gt summary

Tablas analíticas y de resumen listas para publicar con R

Cheat Sheet

Funciones principales para tablas

tbl_summary()
Estadísticos descriptivos para variables continuas, categóricas y dicotómicas.

tbl_regression()
Convierte un objeto de modelo de regresión en una tabla formateada y personalizada.

tbl_survfit()
Convierte un objeto survfit en una tabla con estimaciones de tiempo hasta el evento.

tbl_summary() resume columnas específicas de los datos con sintaxis tidyverse y personalización flexible (¡Revisa la viñeta!)

Código básico

```
trial %>% select(trt, age, grade, response) %>% tbl_summary()
```

Tabla básica

Characteristic	N = 200[1]
Chemotherapy Treatment	
Drug A	98 (49%)
Drug B	102 (51%)
Age, yrs	47 (38, 57)
Unknown	11
Grade	
I	68 (34%)
II	68 (34%)
III	64 (32%)
Tumor Response	61 (32%)
Unknown	7

[1] n (%); Median (IQR)

tbl_svysummary() para objetos de encuesta

Igual que tbl_summary(), pero recibe un objeto de encuesta y toma en cuenta los pesos y el diseño de la encuesta.

Más información en:
http://www.danieldsjoberg.com/gtsummary/reference/tbl_svysummary.html

Código personalizado

```
trial %>%
 select(trt, age, grade, response) %>%
 tbl_summary(
  by = trt,
  label = list(age ~ "Age (years)",
               grade ~ "Tumor grade"),
  percent = "row",
  digits = list(age ~ 2),
  statistic = list(age ~ "{mean} ({sd})",
                   response ~ "{n}/{N} ({p}%)"),
  type = list(response ~ "categorical"),
  missing = "always",
  missing_text = "Missing"
  )
```

Tabla personalizada

Characteristic	Drug A, N = 98[1]	Drug B, N = 102[1]
Age (years)	47.01 (14.71)	47.45 (14.01)
Missing	7	4
Tumor grade		
I	35 (51%)	33 (49%)
II	32 (47%)	36 (53%)
III	31 (48%)	33 (52%)
Missing	0	0
Tumor Response		
0	67/132 (51%)	65/132 (49%)
1	28/61 (46%)	33/61 (54%)
Missing	3	4

[1] Mean (SD); n (%); n/N (%)

Más información sobre argumentos y opciones de personalización en http://www.danieldsjoberg.com/gtsummary/reference/tbl_summary.html

Opciones de personalización

Argumento	Entrada	Efecto en la tabla
by	Columna para la tabulación cruzada	Los estadísticos de resumen se calcularán por separado para cada nivel de la variable
label	Lista de fórmulas con las etiquetas de las variables	Cambia el texto del nombre de la variable en la tabla
statistic	Lista de fórmulas con el tipo de estadístico de resumen para cada variable	Cambia los estadísticos de resumen mostrados para las variables indicadas
digits	Lista de fórmulas con el número de decimales a mostrar	Cambia el número de decimales redondeados en la tabla para las variables continuas indicadas
type	Lista de fórmulas que especifica el tipo de cada variable	Cambia el tipo de las variables indicadas, lo que define qué estadísticos de resumen se muestran
value	Lista de fórmulas con el valor a mostrar para variables dicotómicas	Cambia el valor mostrado para las variables de tipo dicotómico
missing	“no”, “ifany”, “always”	Define si se reportan las observaciones faltantes
missing_text	Texto a mostrar para el conteo de observaciones faltantes	Cambia el nombre del nivel de datos faltantes en las variables correspondientes
sort	Lista de fórmulas con el tipo de ordenamiento a aplicar (“frequency” o “alphanumeric”)	Cambia el tipo de ordenamiento de las variables categóricas
percent	“column”, “row” o “cell”	Cambia el cálculo y la visualización de porcentajes

Funciones auxiliares extensiones útiles para tbl_summary()

Código ampliado

```
trial %>%
  select(trt, age,
         response) %>%
  tbl_summary(
    by = trt,
    missing= "no"
    ) %>%
  add_n() %>%
  add_overall() %>%
  add_p()
```

Tabla ampliada

Characteristic	N	Overall, N = 200[1]	Drug A, N = 98[1]	Drug B, N = 102[1]	p-value[2]
Age, yrs	189	47 (38, 57)	46 (37, 59)	48 (39, 56)	0.7
Tumor Response	193	61 (32%)	28 (29%)	33 (34%)	0.5

[1] Median (IQR); n (%)
[2] Wilcoxon rank sum test; Pearson's Chi-squared test

add_n()
Añade columna con el total de observaciones no faltantes

add_overall()
Añade una columna con los estadísticos de resumen globales

Ver también:
add_q()
bold_p()
bold_labels()
add_stat()

add_p() Añade una columna con valores p de pruebas de diferencias entre grupos. Argumentos:

Argumento	Predeterminado	Entrada	Efecto en la tabla
test	Continuas: "wilcox.test" (2 grupos), "kruskal.test" (>2); Categóricas (≥5): "chisq.test.no.correct"; Categóricas (<5): "fisher.test"	Lista de fórmulas con la prueba para cada variable: "t.test", "aov", "wilcox.test", "kruskal.test", "chisq.test" y "lme4"; también pruebas personalizadas	Cambia el valor p de la tabla según la prueba estadística indicada
pvalue_fun	style_pvalue()	Función para redondear y dar formato a los valores p	Cambia el formato de los valores p en la tabla

Para más información, visita http://www.danieldsjoberg.com/gtsummary/reference/index.html

tbl_regression() Presenta un objeto de modelo de regresión en una tabla lista para publicar

Modelo de Cox: código básico

```
library(survival)
cox1 = coxph(Surv(ttdeath, death) ~ age + marker,
             data = trial)
tbl_regression(cox1, exponentiate = TRUE)
```

Modelo de Cox: tabla básica

Characteristic	HR[1]	95% CI[1]	p-value
Age, yrs	1.01	0.99, 1.02	0.4
Marker Level, ng/mL	0.96	0.76, 1.21	0.7

[1] HR = Hazard Ratio, CI = Confidence Interval

1. **Construye el modelo de interés**
2. **Muestra resultados con tbl_regression()**

tbl_regression() admite los modelos de regresión más usados y usa `broom::tidy(x)` para ordenar inicialmente los resultados.

Más información:
http://www.danieldsjoberg.com/gtsummary/reference/tbl_regression.html

Modelo glm: código básico

```
m1 <- glm(response ~ age + stage, data = trial,
          family = binomial)
tbl_regression(m1, exponentiate = TRUE)
```

Modelo glm: tabla básica

Characteristic	OR[1]	95% CI[1]	p-value
Age, yrs	1.02	1.00, 1.04	0.091
T Stage			
T1	—	—	
T2	0.58	0.24, 1.37	0.2
T3	0.94	0.39, 2.28	0.9
T4	0.79	0.33, 1.90	0.6

[1] OR = Odds Ratio, CI = Confidence Interval

Modelo glm: código personalizado con funciones auxiliares

```
m1 %>%
  tbl_regression(exponentiate = TRUE) %>%
  add_global_p() %>%
  bold_p(t = 0.10) %>%
  bold_labels() %>%
  italicize_levels()
```

Auxiliares

Modelo glm: tabla a medida

Characteristic	OR[1]	95% CI[1]	p-value
Age, yrs	1.02	1.00, 1.04	**0.087**
T Stage			0.6
T1	—	—	
T2	0.58	0.24, 1.37	
T3	0.94	0.39, 2.28	
T4	0.79	0.33, 1.90	

[1] OR = Odds Ratio, CI = Confidence Interval

tbl_uvregression() muestra varios modelos de regresión univariados a la vez

Modelos de Cox univariados: código

```
library(survival)
tbl_uvregression(
    trial %>% select(ttdeath, death, age,
                     grade, response),
    method = coxph,
    y = Surv(ttdeath, death),
    exponentiate = TRUE,
    label = list(grade ~ "Tumor grade")
    ) %>%
  add_global_p() %>%
  add_n(location = "level") %>%
  add_nevent(location = "level")
```

Modelos de Cox univariados: tabla

Characteristic	N	Event N	HR[1]	95% CI[1]	p-value
Age, yrs	189	103	1.01	0.99, 1.02	0.3
Tumor grade					0.075
I	68	33	—	—	
II	68	36	1.28	0.80, 2.05	
III	64	43	1.69	1.07, 2.66	
Tumor Response	193	107	0.50	0.31, 0.78	0.001

[1] HR = Hazard Ratio, CI = Confidence Interval

Requiere el parámetro "`method`" que especifica el tipo de modelo. Puede estimar modelos de regresión univariados manteniendo constante el desenlace ("y"), la covariable ("x") o ambos (ver el parámetro "`formula`"). Para más información sobre "`formula`" y otros parámetros, ver:
https://www.danieldsjoberg.com/gtsummary/reference/tbl_uvregression.html

tbl_survfit() Muestra un objeto survfit con estimaciones a medida en tabla para publicar

```
library(survival)
tbl_survfit(
 list(
    survfit(Surv(ttdeath, death) ~ 1, trial),
    survfit(Surv(ttdeath, death) ~ trt, trial)
 ),
  times = c(12, 24),
  label_header = "**{time} Month**"
)
```

Opciones para definir el modelo:
1. x=modelo survfit explícito (o lista de modelos) a partir de un data frame
2. x=data frame, indicando y=objeto Surv e include=covariables del modelo

Characteristic	12 Month	24 Month
Overall	88% (84%, 93%)	44% (38%, 51%)
Chemotherapy Treatment		
Drug A	91% (85%, 97%)	47% (38%, 58%)
Drug B	86% (80%, 93%)	41% (33%, 52%)

Opciones de salida:
1. supervivencia en tiempos específicos con times=tiempos de interés
2. cuantiles de supervivencia (p. ej., mediana) con probs=cuantil

Más información:
http://www.danieldsjoberg.com/gtsummary/reference/tbl_survfit.html

```
library(survival)
tbl_survfit(
  trial,
  y = Surv(ttdeath, death),
  include = c(trt, grade),
  probs = 0.5,
  label_header = "**Median Survival**"
) %>% add_p()
```

Characteristic	Median Survival	p-value[1]
Chemotherapy Treatment		0.2
Drug A	24 (21, —)	
Drug B	21 (18, —)	
Grade		0.072
I	— (22, —)	
II	22 (18, —)	
III	20 (18, 23)	

[1] Log-rank test

tbl_merge(), tbl_stack() Combina tablas por filas o por columnas

```
t1 = tbl_survfit(
 list(survfit(Surv(ttdeath, death) ~ trt +
grade, trial),
  times = c(12, 24),
  label_header = "**{time} Month**"
)
```

```
t2 = tbl_survfit(
  trial,
  y = Surv(ttdeath, death),
  include = c(trt, grade),
  probs = 0.5,
  label_header = "**Median Survival**"
) %>% add_p()
```

```
tbl_merge(list(t1,t2), tab_spanner = FALSE)
```

Characteristic	12 Month	24 Month	Median Survival	p-value[1]
Chemotherapy Treatment				0.2
Drug A	91% (85%, 97%)	47% (38%, 58%)	24 (21, —)	
Drug B	86% (80%, 93%)	41% (33%, 52%)	21 (18, —)	
Grade				0.072
I	97% (93%, 100%)	51% (41%, 65%)	— (22, —)	
II	82% (74%, 92%)	47% (37%, 61%)	22 (18, —)	
III	86% (78%, 95%)	33% (23%, 47%)	20 (18, 23)	

[1] Log-rank test

tbl_merge combina columnas y tbl_stack() filas. Más información en
https://www.danieldsjoberg.com/gtsummary/reference/