MONASH University

ETC5523: Communicating with Data

Data tables

Lecturer: *Emi Tanaka*

Department of Econometrics and Business Statistics

emi.tanaka@monash.edu

Week 5B

cwd.numbat.space

Aim

- Recall and understand the motivations of various common tables that present statistical information
- Learn guidelines for communicating data with tables
- Make data tables with R

Why

- Your choice of statistical tables to present should consider the motivation of your audience
- You need to be aware of the common convention of how to present information in tables

Some common tables

Descriptive summary statistics tables

```
1 library(tidyverse)
2 mtcars %>%
3   select(mpg, wt, vs, cyl) %>%
4   glimpse()
```

```
Rows: 32
Columns: 4
$ mpg <dbl> 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8, 19.2, 17.8, …
$ wt  <dbl> 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3.150, 3.4…
$ vs  <dbl> 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, …
$ cyl <dbl> 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 8, 4, 4, 4, 4, 8, …
```

	4 (N=11)	**6 (N=7)**	**8 (N=14)**	**Overall (N=32)**
mpg				
Mean (SD)	26.7 (4.51)	19.7 (1.45)	15.1 (2.56)	20.1 (6.03)
Median [Min, Max]	26.0 [21.4, 33.9]	19.7 [17.8, 21.4]	15.2 [10.4, 19.2]	19.2 [10.4, 33.9]
wt				
Mean (SD)	2.29 (0.570)	3.12 (0.356)	4.00 (0.759)	3.22 (0.978)
Median [Min, Max]	2.20 [1.51, 3.19]	3.22 [2.62, 3.46]	3.76 [3.17, 5.42]	3.33 [1.51, 5.42]
vs				
0	1 (9.1%)	3 (42.9%)	14 (100%)	18 (56.3%)
1	10 (90.9%)	4 (57.1%)	0 (0%)	14 (43.8%)

Regression results tables

```
df <- mutate(mtcars,
             cyl = as.factor(cyl))
fit1 <- lm(mpg ~ wt + vs + cyl, data = df)
fit2 <- lm(mpg ~ wt + cyl,  data = df)
```

	fit1			fit2		
Characteristic	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**
wt	-3.3	-4.9, -1.7	<0.001	-3.2	-4.7, -1.7	<0.001
vs	0.86	-2.5, 4.2	0.6			
cyl						
4	—	—		—	—	
6	-3.9	-7.1, -0.67	0.020	-4.3	-7.1, -1.4	0.005
8	-5.1	-10, -0.10	0.046	-6.1	-9.5, -2.7	<0.001

[1] CI = Confidence Interval

What statistics to present?

- **This depends on what you want to convey and your audience.**
- There are two key purposes of the table:

  1. *display* information; and

  2. *communicate* information.

- In general, tables tend to be about display of information and graphs are preferred for communication.
- However, if precision matters then tables can be better at communicating this than graphs.

Descriptive summary statistics

- The main goal is to give a numerical summary to give a “feel” of what the data contains. These generally should convey:
  - variables in the data,
  - the number of observations for each variable,
  - missing values (if any),
  - the distribution, e.g. in the form of five number of summaries or counts/percentages for categorical variables.

Descriptive summary statistics

- For numerical variables, you may have a **correlation table** which displays the correlation coefficients of every pair of variables. Because it is symmetrical, you can omit out the upper triangle.

Variables	Mean	SD	1.	2.
1. Miles per gallon	20.09	6.03		
2. Weight	3.22	0.98	-0.87	
3. Horsepower	146.69	68.56	-0.78	0.66

Descriptive summary statistics

- You may have **cross-tabulations** (also called **contingency tables**)

cyl/gear	3	4	5	Total
4	3.1% (1)	25.0% (8)	6.2% (2)	34.4% (11)
6	6.2% (2)	12.5% (4)	3.1% (1)	21.9% (7)
8	37.5% (12)	0.0% (0)	6.2% (2)	43.8% (14)
Total	46.9% (15)	37.5% (12)	15.6% (5)	100.0% (32)

Regression results

- Or more generally, the important numerical characteristics of your model. This may include:
  - parameter estimates of your model,
  - the standard error or confidence/credible interval of your estimates,
  - model fit quality, e.g. , AIC, BIC and so on.
- You see often the inclusion of -value, usually from the significance test of the corresponding variable and some indicate the significance level by the amount of $*$.
- It is important to ***convey the uncertainty*** for the estimates or predictions from your model.

Regression results

- There is also what is called the **ANOVA table** that shows the variation according to different sources.

```
1 glimpse(ToothGrowth)
```

```
Rows: 60
Columns: 3
$ len  <dbl> 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16.5, 16.5,…
$ supp <fct> VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, V…
$ dose <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0, 1.0, …
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	205.350	205.35000	15.571979	0.0002312
factor(dose)	2	2426.434	1213.21717	91.999965	0.0000000
supp:factor(dose)	2	108.319	54.15950	4.106991	0.0218603
Residuals	54	712.106	13.18715		

Communication considerations for tables

Numerical precision

- Select an appropriate precision for your goal and audience:

	Weight		
Car brand	**5 d.p.**	**2 d.p.**	**0 d.p.**
Mazda RX4	2.61937	2.62	3
Mazda RX4 Wag	2.87518	2.88	3
Datsun 710	2.31916	2.32	2
Hornet 4 Drive	3.21660	3.21	3

Numerical precision

- Display trailing zeroes to match selected precision of the column:

Trailing zeroes	
Yes	No
0.233	0.233
0.320	0.32
0.400	0.4
0.343	0.343

Numerical precision

- Change and display units as appropriate:

Car brand	Weight (lbs)	Weight (1000 lbs)	Weight (kg)	Weight (g)	Weight (mg)
Mazda RX4	2620	2.620	1188	1190000	1,188,000,000
Mazda RX4 Wag	2875	2.875	1304	1300000	1,304,000,000
Datsun 710	2320	2.320	1052	1050000	1,052,000,000
Hornet 4 Drive	3215	3.215	1458	1460000	1,458,000,000

- Show comma every 3 digits (or other marks as needed).
  E.g. `1000000` is harder to read than `1,000,000`.

```
1 scales::comma(c(439024, 4900343), accuracy = 1000)
```

```
[1] "439,000"   "4,900,000"
```

Column alignment

- Spanner labels are usually aligned in center.
- Right-align numbers
- Left-align texts

Car brand			Horsepower		
Left	Center	Right	Left	Center	Right
Mazda RX4	Mazda RX4	Mazda RX4	110	110	110
Mazda RX4 Wag	Mazda RX4 Wag	Mazda RX4 Wag	110	110	110
Datsun 710	Datsun 710	Datsun 710	93	93	93
Hornet 4 Drive	Hornet 4 Drive	Hornet 4 Drive	110	110	110

Labels within tables

- It is possibly obvious, but tables designed as final product (e.g. in report) should have polished labels
- For columns, the unit may be written in the column header label
- You shouldn't label the unit within the table

✅

Car brand	Weight (1000 lbs)	Displacement (cubic inches)
Mazda RX4	2.620	160
Mazda RX4 Wag	2.875	160
Datsun 710	2.320	108
Hornet 4 Drive	3.215	258
Hornet Sportabout	3.440	360
Valiant	3.460	225

❌

car	wt	disp
Mazda RX4	2620 lbs	160 cubic inches
Mazda RX4 Wag	2875 lbs	160 cubic inches
Datsun 710	2320 lbs	108 cubic inches
Hornet 4 Drive	3215 lbs	258 cubic inches
Hornet Sportabout	3440 lbs	360 cubic inches
Valiant	3460 lbs	225 cubic inches

Texts accompanying tables

- Besides the contents of table, a table may be accompanied with: ***table header***, ***caption***, ***footnotes*** and/or ***source notes***.
- The conventions of how and what to write will depend on your audience and medium of report
- Generally if you are *communicating* information, your caption should:
  - summarise the take-away message, in other words, why should the audience care about this table?
  - give context of the table (e.g. " means that the virus is more infectious") $R_0 > 1$

Making tables with R

How to make tables in R?

- There are many packages that make table in R, including ones that wrangle the data for you to make specialised table output. E.g. `knitr::kable`, `kableExtra`, `formattable`, `gt`, `DT`, `pander`, `xtable`, `stargazer`.
- You can read the documentation for each packages to make the table you want (I mainly use `knitr::kable`, `kableExtra` and `DT`).
- Whatever package you use, it's important that you understand the output and how it works with the medium you are trying to display the table.

Table in Markdown

In markdown file:

```
First Header  | Second Header
------------- | -------------
Content Cell  | Content Cell
Content Cell  | Content Cell
```

Possible display:

First Header	Second Header
Content Cell	Content Cell
Content Cell	Content Cell

Table in HTML & PDF

HTML → Web browser

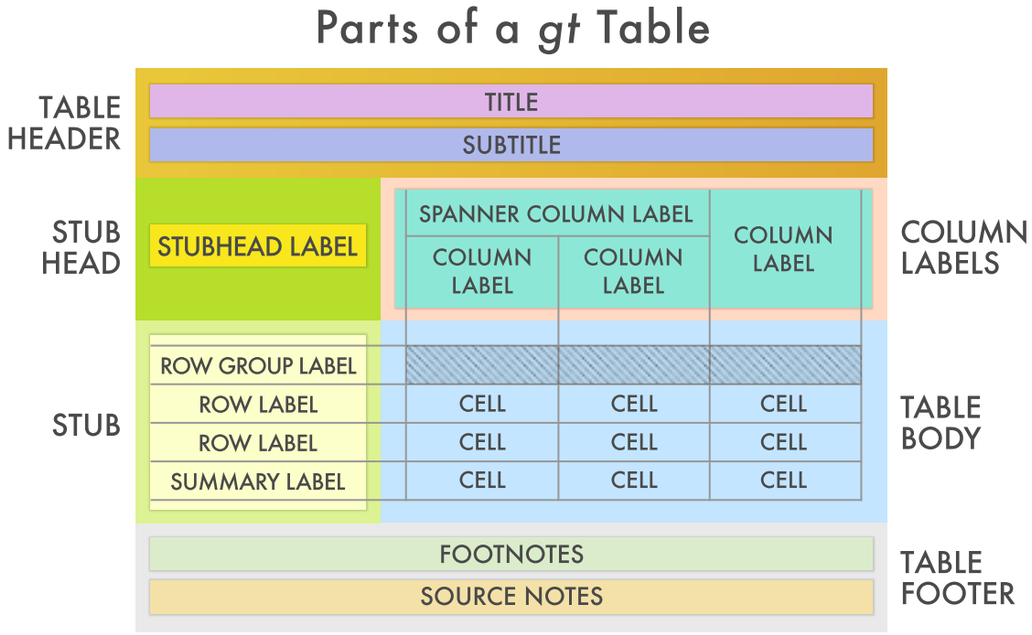
```html
<table>
<thead>
<tr>
<th>First Header</th> <th>Second Header</th
</tr>
</thead>
<tbody>
<tr>
<td>Content Cell</td> <td>Content Cell</td>
</tr>
<tr>
<td>Content Cell</td> <td>Content Cell</td>
</tr>
</tbody>
</table>
```

LaTeX → PDF

```latex
\begin{tabular}{cc}
\hline
First Header & Second Header\\
\hline
Content Cell & Content Cell\\
Content Cell & Content Cell\\
\hline
\end{tabular}
```

- HTML is for HTML output
- LaTeX is for PDF output
- Most HTML elements do *not* work in LaTeX and vice versa.

Case study: building static tables by components with `gt`



Source: https://gt.rstudio.com/

Case study: building static tables by components with `gt`

```
1 library(gt)
2 df <- select(head(mtcars), wt, disp, cyl)
```

Case study: building static tables by components with gt

```
1 library(gt)
2 df <- select(head(mtcars), wt, disp, cyl)
3 gt(df)
```

wt	disp	cyl
2.620	160	6
2.875	160	6
2.320	108	4
3.215	258	6
3.440	360	8
3.460	225	6

Case study: building static tables by components with gt

```
library(gt)
df <- select(head(mtcars), wt, disp, cyl)
gt(df) %>%
  tab_header(title = "Motor Trend Car Road Tests",
             subtitle = "Design and performance of 1973-74 automobile models")
```

Motor Trend Car Road Tests

Design and performance of 1973-74 automobile models

wt	disp	cyl
2.620	160	6
2.875	160	6
2.320	108	4
3.215	258	6
3.440	360	8
3.460	225	6

Case study: building static tables by components with gt

```
library(gt)
df <- select(head(mtcars), wt, disp, cyl)
gt(df) %>%
  tab_header(title = "Motor Trend Car Road Tests",
             subtitle = "Design and performance of 1973-74 automobile models") %>%
  tab_source_note(md("Source: 1974 *Motor Trend* US magazine"))
```

Motor Trend Car Road Tests

Design and performance of 1973-74 automobile models

wt	disp	cyl
2.620	160	6
2.875	160	6
2.320	108	4
3.215	258	6
3.440	360	8
3.460	225	6

Source: 1974 *Motor Trend* US magazine

Case study: building static tables by components with gt

```
 1  library(gt)
 2  df <- select(head(mtcars), wt, disp, cyl)
 3  gt(df) %>%
 4    tab_header(title = "Motor Trend Car Road Tests",
 5               subtitle = "Design and performance of 1973-74 automobile models") %>%
 6    tab_source_note(md("Source: 1974 *Motor Trend* US magazine")) %>%
 7    cols_label(
 8      wt = html("Weight<br>(1000lbs)"),
 9      disp = html("Displacement<br> (inch<sup>3</sup>)"),
10      cyl = html("Number of<br>cylinders")
11    )
```

Motor Trend Car Road Tests

Design and performance of 1973-74 automobile models

Weight (1000lbs)	Displacement ($inch^3$)	Number of cylinders
2.620	160	6
2.875	160	6
2.320	108	4

Source: 1974 *Motor Trend* US magazine

Motor Trend Car Road Tests

Design and performance of 1973-74 automobile models

3.215	258	6
3.440	360	8
3.460	225	6

Source: 1974 *Motor Trend* US magazine

Case study: interactive tables with DT

```
1  DT::datatable(mtcars, options = list(pageLength = 4))
```

Show 4 entries Search:

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1

Showing 1 to 4 of 32 entries Previous 1 2 3 4 5 … 8 Next

When do you make tables over plots?

- When you want to ***show exact values*** or the ***accuracy of the values are important*** to convey.

- You *can* combine plots with tables!

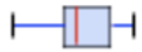
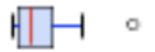
Case study heatmap table with `formattable`

```
1 library(formattable)
2 mtcars %>%
3   select(mpg, disp, wt, hp) %>%
4   cor() %>%
5   as.data.frame() %>%
6   rownames_to_column("Variables") %>%
7   formattable(list(area(col = 2:5) ~ color_tile("#F5B7B1", "#7DCEA0")))
```

Variables	mpg	disp	wt	hp
mpg	1.0000000	-0.8475514	-0.8676594	-0.7761684
disp	-0.8475514	1.0000000	0.8879799	0.7909486
wt	-0.8676594	0.8879799	1.0000000	0.6587479
hp	-0.7761684	0.7909486	0.6587479	1.0000000

Case study: inline plots with `sparkline`

- **Sparkline** refers to a small chart drawn without axes or coordinates.
- This will be out of scope for this unit.

Stock	Sparkline	Boxplot
x		
y		

Week 5B Lesson

Summary

- You saw various common tables that present information and the motivation behind it
- We went through some guidelines for best way to communicate with tables

Resources

- `kableExtra` documentation
- Introduction to Creating `gt` Tables
- `DT` documentation
- `gtsummary` for summary tables
- `modelsummary` for regression tables