%)	34 (55%)	
Papúa	0 (0%)	0 (0%)	
<code>largo_pico_mm</code>	42.5 (37.0, 46.4)	49.1 (40.6, 51.2)	<0.001
<code>alto_pico_mm</code>	17.80 (17.00, 18.20)	19.00 (18.50, 19.70)	<0.001
<code>largo_aleta_mm</code>	190 (187, 195)	196 (191, 201)	<0.001
<code>masa_corporal_g</code>	3,450 (3,300, 3,650)	3,950 (3,756, 4,250)	<0.001
¹ n (%); Median (IQR)			
² Fisher's exact test; Wilcoxon rank sum test			

Formato numérico español

La función `theme_gtsummary_language` permite traducir el **formato numérico a español**, y permite elegir los separadores para miles y decimales.

Es posible modificar el texto que aparece en la tabla. En este caso se cambió la etiqueta de *especie*, aquellas relacionadas con el tamaño del pingüino, y *Characteristic* por *Variable*.

Se cambiaron las **medidas de resumen y test de hipótesis** para variables continuas a promedio, desviación estándar y prueba T de Student, respectivamente, y se agregó un **encabezado** para agrupar *macho* y *hembra* (figura 3).

Figura 3

Variable	Sexo		valor p ²
	hembra, N = 61 ¹	macho, N = 62 ¹	
Especie			>0,9
Adelia	27 (44%)	28 (45%)	
Barbijo	34 (56%)	34 (55%)	
Papúa	0 (0%)	0 (0%)	
Largo del pico (mm)	42,3 (5,5)	46,1 (5,8)	<0,001
Alto del pico (mm)	18 (0,8)	19 (0,9)	<0,001
Largo de la aleta (mm)	190 (6)	196 (7)	<0,001
Masa corporal (g)	3.446 (270)	3.987 (350)	<0,001
¹ n (%); Media (DE)			
² test exacto de Fisher; t de Student			

Reproducibilidad

Si fuera necesario **replicar** esta tabla, incluyendo los datos de todas las islas, la tabla resultante sería la Tabla 4. En ella se mantienen las medidas resumen elegidas en el paso anterior.

Como el tamaño muestral es diferente, cambian los test estadísticos que aparecen por defecto. Para este caso, la prueba de hipótesis de la variable categórica *especie* cambia de prueba exacta de Fisher a prueba de Chi cuadrado.

Otras consideraciones

La **nota al pie** de la tabla que incluye por defecto las medidas resumen de estadística descriptiva y los test de hipótesis se puede eliminar o sobrescribir manualmente según sea necesario con la función `modify_footnote`.

En figura 5 se observa la parte inferior de una tabla similar a la figura 4, con una modificación de la nota al pie.

Figura 4

Variable	Sexo		valor p ²
	hembra, N = 165 ¹	macho, N = 168 ¹	
Especie			>0,9
Adelia	73 (44%)	73 (43%)	
Barbijo	34 (21%)	34 (20%)	
Papúa	58 (35%)	61 (36%)	
Isla			>0,9
Biscoe	80 (48%)	83 (49%)	
Dream	61 (37%)	62 (37%)	
Torgersen	24 (15%)	23 (14%)	
Largo del pico (mm)	42,1 (4,9)	45,9 (5,4)	<0,001
Alto del pico (mm)	16 (1,8)	18 (1,9)	<0,001
Largo de la aleta (mm)	197 (13)	205 (15)	<0,001
Masa corporal (g)	3.862 (666)	4.546 (788)	<0,001
¹ n (%); Media (DE)			
² prueba chi cuadrado de independencia; t de Student			

Figura 5

Largo de la aleta (mm)	197 (13)	205 (15)	<0,001
Masa corporal (g)	3.862 (666)	4.546 (788)	<0,001
¹ Recuentos (%); media (desviación estándar)			
² Prueba Chi cuadrado; prueba T para muestras independientes			

Referencias

- Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. "Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data." Manual. <https://allisonhorst.github.io/palmerpenguins/>.
- Quiroga, Riva, Edgar Ruiz, Mauricio Vargas, and Mauro Lepore. 2021. "Datos: Traduce Al Español Varios Conjuntos de Datos de Práctica." Manual. <https://github.com/cienciadedatos/datos>.
- Sjoberg, Daniel D., Michael Curry, Margie Hannum, Karissa Whiting, and Emily C. Zabor. 2021. "Gtsummary: Presentation-Ready Data Summary and Analytic Result Tables." Manual. <https://CRAN.R-project.org/package=gtsummary>.