

Statistical consulting Homework2

s76134124何佩勳

2025-03-18

目錄

Question 1	1
Question 2:	23

Question 1

Please use Quarto to generate a summary report for the mushroom dataset.

資料來源:Mushroom species drawn from source book:Patrick Hardin.Mushrooms & Toadstools.Zondervan, 1999

以下表格為所有變數的定義整理(coding book)：

Variable Name	Data Type	Definition	Note
family	nominal	Mushroom family name	Multinomial family category
name	nominal	Scientific name of mushroom	Multinomial species category
class	nominal	Edibility classification	p = poisonous, e = edible
cap-diameter	continuous	Cap diameter size in cm	Float value or range min-max or mean
cap-shape	nominal	Shape of the mushroom cap	b = bell, c = conical, x = convex, f = flat, s = sunken, p = spherical, o = others
cap-surface	nominal	Surface texture of the cap	i = fibrous, g = grooves, y = scaly, s = smooth, h = shiny, l = leathery, k = silky, t = sticky, w = wrinkled, e = fleshy
cap-color	nominal	Color of the mushroom cap	n = brown, b = buff, g = gray, r = green, p = pink, u = purple, e = red, w = white, y = yellow, l = blue, o = orange, k = black

Variable Name	Data Type	Definition	Note
does-bruise-bleed	nominal	Does it bruise or bleed	t = bruises or bleeding, f = no bruising or bleeding
gill-attachment	nominal	Attachment of gills to stem	a = adnate, x = adnexed, d = decurrent, e = free, s = sinuate, p = pores, f = none, ? = unknown
gill-spacing	nominal	Spacing between gills	c = close, d = distant, f = none
gill-color	nominal	Color of the gills	Same as cap-color plus f = none
stem-height	continuous	Stem height in cm	Float value or range min-max or mean
stem-width	continuous	Stem width in mm	Float value or range min-max or mean
stem-root	nominal	Type of stem root	b = bulbous, s = swollen, c = club, u = cup, e = equal, z = rhizomorphs, r = rooted
stem-surface	nominal	Surface texture of the stem	Same as cap-surface plus f = none
stem-color	nominal	Color of the stem	Same as cap-color plus f = none
veil-type	nominal	Type of veil covering	p = partial, u = universal
veil-color	nominal	Color of the veil	Same as cap-color plus f = none
has-ring	nominal	Presence of a ring	t = ring present, f = none
ring-type	nominal	Type of ring	c = cobwebby, e = evanescent, r = flaring, g = grooved, l = large, p = pendant, s = sheathing, z = zone, y = scaly, m = movable, f = none, ? = unknown
spore-print-color	nominal	Color of spore print	Same as cap-color
habitat	nominal	Where it is found	g = grasses, l = leaves, m = meadows, p = paths, h = heaths, u = urban, w = waste, d = woods
season	nominal	Season when it grows	s = spring, u = summer, a = autumn, w = winter

資料整理:

```
library(dplyr)
file_path <- "primary_data.csv"
data <- read.csv(file_path, sep=";", stringsAsFactors=FALSE)
```

```
data <- data %>% mutate(across(everything(), trimws))
head(data)
```

	family	name	class	cap.diameter	cap.shape	Cap.surface
1	Amanita Family	Fly Agaric	p	[10, 20]	[x, f]	[g, h]
2	Amanita Family	Panther Cap	p	[5, 10]	[p, x]	[g]
3	Amanita Family	False Panther Cap	p	[10, 15]	[x, f]	
4	Amanita Family	The Blusher	e	[5, 15]	[x, f]	
5	Amanita Family	Death Cap	p	[5, 12]	[x, f]	[h]
6	Amanita Family	False Death Cap	e	[4, 9]	[x]	

	cap.color	does.bruise.or.bleed	gill.attachment	gill.spacing	gill.color
1	[e, o]		[f]	[e]	[w]
2	[n]		[f]	[e]	[w]
3	[g, n]		[f]	[e]	[w]
4	[n]		[t]		[w]
5	[r]		[f]		[w]
6	[w, y]		[f]	[e]	[w]

	stem.height	stem.width	stem.root	stem.surface	stem.color	veil.type	veil.color
1	[15, 20]	[15, 20]	[s]	[y]	[w]	[u]	[w]
2	[6, 10]	[10, 20]		[y]	[w]	[u]	[w]
3	[10, 12]	[10, 20]			[w]	[u]	[w]
4	[7, 15]	[10, 25]	[b]		[w]	[u]	[w]
5	[10, 12]	[10, 20]			[w]	[u]	[w]
6	[5, 7]	[10, 15]	[b]		[w, y]	[u]	[y, w]

	has.ring	ring.type	Spore.print.color	habitat	season
1	[t]	[g, p]		[d]	[u, a, w]
2	[t]	[p]		[d]	[u, a]
3	[t]	[e, g]		[d]	[u, a]
4	[t]	[g]		[d]	[u, a]
5	[t]	[g, p]		[d]	[u, a]
6	[t]	[g]		[d]	[u, a]

描述性統計：

有missing data的變項:

```
colSums(data == "" | is.na(data))
```

family	name	class
0	0	0
cap.diameter	cap.shape	Cap.surface
0	0	40
cap.color	does.bruise.or.bleed	gill.attachment
0	0	28
gill.spacing	gill.color	stem.height
71	0	0
stem.width	stem.root	stem.surface
0	146	108
stem.color	veil.type	veil.color
0	164	152
has.ring	ring.type	Spore.print.color
0	7	155
habitat	season	
0	0	

本資料一個樣本的一個變項可能有多個level,如:cap.shape變項中可以既x又f的level,故需要進一步整理表格。

- (1) 類別變項整理:將一個樣本的一個變項有重複的level的這些變項,將這些變項的level拆成單獨的變項,如:cap.shape拆成cap.shape_x, cap.shape_f, cap.shape_p, cap.shape_b, cap.shape_c, cap.shape_s, cap.shape_o。

```
#set function to product a table of variable
process_column_to_matrix <- function(data, column_name) {

  if (!(column_name %in% names(data))) {
    stop(paste(" ", column_name, " "))
  }

  data[[column_name]] <- ifelse(is.na(data[[column_name]]), "",
                                gsub("[^a-zA-Z,]", "", data[[column_name]]))

  level <- unique(unlist(strsplit(data[[column_name]], ",")))

  ta <- matrix(NA, nrow = nrow(data), ncol = length(level))
  colnames(ta) <- paste(column_name, level, sep = "_")

  for (i in seq_along(level)) {
    ta[which(grepl(paste0("\\b", level[i], "\\b"), data[[column_name]])), i] <- level[i]
  }

  return(as.data.frame(ta))
}

#use these columues
names(data) <- tolower(names(data))
columns_to_process <- c("cap.shape", "cap.surface", "cap.color", "gill.attachment",
                        "gill.color", "stem.surface", "stem.color", "veil.color",
                        "ring.type", "spore.print.color", "habitat", "season")

processed_tables <- lapply(columns_to_process,
                           function(col) process_column_to_matrix(data, col))

final_table <- do.call(cbind, processed_tables)
```

(註):以上有寫簡單的一個變項整理再用AI工具協助修飾code。

- (2) 連續變項整理:有兩個數值的取平均數並取代:

```
process_numbers <- function(x) {
  x <- gsub("\\[|\\]", "", x)
  numbers <- as.numeric(unlist(strsplit(x, ",")))
  if (length(numbers) == 2) {
    return(mean(numbers))
  } else {
    return(numbers)
  }
}
```

```
}

data <- data %>%
  mutate(across(c(4, 12, 13), ~sapply(., process_numbers)))
```

(註):以上有寫簡單的一個變項整理再用AI工具協助修飾code。

(3) 將類別變項和連續變項合併成最終資料:

```
final_data <- cbind(data[,c(1,2,3,4,12,13,8,10,14,17,19)],final_table)
```

(4) 完整描述性統計資料:

```
library(reticulate)
library(Hmisc)
latex(describe(final_data), file="")
```

		final_data	
		114 Variables	173 Observations
family			
n	missing	distinct	
173	0	23	
lowest :	Amanita Family	Bolbitius Family	Bolete Family
highest:	Russula Family	Saddle-Cup Family	Stropharia Family
			Bracket Fungi
			Tricholoma Family
			Chanterelle Family
			Wax Gill Family
name			
n	missing	distinct	
173	0	173	
lowest :	Amethyst Deceiver	Aniseed Funnel Cap	Apricot Fungus
highest:	Yellow-gilled Russula	Yellow-staining Mushroom	Yellow-stemmed Bell Cap
			Bare-toothed Russula
			Yellow Swamp Russula
			Bay Bolete
			Yellow Wax cap
class			
n	missing	distinct	
173	0	2	
Value	e	p	
Frequency	77	96	
Proportion	0.445	0.555	
cap.diameter			
n	missing	distinct	Info
173	0	35	0.997
Mean			6.739
pMedian			6.25
Gmd			4.755
.05			1.5
.10			2.0
.25			3.5
.50			6.0
.75			8.5
.90			11.4
.95			15.0
lowest :	0.7	0.75	1
	1.25	1.5	
highest:	16.5	17.5	18.5
	19	50	
stem.height			
n	missing	distinct	Info
173	0	30	0.993
Mean			6.59
pMedian			6.25
Gmd			3.228
.05			3.0
.10			3.5
.25			4.5
.50			6.0
.75			7.5
.90			10.0
.95			12.2
lowest :	0	1.5	2
	2.5	3	
highest:	16	16.5	17.5
	19	25	

stem.width

n	missing	distinct	Info	Mean	pMedian	Gmd	.05	.10	.25	.50	.75	.90	.95
173	0	35	0.997	12.16	11	10.09	1.5	2.5	5.5	10.0	15.0	25.0	30.0

lowest : 0 0.75 1 1.5 2 , highest: 30 35 40 50 70

does.bruise.or.bleed

n	missing	distinct
173	0	2

Value	[f]	[t]
Frequency	143	30
Proportion	0.827	0.173

gill.spacing

n	missing	distinct
102	71	3

Value	[c]	[d]	[f]
Frequency	70	22	10
Proportion	0.686	0.216	0.098

stem.root

n	missing	distinct
27	146	5

Value	[b]	[c]	[f]	[r]	[s]
Frequency	9	2	3	4	9
Proportion	0.333	0.074	0.111	0.148	0.333

veil.type

n	missing	distinct	value
9	164	1	[u]

Value	[u]
Frequency	9
Proportion	1

has.ring

n	missing	distinct
173	0	2

Value	[f]	[t]
Frequency	130	43
Proportion	0.751	0.249

cap.shape_x

n	missing	distinct	value
110	63	1	x

Value	x
Frequency	110
Proportion	1

cap.shape_f

n	missing	distinct	value
74	99	1	f

Value	f
Frequency	74
Proportion	1

cap.shape_p

n	missing	distinct	value
15	158	1	p

Value	p
Frequency	15
Proportion	1

cap.shape_b

n	missing	distinct	value
23	150	1	b

Value	b
Frequency	23
Proportion	1

cap.shape_c

n	missing	distinct	value
8	165	1	c

Value	c
Frequency	8
Proportion	1

cap.shape_s

n	missing	distinct	value
36	137	1	s

Value	s
Frequency	36
Proportion	1

cap.shape_o

n	missing	distinct	value
12	161	1	o

Value	o
Frequency	12
Proportion	1

cap.surface_g

n	missing	distinct	value
16	157	1	g

Value	g
Frequency	16
Proportion	1

cap.surface_h

n	missing	distinct	value
26	147	1	h

Value	h
Frequency	26
Proportion	1

cap.surface_t

n	missing	distinct	value
37	136	1	t

Value	t
Frequency	37
Proportion	1

cap.surface_y

n	missing	distinct	value
23	150	1	y

Value	y
Frequency	23
Proportion	1

cap.surface_e

n	missing	distinct	value
11	162	1	e

Value	e
Frequency	11
Proportion	1

cap.surface_s

n	missing	distinct	value
33	140	1	s

Value	s
Frequency	33
Proportion	1

cap.surface_l

n	missing	distinct	value
4	169	1	l

Value	l
Frequency	4
Proportion	1

cap.surface_d

n	missing	distinct	value
18	155	1	d

Value	d
Frequency	18
Proportion	1

cap.surface_w

n	missing	distinct	value
8	165	1	w

Value	w
Frequency	8
Proportion	1

cap.surface_i

n	missing	distinct	value
9	164	1	i

Value	i
Frequency	9
Proportion	1

cap.surface_k

n	missing	distinct	value
10	163	1	k

Value	k
Frequency	10
Proportion	1

cap.color_e

n	missing	distinct	value
25	148	1	e

Value	e
Frequency	25
Proportion	1

cap.color_o

n	missing	distinct	value
22	151	1	o

Value	o
Frequency	22
Proportion	1

cap.color_n

n	missing	distinct	value
110	63	1	n

Value	n
Frequency	110
Proportion	1

cap.color_g

n	missing	distinct	value
28	145	1	g

Value	g
Frequency	28
Proportion	1

cap.color_r

n	missing	distinct	value
13	160	1	r

Value	r
Frequency	13
Proportion	1

cap.color_w

n	missing	distinct	value
35	138	1	w

Value	w
Frequency	35
Proportion	1

cap.color_y

n	missing	distinct	value
44	129	1	y

Value	y
Frequency	44
Proportion	1

cap.color_p

n	missing	distinct	value
11	162	1	p

Value	p
Frequency	11
Proportion	1

cap.color_b

n	missing	distinct	value
7	166	1	b

Value	b
Frequency	7
Proportion	1

cap.color_u

n	missing	distinct	value
10	163	1	u

Value	u
Frequency	10
Proportion	1

cap.color_l

n	missing	distinct	value
6	167	1	l

Value	l
Frequency	6
Proportion	1

cap.color_k

n	missing	distinct	value
9	164	1	k

Value	k
Frequency	9
Proportion	1

gill.attachment_e

n	missing	distinct	value
16	157	1	e

Value	e
Frequency	16
Proportion	1

gill.attachment_a

n	missing	distinct	value
40	133	1	a

Value	a
Frequency	40
Proportion	1

gill.attachment_d

n	missing	distinct	value
33	140	1	d

Value	d
Frequency	33
Proportion	1

gill.attachment_s

n	missing	distinct	value
16	157	1	s

Value	s
Frequency	16
Proportion	1

gill.attachment_x

n	missing	distinct	value
21	152	1	x

Value	x
Frequency	21
Proportion	1

gill.attachment_p

n	missing	distinct	value
17	156	1	p

Value	p
Frequency	17
Proportion	1

gill.attachment_f

n	missing	distinct	value
10	163	1	f

Value	f
Frequency	10
Proportion	1

gill.color_w

n	missing	distinct	value
73	100	1	w

Value	w
Frequency	73
Proportion	1

gill.color_n

n	missing	distinct	value
47	126	1	n

Value	n
Frequency	47
Proportion	1

gill.color_p

n	missing	distinct	value
28	145	1	p

Value	p
Frequency	28
Proportion	1

gill.color_u

n	missing	distinct	value
7	166	1	u

Value	u
Frequency	7
Proportion	1

gill.color_b

n	missing	distinct	value
5	168	1	b

Value	b
Frequency	5
Proportion	1

gill.color_g

n	missing	distinct	value
23	150	1	g

Value	g
Frequency	23
Proportion	1

gill.color_y

n	missing	distinct	value
44	129	1	y

Value	y
Frequency	44
Proportion	1

gill.color_r

n	missing	distinct	value
8	165	1	r

Value	r
Frequency	8
Proportion	1

gill.color_e

n	missing	distinct	value
6	167	1	e

Value	e
Frequency	6
Proportion	1

gill.color_o

n	missing	distinct	value
13	160	1	o

Value	o
Frequency	13
Proportion	1

gill.color_k

n	missing	distinct	value
15	158	1	k

Value	k
Frequency	15
Proportion	1

gill.color_f

n	missing	distinct	value
10	163	1	f

Value	f
Frequency	10
Proportion	1

stem.surface_y

n	missing	distinct	value
15	158	1	y

Value	y
Frequency	15
Proportion	1

stem.surface_s

n	missing	distinct	value
19	154	1	s

Value	s
Frequency	19
Proportion	1

stem.surface_k

n	missing	distinct	value
5	168	1	k

Value	k
Frequency	5
Proportion	1

stem.surface_i

n	missing	distinct	value
14	159	1	i

Value	i
Frequency	14
Proportion	1

stem.surface_h

n	missing	distinct	value
2	171	1	h

Value	h
Frequency	2
Proportion	1

stem.surface_t

n	missing	distinct	value
8	165	1	t

Value	t
Frequency	8
Proportion	1

stem.surface_g

n	missing	distinct	value
5	168	1	g

Value	g
Frequency	5
Proportion	1

stem.surface_f

n	missing	distinct	value
3	170	1	f

Value	f
Frequency	3
Proportion	1

stem.color_w

n	missing	distinct	value
74	99	1	w

Value	w
Frequency	74
Proportion	1

stem.color_y

n	missing	distinct	value
32	141	1	y

Value	y
Frequency	32
Proportion	1

stem.color_n

n	missing	distinct	value
70	103	1	n

Value	n
Frequency	70
Proportion	1

stem.color_b

n	missing	distinct	value
1	172	1	b

Value	b
Frequency	1
Proportion	1

stem.color_u

n	missing	distinct	value
7	166	1	u

Value	u
Frequency	7
Proportion	1

stem.color_l

n	missing	distinct	value
2	171	1	l

Value	l
Frequency	2
Proportion	1

stem.color_r

n	missing	distinct	value
4	169	1	r

Value	r
Frequency	4
Proportion	1

stem.color_p

n	missing	distinct	value
4	169	1	p

Value	p
Frequency	4
Proportion	1

stem.color_e

n	missing	distinct	value
11	162	1	e

Value	e
Frequency	11
Proportion	1

stem.color_k

n	missing	distinct	value
4	169	1	k

Value	k
Frequency	4
Proportion	1

stem.color_g

n	missing	distinct	value
14	159	1	g

Value	g
Frequency	14
Proportion	1

stem.color_o

n	missing	distinct	value
12	161	1	o

Value	o
Frequency	12
Proportion	1

stem.color_f

n	missing	distinct	value
3	170	1	f

Value	f
Frequency	3
Proportion	1

veil.color_w

n	missing	distinct	value
16	157	1	w

Value	w
Frequency	16
Proportion	1

veil.color_y

n	missing	distinct	value
2	171	1	y

Value	y
Frequency	2
Proportion	1

veil.color_e

n	missing	distinct	value
1	172	1	e

Value	e
Frequency	1
Proportion	1

veil.color_n

n	missing	distinct	value
2	171	1	n

Value	n
Frequency	2
Proportion	1

veil.color_u

n	missing	distinct	value
1	172	1	u

Value	u
Frequency	1
Proportion	1

veil.color_k

n	missing	distinct	value
1	172	1	k

Value	k
Frequency	1
Proportion	1

ring.type_g

n	missing	distinct	value
5	168	1	g

Value	g
Frequency	5
Proportion	1

ring.type_p

n	missing	distinct	value
5	168	1	p

Value	p
Frequency	5
Proportion	1

ring.type_e

n	missing	distinct	value
8	165	1	e

Value	e
Frequency	8
Proportion	1

ring.type_l

n	missing	distinct	value
6	167	1	l

Value	l
Frequency	6
Proportion	1

ring.type_f

n	missing	distinct	value
137	36	1	f

Value	f
Frequency	137
Proportion	1

ring.type_m

n	missing	distinct	value
1	172	1	m

Value	m
Frequency	1
Proportion	1

ring.type_r

n	missing	distinct	value
5	168	1	r

Value	r
Frequency	5
Proportion	1

ring.type_z

n	missing	distinct	value
6	167	1	z

Value	z
Frequency	6
Proportion	1

spore.print.color_w

n	missing	distinct	value
4	169	1	w

Value	w
Frequency	4
Proportion	1

spore.print.color_p

n	missing	distinct	value
4	169	1	p

Value	p
Frequency	4
Proportion	1

spore.print.color_k

n	missing	distinct	value
7	166	1	k

Value	k
Frequency	7
Proportion	1

spore.print.color_r

n	missing	distinct	value
1	172	1	r

Value	r
Frequency	1
Proportion	1

spore.print.color_u

n	missing	distinct	value
1	172	1	u

Value	u
Frequency	1
Proportion	1

spore.print.color_n

n	missing	distinct	value
3	170	1	n

Value	n
Frequency	3
Proportion	1

spore.print.color_g

n	missing	distinct	value
1	172	1	g

Value	g
Frequency	1
Proportion	1

habitat_d

n	missing	distinct	value
151	22	1	d

Value	d
Frequency	151
Proportion	1

habitat_m

n	missing	distinct	value
17	156	1	m

Value	m
Frequency	17
Proportion	1

habitat_g

n	missing	distinct	value
38	135	1	g

Value	g
Frequency	38
Proportion	1

habitat_h

n	missing	distinct	value
13	160	1	h

Value	h
Frequency	13
Proportion	1

habitat_l

n	missing	distinct	value
18	155	1	l

Value	l
Frequency	18
Proportion	1

habitat_p

n	missing	distinct	value
2	171	1	p

Value	p
Frequency	2
Proportion	1

habitat_w

n	missing	distinct	value
1	172	1	w

Value	w
Frequency	1
Proportion	1

habitat_u

n	missing	distinct	value
1	172	1	u

Value	u
Frequency	1
Proportion	1

season_u

n	missing	distinct	value
140	33	1	u

Value	u
Frequency	140
Proportion	1

season_a

n	missing	distinct	value
168	5	1	a

Value	a
Frequency	168
Proportion	1

season_w

n	missing	distinct	value
41	132	1	w

Value	w
Frequency	41
Proportion	1

season_s

n	missing	distinct	value
23	150	1	s

Value	s
Frequency	23
Proportion	1

有重複的level的占比:

```
percentage_list <- list()

for (k in 1:length(columns_to_process)) {
  b <- grepl(paste0("^", columns_to_process[k]), colnames(final_data), ignore.case = TRUE)
  percentage <- round(colSums(!is.na(final_data[,b]))/173*100,2)
  percentage_list[[columns_to_process[k]]] <- percentage
}

print(percentage_list)
```

\$cap.shape

cap.shape_x	cap.shape_f	cap.shape_p	cap.shape_b	cap.shape_c	cap.shape_s
63.58	42.77	8.67	13.29	4.62	20.81
cap.shape_o					
6.94					

\$cap.surface

cap.surface_g	cap.surface_h	cap.surface_t	cap.surface_y	cap.surface_e	
9.25	15.03	21.39	13.29	6.36	
cap.surface_s	cap.surface_l	cap.surface_d	cap.surface_w	cap.surface_i	
19.08	2.31	10.40	4.62	5.20	
cap.surface_k					
5.78					

\$cap.color

cap.color_e	cap.color_o	cap.color_n	cap.color_g	cap.color_r	cap.color_w
14.45	12.72	63.58	16.18	7.51	20.23
cap.color_y	cap.color_p	cap.color_b	cap.color_u	cap.color_l	cap.color_k
25.43	6.36	4.05	5.78	3.47	5.20

\$gill.attachment

gill.attachment_e	gill.attachment_a	gill.attachment_d	gill.attachment_s		
9.25	23.12	19.08	9.25		
gill.attachment_x	gill.attachment_p	gill.attachment_f			
12.14	9.83	5.78			

\$gill.color

gill.color_w	gill.color_n	gill.color_p	gill.color_u	gill.color_b	gill.color_g
42.20	27.17	16.18	4.05	2.89	13.29
gill.color_y	gill.color_r	gill.color_e	gill.color_o	gill.color_k	gill.color_f
25.43	4.62	3.47	7.51	8.67	5.78

\$stem.surface

stem.surface_y	stem.surface_s	stem.surface_k	stem.surface_i	stem.surface_h
8.67	10.98	2.89	8.09	1.16
stem.surface_t	stem.surface_g	stem.surface_f		
4.62	2.89	1.73		

\$stem.color

stem.color_w	stem.color_y	stem.color_n	stem.color_b	stem.color_u	stem.color_l
42.77	18.50	40.46	0.58	4.05	1.16
stem.color_r	stem.color_p	stem.color_e	stem.color_k	stem.color_g	stem.color_o
2.31	2.31	6.36	2.31	8.09	6.94
stem.color_f					
1.73					

\$veil.color

veil.color_w	veil.color_y	veil.color_e	veil.color_n	veil.color_u	veil.color_k
9.25	1.16	0.58	1.16	0.58	0.58

\$ring.type

ring.type_g	ring.type_p	ring.type_e	ring.type_l	ring.type_f	ring.type_m
2.89	2.89	4.62	3.47	79.19	0.58
ring.type_r	ring.type_z				
2.89	3.47				

\$spore.print.color

spore.print.color_w	spore.print.color_p	spore.print.color_k	spore.print.color_r
2.31	2.31	4.05	0.58
spore.print.color_u	spore.print.color_n	spore.print.color_g	
0.58	1.73	0.58	

\$habitat

habitat_d	habitat_m	habitat_g	habitat_h	habitat_l	habitat_p	habitat_w	habitat_u
87.28	9.83	21.97	7.51	10.40	1.16	0.58	0.58

\$season

season_u	season_a	season_w	season_s
80.92	97.11	23.70	13.29

簡述此資料:

1. 此資料有173個樣本,23個變數,其中將一個樣本的一個變項有重複的level的這些變項,將這些變項的level拆成單獨的變項後,因此報表呈現有114個variable。

2. 有重複level的佔比:有重複的level的變項有cap.shape, cap.surface, cap.color, gill.attachment, gill.color, stem.surface, stem.color, veil.color, ring.type, spore.print.color, habitat, season, 單獨去計算在173個樣本中出現幾次,如cap.shape中x在173個菇中有110個,佔63.58%。
3. 有毒的菇較多佔55.5%。
4. 平均莖長6.59cm,莖寬12.16mm,蕈帽直徑6.739cm。

Question 2:

Create a Table one with a single column (variable name: class)

```
library(table1)
final_data <- final_data %>%
  mutate(across(7:114, ~ replace(., is.na(.), "None")))
final_data <- final_data %>%
  mutate(across(7:11, ~ replace(., . == "", "Unknow")))
final_data <- final_data %>%
  mutate(across(7:114, as.factor))
data_filter <- final_data %>% select(-name, -family)
table1(~.|class,data = data_filter)
```

	e	p	Overall
	(N=77)	(N=96)	(N=173)
cap.diameter			
Mean (SD)	7.81 (6.26)	5.88 (3.85)	6.74 (5.14)
Median [Min, Max]	6.50 [1.00, 50.0]	5.00 [0.700, 19.0]	6.00 [0.700, 50.0]
stem.height			
Mean (SD)	7.05 (3.48)	6.22 (3.05)	6.59 (3.26)
Median [Min, Max]	6.00 [2.50, 25.0]	5.50 [0, 17.5]	6.00 [0, 25.0]
stem.width			
Mean (SD)	14.4 (10.8)	10.4 (8.66)	12.2 (9.86)
Median [Min, Max]	12.5 [1.00, 70.0]	7.50 [0, 40.0]	10.0 [0, 70.0]
does.bruise.or.bleed			
[f]	63 (81.8%)	80 (83.3%)	143 (82.7%)
[t]	14 (18.2%)	16 (16.7%)	30 (17.3%)
gill.spacing			
[c]	29 (37.7%)	41 (42.7%)	70 (40.5%)
[d]	13 (16.9%)	9 (9.4%)	22 (12.7%)
[f]	4 (5.2%)	6 (6.3%)	10 (5.8%)
Unknow	31 (40.3%)	40 (41.7%)	71 (41.0%)
stem.root			
[b]	6 (7.8%)	3 (3.1%)	9 (5.2%)
[c]	0 (0%)	2 (2.1%)	2 (1.2%)
[f]	0 (0%)	3 (3.1%)	3 (1.7%)
[r]	0 (0%)	4 (4.2%)	4 (2.3%)
[s]	4 (5.2%)	5 (5.2%)	9 (5.2%)
Unknow	67 (87.0%)	79 (82.3%)	146 (84.4%)
veil.type			
[u]	3 (3.9%)	6 (6.3%)	9 (5.2%)
Unknow	74 (96.1%)	90 (93.8%)	164 (94.8%)
has.ring			

	e	p	Overall
[f]	60 (77.9%)	70 (72.9%)	130 (75.1%)
[t]	17 (22.1%)	26 (27.1%)	43 (24.9%)
cap.shape_x			
None	23 (29.9%)	40 (41.7%)	63 (36.4%)
x	54 (70.1%)	56 (58.3%)	110 (63.6%)
cap.shape_f			
f	36 (46.8%)	38 (39.6%)	74 (42.8%)
None	41 (53.2%)	58 (60.4%)	99 (57.2%)
cap.shape_p			
None	67 (87.0%)	91 (94.8%)	158 (91.3%)
p	10 (13.0%)	5 (5.2%)	15 (8.7%)
cap.shape_b			
b	5 (6.5%)	18 (18.8%)	23 (13.3%)
None	72 (93.5%)	78 (81.3%)	150 (86.7%)
cap.shape_c			
c	4 (5.2%)	4 (4.2%)	8 (4.6%)
None	73 (94.8%)	92 (95.8%)	165 (95.4%)
cap.shape_s			
None	60 (77.9%)	77 (80.2%)	137 (79.2%)
s	17 (22.1%)	19 (19.8%)	36 (20.8%)
cap.shape_o			
None	73 (94.8%)	88 (91.7%)	161 (93.1%)
o	4 (5.2%)	8 (8.3%)	12 (6.9%)
cap.surface_g			
g	7 (9.1%)	9 (9.4%)	16 (9.2%)
None	70 (90.9%)	87 (90.6%)	157 (90.8%)
cap.surface_h			
h	13 (16.9%)	13 (13.5%)	26 (15.0%)
None	64 (83.1%)	83 (86.5%)	147 (85.0%)
cap.surface_t			
None	62 (80.5%)	74 (77.1%)	136 (78.6%)
t	15 (19.5%)	22 (22.9%)	37 (21.4%)
cap.surface_y			
None	65 (84.4%)	85 (88.5%)	150 (86.7%)
y	12 (15.6%)	11 (11.5%)	23 (13.3%)
cap.surface_e			
e	4 (5.2%)	7 (7.3%)	11 (6.4%)
None	73 (94.8%)	89 (92.7%)	162 (93.6%)
cap.surface_s			
None	59 (76.6%)	81 (84.4%)	140 (80.9%)
s	18 (23.4%)	15 (15.6%)	33 (19.1%)
cap.surface_l			
l	2 (2.6%)	2 (2.1%)	4 (2.3%)
None	75 (97.4%)	94 (97.9%)	169 (97.7%)
cap.surface_d			
d	8 (10.4%)	10 (10.4%)	18 (10.4%)
None	69 (89.6%)	86 (89.6%)	155 (89.6%)
cap.surface_w			
None	74 (96.1%)	91 (94.8%)	165 (95.4%)
w	3 (3.9%)	5 (5.2%)	8 (4.6%)
cap.surface_i			
i	2 (2.6%)	7 (7.3%)	9 (5.2%)

	e	p	Overall
None	75 (97.4%)	89 (92.7%)	164 (94.8%)
cap.surface_k	1 (1.3%)	9 (9.4%)	10 (5.8%)
None	76 (98.7%)	87 (90.6%)	163 (94.2%)
cap.color_e	7 (9.1%)	18 (18.8%)	25 (14.5%)
None	70 (90.9%)	78 (81.3%)	148 (85.5%)
cap.color_o	70 (90.9%)	81 (84.4%)	151 (87.3%)
None	7 (9.1%)	15 (15.6%)	22 (12.7%)
cap.color_n	53 (68.8%)	57 (59.4%)	110 (63.6%)
None	24 (31.2%)	39 (40.6%)	63 (36.4%)
cap.color_g	14 (18.2%)	14 (14.6%)	28 (16.2%)
None	63 (81.8%)	82 (85.4%)	145 (83.8%)
cap.color_r	75 (97.4%)	85 (88.5%)	160 (92.5%)
None	2 (2.6%)	11 (11.5%)	13 (7.5%)
cap.color_w	60 (77.9%)	78 (81.3%)	138 (79.8%)
None	17 (22.1%)	18 (18.8%)	35 (20.2%)
cap.color_y	61 (79.2%)	68 (70.8%)	129 (74.6%)
None	16 (20.8%)	28 (29.2%)	44 (25.4%)
cap.color_p	73 (94.8%)	89 (92.7%)	162 (93.6%)
None	4 (5.2%)	7 (7.3%)	11 (6.4%)
cap.color_b	5 (6.5%)	2 (2.1%)	7 (4.0%)
None	72 (93.5%)	94 (97.9%)	166 (96.0%)
cap.color_u	72 (93.5%)	91 (94.8%)	163 (94.2%)
None	5 (6.5%)	5 (5.2%)	10 (5.8%)
cap.color_l	4 (5.2%)	2 (2.1%)	6 (3.5%)
None	73 (94.8%)	94 (97.9%)	167 (96.5%)
cap.color_k	3 (3.9%)	6 (6.3%)	9 (5.2%)
None	74 (96.1%)	90 (93.8%)	164 (94.8%)
gill.attachment_e	10 (13.0%)	6 (6.3%)	16 (9.2%)
None	67 (87.0%)	90 (93.8%)	157 (90.8%)
gill.attachment_a	16 (20.8%)	24 (25.0%)	40 (23.1%)
None	61 (79.2%)	72 (75.0%)	133 (76.9%)
gill.attachment_d	14 (18.2%)	19 (19.8%)	33 (19.1%)
None	63 (81.8%)	77 (80.2%)	140 (80.9%)
gill.attachment_s	70 (90.9%)	87 (90.6%)	157 (90.8%)
None	7 (9.1%)	9 (9.4%)	16 (9.2%)

	e	p	Overall
gill.attachment_x			
None	68 (88.3%)	84 (87.5%)	152 (87.9%)
x	9 (11.7%)	12 (12.5%)	21 (12.1%)
gill.attachment_p			
None	65 (84.4%)	91 (94.8%)	156 (90.2%)
p	12 (15.6%)	5 (5.2%)	17 (9.8%)
gill.attachment_f			
f	4 (5.2%)	6 (6.3%)	10 (5.8%)
None	73 (94.8%)	90 (93.8%)	163 (94.2%)
gill.color_w			
None	39 (50.6%)	61 (63.5%)	100 (57.8%)
w	38 (49.4%)	35 (36.5%)	73 (42.2%)
gill.color_n			
n	15 (19.5%)	32 (33.3%)	47 (27.2%)
None	62 (80.5%)	64 (66.7%)	126 (72.8%)
gill.color_p			
None	65 (84.4%)	80 (83.3%)	145 (83.8%)
p	12 (15.6%)	16 (16.7%)	28 (16.2%)
gill.color_u			
None	74 (96.1%)	92 (95.8%)	166 (96.0%)
u	3 (3.9%)	4 (4.2%)	7 (4.0%)
gill.color_b			
b	3 (3.9%)	2 (2.1%)	5 (2.9%)
None	74 (96.1%)	94 (97.9%)	168 (97.1%)
gill.color_g			
g	10 (13.0%)	13 (13.5%)	23 (13.3%)
None	67 (87.0%)	83 (86.5%)	150 (86.7%)
gill.color_y			
None	60 (77.9%)	69 (71.9%)	129 (74.6%)
y	17 (22.1%)	27 (28.1%)	44 (25.4%)
gill.color_r			
None	75 (97.4%)	90 (93.8%)	165 (95.4%)
r	2 (2.6%)	6 (6.3%)	8 (4.6%)
gill.color_e			
e	2 (2.6%)	4 (4.2%)	6 (3.5%)
None	75 (97.4%)	92 (95.8%)	167 (96.5%)
gill.color_o			
None	72 (93.5%)	88 (91.7%)	160 (92.5%)
o	5 (6.5%)	8 (8.3%)	13 (7.5%)
gill.color_k			
k	6 (7.8%)	9 (9.4%)	15 (8.7%)
None	71 (92.2%)	87 (90.6%)	158 (91.3%)
gill.color_f			
f	4 (5.2%)	6 (6.3%)	10 (5.8%)
None	73 (94.8%)	90 (93.8%)	163 (94.2%)
stem.surface_y			
None	72 (93.5%)	86 (89.6%)	158 (91.3%)
y	5 (6.5%)	10 (10.4%)	15 (8.7%)
stem.surface_s			
None	66 (85.7%)	88 (91.7%)	154 (89.0%)
s	11 (14.3%)	8 (8.3%)	19 (11.0%)
stem.surface_k			

	e	p	Overall
k	2 (2.6%)	3 (3.1%)	5 (2.9%)
None	75 (97.4%)	93 (96.9%)	168 (97.1%)
stem.surface_i			
i	5 (6.5%)	9 (9.4%)	14 (8.1%)
None	72 (93.5%)	87 (90.6%)	159 (91.9%)
stem.surface_h			
h	0 (0%)	2 (2.1%)	2 (1.2%)
None	77 (100%)	94 (97.9%)	171 (98.8%)
stem.surface_t			
None	73 (94.8%)	92 (95.8%)	165 (95.4%)
t	4 (5.2%)	4 (4.2%)	8 (4.6%)
stem.surface_g			
g	0 (0%)	5 (5.2%)	5 (2.9%)
None	77 (100%)	91 (94.8%)	168 (97.1%)
stem.surface_f			
f	0 (0%)	3 (3.1%)	3 (1.7%)
None	77 (100%)	93 (96.9%)	170 (98.3%)
stem.color_w			
None	34 (44.2%)	65 (67.7%)	99 (57.2%)
w	43 (55.8%)	31 (32.3%)	74 (42.8%)
stem.color_y			
None	68 (88.3%)	73 (76.0%)	141 (81.5%)
y	9 (11.7%)	23 (24.0%)	32 (18.5%)
stem.color_n			
n	27 (35.1%)	43 (44.8%)	70 (40.5%)
None	50 (64.9%)	53 (55.2%)	103 (59.5%)
stem.color_b			
b	1 (1.3%)	0 (0%)	1 (0.6%)
None	76 (98.7%)	96 (100%)	172 (99.4%)
stem.color_u			
None	75 (97.4%)	91 (94.8%)	166 (96.0%)
u	2 (2.6%)	5 (5.2%)	7 (4.0%)
stem.color_l			
l	1 (1.3%)	1 (1.0%)	2 (1.2%)
None	76 (98.7%)	95 (99.0%)	171 (98.8%)
stem.color_r			
None	76 (98.7%)	93 (96.9%)	169 (97.7%)
r	1 (1.3%)	3 (3.1%)	4 (2.3%)
stem.color_p			
None	76 (98.7%)	93 (96.9%)	169 (97.7%)
p	1 (1.3%)	3 (3.1%)	4 (2.3%)
stem.color_e			
e	3 (3.9%)	8 (8.3%)	11 (6.4%)
None	74 (96.1%)	88 (91.7%)	162 (93.6%)
stem.color_k			
k	1 (1.3%)	3 (3.1%)	4 (2.3%)
None	76 (98.7%)	93 (96.9%)	169 (97.7%)
stem.color_g			
g	7 (9.1%)	7 (7.3%)	14 (8.1%)
None	70 (90.9%)	89 (92.7%)	159 (91.9%)
stem.color_o			
None	72 (93.5%)	89 (92.7%)	161 (93.1%)

	e	p	Overall
o	5 (6.5%)	7 (7.3%)	12 (6.9%)
stem.color_f			
f	0 (0%)	3 (3.1%)	3 (1.7%)
None	77 (100%)	93 (96.9%)	170 (98.3%)
veil.color_w			
None	69 (89.6%)	88 (91.7%)	157 (90.8%)
w	8 (10.4%)	8 (8.3%)	16 (9.2%)
veil.color_y			
None	75 (97.4%)	96 (100%)	171 (98.8%)
y	2 (2.6%)	0 (0%)	2 (1.2%)
veil.color_e			
e	0 (0%)	1 (1.0%)	1 (0.6%)
None	77 (100%)	95 (99.0%)	172 (99.4%)
veil.color_n			
n	0 (0%)	2 (2.1%)	2 (1.2%)
None	77 (100%)	94 (97.9%)	171 (98.8%)
veil.color_u			
None	77 (100%)	95 (99.0%)	172 (99.4%)
u	0 (0%)	1 (1.0%)	1 (0.6%)
veil.color_k			
k	0 (0%)	1 (1.0%)	1 (0.6%)
None	77 (100%)	95 (99.0%)	172 (99.4%)
ring.type_g			
g	2 (2.6%)	3 (3.1%)	5 (2.9%)
None	75 (97.4%)	93 (96.9%)	168 (97.1%)
ring.type_p			
None	75 (97.4%)	93 (96.9%)	168 (97.1%)
p	2 (2.6%)	3 (3.1%)	5 (2.9%)
ring.type_e			
e	3 (3.9%)	5 (5.2%)	8 (4.6%)
None	74 (96.1%)	91 (94.8%)	165 (95.4%)
ring.type_l			
l	4 (5.2%)	2 (2.1%)	6 (3.5%)
None	73 (94.8%)	94 (97.9%)	167 (96.5%)
ring.type_f			
f	61 (79.2%)	76 (79.2%)	137 (79.2%)
None	16 (20.8%)	20 (20.8%)	36 (20.8%)
ring.type_m			
m	1 (1.3%)	0 (0%)	1 (0.6%)
None	76 (98.7%)	96 (100%)	172 (99.4%)
ring.type_r			
None	74 (96.1%)	94 (97.9%)	168 (97.1%)
r	3 (3.9%)	2 (2.1%)	5 (2.9%)
ring.type_z			
None	77 (100%)	90 (93.8%)	167 (96.5%)
z	0 (0%)	6 (6.3%)	6 (3.5%)
spore.print.color_w			
None	75 (97.4%)	94 (97.9%)	169 (97.7%)
w	2 (2.6%)	2 (2.1%)	4 (2.3%)
spore.print.color_p			
None	76 (98.7%)	93 (96.9%)	169 (97.7%)
p	1 (1.3%)	3 (3.1%)	4 (2.3%)

	e	p	Overall
spore.print.color_k			
k	1 (1.3%)	6 (6.3%)	7 (4.0%)
None	76 (98.7%)	90 (93.8%)	166 (96.0%)
spore.print.color_r			
None	77 (100%)	95 (99.0%)	172 (99.4%)
r	0 (0%)	1 (1.0%)	1 (0.6%)
spore.print.color_u			
None	77 (100%)	95 (99.0%)	172 (99.4%)
u	0 (0%)	1 (1.0%)	1 (0.6%)
spore.print.color_n			
n	0 (0%)	3 (3.1%)	3 (1.7%)
None	77 (100%)	93 (96.9%)	170 (98.3%)
spore.print.color_g			
g	1 (1.3%)	0 (0%)	1 (0.6%)
None	76 (98.7%)	96 (100%)	172 (99.4%)
habitat_d			
d	69 (89.6%)	82 (85.4%)	151 (87.3%)
None	8 (10.4%)	14 (14.6%)	22 (12.7%)
habitat_m			
m	8 (10.4%)	9 (9.4%)	17 (9.8%)
None	69 (89.6%)	87 (90.6%)	156 (90.2%)
habitat_g			
g	15 (19.5%)	23 (24.0%)	38 (22.0%)
None	62 (80.5%)	73 (76.0%)	135 (78.0%)
habitat_h			
h	5 (6.5%)	8 (8.3%)	13 (7.5%)
None	72 (93.5%)	88 (91.7%)	160 (92.5%)
habitat_l			
l	11 (14.3%)	7 (7.3%)	18 (10.4%)
None	66 (85.7%)	89 (92.7%)	155 (89.6%)
habitat_p			
None	77 (100%)	94 (97.9%)	171 (98.8%)
p	0 (0%)	2 (2.1%)	2 (1.2%)
habitat_w			
None	76 (98.7%)	96 (100%)	172 (99.4%)
w	1 (1.3%)	0 (0%)	1 (0.6%)
habitat_u			
None	76 (98.7%)	96 (100%)	172 (99.4%)
u	1 (1.3%)	0 (0%)	1 (0.6%)
season_u			
None	16 (20.8%)	17 (17.7%)	33 (19.1%)
u	61 (79.2%)	79 (82.3%)	140 (80.9%)
season_a			
a	74 (96.1%)	94 (97.9%)	168 (97.1%)
None	3 (3.9%)	2 (2.1%)	5 (2.9%)
season_w			
None	52 (67.5%)	80 (83.3%)	132 (76.3%)
w	25 (32.5%)	16 (16.7%)	41 (23.7%)
season_s			
None	65 (84.4%)	85 (88.5%)	150 (86.7%)
s	12 (15.6%)	11 (11.5%)	23 (13.3%)

簡述table1:

1. 連續變項:顯示有毒的菇和沒毒的菇那個變項的平均值、中位數、最大值、最小值。
2. 類別變項中,顯示有毒的菇和沒毒的菇那個變項裡每個level的各數及占比,如:gill.spacing變項裡,有c的有41個(42.7%),有d的有9個(9.4%),有f的有6個(6.3%),未知的" Unknown" 有40個佔41%,
3. 有多個重複level的類別變項,將每個level分開後的解讀,如:cap.shape_x的意思為,在有毒的菇中有56個觀察到x這個特徵佔58.3%・沒有x則有40個佔41.7%;而沒毒的菇中有54個觀察到x這個特徵佔70.1%・沒有x則有23個佔29.9%