

My Report

H24106086 JIANG

2024-09-17

Table of contents

Statistical Thinking	1
Summary Staistic	1
Table 1	5
Missing Values	6

Statistical Thinking

Reference: <https://www.fharrell.com/post/rflow/>

Summary Staistic

```
library(Hmisc)
```

Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

```
format.pval, units
```

```
library(palmerpenguins)
```

Warning: package 'palmerpenguins' was built under R version 4.3.3

```
latex(describe(data.frame(penguins_raw)), file = "", caption.placement = "top")
```

data.frame(penguins_raw)
17 Variables 344 Observations

studyName

n	missing	distinct
344	0	3

Value	PAL0708	PAL0809	PAL0910
Frequency	110	114	120
Proportion	0.320	0.331	0.349

Sample.Number

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
344	0	152	1	63.15	46.35	6.15	12.00	29.00	58.00	95.25	121.00	134.85

lowest : 1 2 3 4 5, highest: 148 149 150 151 152

Species

n	missing	distinct
344	0	3

Value	Adelie Penguin (Pygoscelis adeliae)	Chinstrap penguin (Pygoscelis antarctica)
Frequency	152	68
Proportion	0.442	0.198

Value	Gentoo penguin (Pygoscelis papua)
Frequency	124
Proportion	0.360

Region

n	missing	distinct	value
344	0	1	Anvers

Value	Anvers
Frequency	344
Proportion	1

Island

n missing distinct
344 0 3

Value	Biscoe	Dream	Torgersen
Frequency	168	124	52
Proportion	0.488	0.360	0.151

Stage

n missing distinct value
344 0 1 Adult, 1 Egg Stage

Value	Adult, 1 Egg Stage
Frequency	344
Proportion	1

Individual.ID

n missing distinct
344 0 190

lowest : N100A1 N100A2 N10A1 N10A2 N11A1 , highest: N98A2 N99A1 N99A2 N9A1 N9A2

Clutch.Completion

n missing distinct
344 0 2

Value	No	Yes
Frequency	36	308
Proportion	0.105	0.895

Date.Egg



n	missing	distinct	Info	Mean	Gmd	.05	.10
344	0	50	0.999	2008-11-27	328	2007-11-12	2007-11-16
.25	.50	.75	.90	.95			
2007-11-28	2008-11-09	2009-11-16	2009-11-22	2009-11-26			

lowest : 2007-11-09 2007-11-10 2007-11-11 2007-11-12 2007-11-13
highest: 2009-11-22 2009-11-23 2009-11-25 2009-11-27 2009-12-01

Culmen.Length..mm.



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	164	1	43.92	6.274	35.70	36.60	39.23	44.45	48.50	50.80	51.99

lowest : 32.1 33.1 33.5 34 34.1, highest: 55.1 55.8 55.9 58 59.6

Culmen.Depth..mm.



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	80	1	17.15	2.267	13.9	14.3	15.6	17.3	18.7	19.5	20.0

lowest : 13.1 13.2 13.3 13.4 13.5, highest: 20.7 20.8 21.1 21.2 21.5

Flipper.Length..mm.



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	55	0.999	200.9	16.03	181.0	185.0	190.0	197.0	213.0	220.9	225.0

lowest : 172 174 176 178 179, highest: 226 228 229 230 231

Body.Mass..g.



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	94	1	4202	911.8	3150	3300	3550	4050	4750	5400	5650

lowest : 2700 2850 2900 2925 2975, highest: 5850 5950 6000 6050 6300

Sex

n	missing	distinct
333	11	2

Value	FEMALE	MALE
Frequency	165	168
Proportion	0.495	0.505

Δ.15.N..o.oo.:



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
330	14	330	1	8.733	0.6323	7.897	8.047	8.300	8.652	9.172	9.491	9.689

lowest : 7.6322 7.63452 7.63884 7.68528 7.6887 , highest: 9.93727 9.98044 10.0202 10.0237 10.0254

Δ.13.C..o.oo.:



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
331	13	331	1	-25.69	0.9093	-26.79	-26.69	-26.32	-25.83	-25.06	-24.53	-24.36

lowest : -27.0185 -26.9547 -26.8964 -26.8648 -26.8635, highest: -24.1657 -24.1026 -23.9031 -23.8902 -23.7877

Comments



n	missing	distinct
54	290	10

lowest : Adult not sampled.
highest: No blood sample obtained.

Adult not sampled. Nest never observed with ful
No delta15N data received from lab.

中文

Table 1

```
library(table1)
df <- data.frame(penguins_raw)
str(df)
```

```
'data.frame': 344 obs. of 17 variables:
 $ studyName      : chr  "PAL0708" "PAL0708" "PAL0708" "PAL0708" ...
 $ Sample.Number  : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Species        : chr  "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pygoscelis adeliae)" ...
 $ Region         : chr  "Anvers" "Anvers" "Anvers" "Anvers" ...
 $ Island         : chr  "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
 $ Stage          : chr  "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" ...
 $ Individual.ID  : chr  "N1A1" "N1A2" "N2A1" "N2A2" ...
 $ Clutch.Completion : chr  "Yes" "Yes" "Yes" "Yes" ...
 $ Date.Egg       : Date, format: "2007-11-11" "2007-11-11" ...
 $ Culmen.Length..mm. : num  39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
 $ Culmen.Depth..mm. : num  18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
 $ Flipper.Length..mm.: num  181 186 195 NA 193 190 181 195 193 190 ...
 $ Body.Mass..g.    : num  3750 3800 3250 NA 3450 ...
 $ Sex            : chr  "MALE" "FEMALE" "FEMALE" NA ...
 $ Delta.15.N..o.o. : num  NA 8.95 8.37 NA 8.77 ...
 $ Delta.13.C..o.o. : num  NA -24.7 -25.3 NA -25.3 ...
 $ Comments       : chr  "Not enough blood for isotopes." NA NA "Adult not sampled." ...
```

```
table1(~ (df$Culmen.Length..mm.)+(df$Culmen.Depth..mm.)| df$Clutch.Completion,data=df)
```

	No	Yes	Overall
	(N=36)	(N=308)	(N=344)
dfCulmen.Length..mm. Mean(SD)	44.6(5.77)	43.8(5.43)	43.9(5.46) Median[Min, Max]
Mean (SD)	17.7 (1.86)	17.1 (1.98)	17.2 (1.97)
Median [Min, Max]	17.9 [13.7, 20.0]	17.2 [13.1, 21.5]	17.3 [13.1, 21.5]
Missing	1 (2.8%)	1 (0.3%)	2 (0.6%)

```
# help(table1)
```

Missing Values

```
library(Hmisc)
library(DataExplorer)
```

Warning: package 'DataExplorer' was built under R version 4.3.3

```
plot_missing(df)
```

