HW2

Summary Report for Mushroom Dataset

Chun-Ju Chou

2025-03-18

目錄

Data information

This dataset was compiled by Dennis Wagner on 05 September 2020. It includes 173 species of mushrooms with caps from various families and one entry for each species. Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended (the latter class was combined with the poisonous class). Of the 20 variables, 17 are nominal and 3 are metrical. The values of each nominal variable are a set of possible values and for the metrical variables a range of possible values.

Variable Definition

Variable (Data Type)	Definition
family (multinomial)	String of the name of the family of mushroom species
name (multinomial)	String of the of the mushroom species
class (binary)	poisonous=p, edibile=e
cap-diameter (metrical)	float number(s) in cm, two values=min max, one value=mean
cap-shape (nominal	bell=b, conical=c, convex=x, flat=f, sunken=s, spherical=p, others=o
cap-surface (nominal)	fibrous=i, grooves=g, scaly=y, smooth=s,shiny=h, leathery=l, silky=k, sticky=t, wrinkled=w, fleshy=e
cap-color (nominal)	brown=n, buff=b, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y, blue=l, orange=o, black=k
does-bruise-bleed (nominal)	bruises-or-bleeding=t,no=f
gill-attachment (nominal)	adnate=a, adnexed=x, decurrent=d, free=e, sinuate=s, pores=p, none=f, unknown=?
gill-spacing (nominal)	close=c, distant=d, none=f
gill-color (nominal)	see cap-color + none=f
stem-height (metrical)	float number(s) in cm, two values=min max, one value=mean
stem-width (metrical)	=bulbous=b, swollen=s, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r
stem-surface (nominal)	see cap-surface + none=f

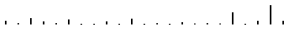
Variable (Data Type)	Definition
stem-color (nominal)	see cap-color + none=f
veil-type (nominal)	partial=p, universal=u
veil-color (nominal)	see cap-color + none=f
has-ring (nominal)	ring=t, none=f
ring-type (nominal)	cobwebby=c, evanescent=e, flaring=r, grooved=g, large=l, pendant=p, sheathing=s, zone=z, scaly=y, movable=m, none=f, unknown=?
spore-print-color (nominal)	see cap color
habitat (nominal)	grasses=g, leaves=l, meadows=m, paths=p, heaths=h, urban=u, waste=w, woods=d
season (nominal)	spring=s, summer=u, autumn=a, winter=w

Data Description

```
# R Interface to Python
library(reticulate)
library(Hmisc)
mushroom <- read.csv("primary_data.csv", sep=';')
latex(describe(mushroom), file="")
```

mushroom
23 Variables 173 Observations

family



n	missing	distinct
173	0	23

```
lowest : Amanita Family     Bolbitius Family   Bolete Family      Bracket Fungi      Chanterelle Family
highest: Russula Family     Saddle-Cup Family  Stropharia Family  Tricholoma Family  Wax Gill Family
```

name

n	missing	distinct
173	0	173

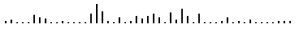
```
lowest : Amethyst Deceiver        Aniseed Funnel Cap       Apricot Fungus           Bare-toothed Russula     Bay Bolete
highest: Yellow-gilled Russula    Yellow-staining Mushroom Yellow-stemmed Bell Cap   Yellow Swamp Russula     Yellow Wax cap
```

class

n	missing	distinct
173	0	2

Value	e	p
Frequency	77	96
Proportion	0.445	0.555

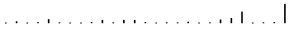
cap.diameter



n	missing	distinct
173	0	51

```
lowest : [0.4, 1]   [0.5, 1.5] [0.5, 1]   [0.7, 1.3] [1, 1.5]
highest: [8, 14]    [8, 15]    [8, 20]    [8, 25]    [8, 30]
```

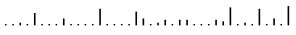
cap.shape



n	missing	distinct
173	0	27

```
lowest : [b, f, s] [b, f]    [b, x, f] [b, x]    [b]
highest: [x, f]    [x, o]    [x, p]    [x, s]    [x]
```

Cap.surface

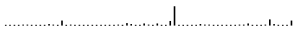


n	missing	distinct
133	40	40

```
lowest : [d, e, y, i] [d, k, s]    [d, k]       [d, s]       [d]
highest: [t]          [w, t]       [w]          [y, s]       [y]
```

---

cap.color



n	missing	distinct
173	0	67

```
lowest : [b, p, e, y]    [b, u]          [b]             [e, n, p, w]    [e, n, y]
highest: [y, n]          [y, o, g, n, r] [y, o, r, n]    [y, o]          [y]
```

---

does.bruise.or.bleed

n	missing	distinct
173	0	2

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

---

gill.attachment

n	missing	distinct
145	28	8

Value	[a, d]	[a]	[d]	[e]	[f]	[p]	[s]	[x]
Frequency	8	32	25	16	10	17	16	21
Proportion	0.055	0.221	0.172	0.110	0.069	0.117	0.110	0.145

---

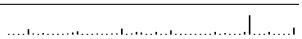
gill.spacing

n	missing	distinct
102	71	3

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

---

gill.color



n	missing	distinct
173	0	59

```
lowest : [b, p, w] [b, u]    [b]       [e]       [f]
highest: [y, o, e] [y, r, k] [y, r]    [y, w]    [y]
```

---

stem.height

n	missing	distinct
173	0	46

```
lowest : [0]      [1, 2]   [1, 3]   [10, 12] [10, 15], highest: [8, 12]  [8, 15]  [8, 20]  [8, 25]  [8, 30]
```

---

stem.width

n	missing	distinct
173	0	48

```
lowest : [0.5, 1] [0]      [1, 2]   [1, 3]   [1]     , highest: [7, 15]  [8, 12]  [8, 15]  [8, 18]  [8, 20]
```

---

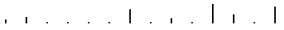
stem.root

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

---

stem.surface



n	missing	distinct
65	108	14

Value	[f]	[g]	[h]	[i, s]	[i, t]	[i, y]	[i]	[k, s]	[k]	[s, h]	[s]	[t]
Frequency	3	5	1	1	1	1	11	1	4	1	15	7
Proportion	0.046	0.077	0.015	0.015	0.015	0.015	0.169	0.015	0.062	0.015	0.231	0.108

Value	[y, s]	[y]
Frequency	1	13
Proportion	0.015	0.200

stem.color

n	missing	distinct
173	0	41

lowest :	[b, u]	[e, n]	[e, u, y]	[e, y]	[e]
highest:	[w]	[y, e, n]	[y, n]	[y, o, k]	[y]

veil.type

n	missing	distinct	value
9	164	1	[u]

Value	[u]
Frequency	9
Proportion	1

veil.color

n	missing	distinct
21	152	7

Value	[e, n]	[k]	[n]	[u]	[w]	[y, w]	[y]
Frequency	1	1	1	1	15	1	1
Proportion	0.048	0.048	0.048	0.048	0.714	0.048	0.048

has.ring

n	missing	distinct
173	0	2

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

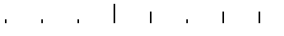
ring.type

n	missing	distinct
166	7	13

Value	[e, g]	[e]	[f]	[g, p]	[g]	[l, e]	[l, p]	[l, r]	[l]	[m]	[p]	[r]
Frequency	1	6	137	2	2	1	1	2	2	1	2	3
Proportion	0.006	0.036	0.825	0.012	0.012	0.006	0.006	0.012	0.012	0.006	0.012	0.018

Value	[z]
Frequency	6
Proportion	0.036

Spore.print.color



n	missing	distinct
18	155	8

Value	[g]	[k, r]	[k, u]	[k]	[n]	[p, w]	[p]	[w]
Frequency	1	1	1	5	3	1	3	3
Proportion	0.056	0.056	0.056	0.278	0.167	0.056	0.167	0.167

habitat

n	missing	distinct
173	0	21

lowest :	[d, h]	[d]	[g, d, h]	[g, d]	[g, h, d]
highest:	[m, d]	[m, h]	[m]	[p, d]	[w]

season

n	missing	distinct
173	0	10

Value	[a, w]	[a]	[s, a, w]	[s, u, a, w]	[s, u, a]	[s, u]
Frequency	15	16	1	13	5	3
Proportion	0.087	0.092	0.006	0.075	0.029	0.017

Value	[s]	[u, a, w]	[u, a]	[u]
Frequency	1	12	106	1
Proportion	0.006	0.069	0.613	0.006

Table One

```
library(table1)
table1(~ class, data = mushroom)
```

	Overall
	(N=173)
class	
e	77 (44.5%)
p	96 (55.5%)