Contraceptive method choice (CMC)

Attribute Information:

1. Wife's age (numerical)  
2. Wife's education (categorical) 1=low, 2, 3, 4=high  
3. Husband's education (categorical) 1=low, 2, 3, 4=high  
4. Number of children ever born (numerical)  
5. Wife's religion (binary) 0=Non-Islam, 1=Islam  
6. Wife's now working? (binary) 0=Yes, 1=No  
7. Husband's occupation (categorical) 1, 2, 3, 4  
8. Standard-of-living index (categorical) 1=low, 2, 3, 4=high  
9. Media exposure (binary) 0=Good, 1=Not good  
10. Contraceptive method used (class attribute) 1=No-use, 2=Long-term, 3=Short-term

Audiology

Number of instances: 200 training cases, 26 test cases

6. Number of attributes: 69 + identifier attribute + class attribute

7. Attribute information:

age\_gt\_60: f, t.

air(): mild,moderate,severe,normal,profound.

airBoneGap: f, t.

ar\_c(): normal,elevated,absent.

ar\_u(): normal,absent,elevated.

bone(): mild,moderate,normal,unmeasured.

boneAbnormal: f, t.

bser(): normal,degraded.

history\_buzzing: f, t.

history\_dizziness: f, t.

history\_fluctuating: f, t.

history\_fullness: f, t.

history\_heredity: f, t.

history\_nausea: f, t.

history\_noise: f, t.

history\_recruitment: f, t.

history\_ringing: f, t.

history\_roaring: f, t.

history\_vomiting: f, t.

late\_wave\_poor: f, t.

m\_at\_2k: f, t.

m\_cond\_lt\_1k: f, t.

m\_gt\_1k: f, t.

m\_m\_gt\_2k: f, t.

m\_m\_sn: f, t.

m\_m\_sn\_gt\_1k: f, t.

m\_m\_sn\_gt\_2k: f, t.

m\_m\_sn\_gt\_500: f, t.

m\_p\_sn\_gt\_2k: f, t.

m\_s\_gt\_500: f, t.

m\_s\_sn: f, t.

m\_s\_sn\_gt\_1k: f, t.

m\_s\_sn\_gt\_2k: f, t.

m\_s\_sn\_gt\_3k: f, t.

m\_s\_sn\_gt\_4k: f, t.

m\_sn\_2\_3k: f, t.

m\_sn\_gt\_1k: f, t.

m\_sn\_gt\_2k: f, t.

m\_sn\_gt\_3k: f, t.

m\_sn\_gt\_4k: f, t.

m\_sn\_gt\_500: f, t.

m\_sn\_gt\_6k: f, t.

m\_sn\_lt\_1k: f, t.

m\_sn\_lt\_2k: f, t.

m\_sn\_lt\_3k: f, t.

middle\_wave\_poor: f, t.

mod\_gt\_4k: f, t.

mod\_mixed: f, t.

mod\_s\_mixed: f, t.

mod\_s\_sn\_gt\_500: f, t.

mod\_sn: f, t.

mod\_sn\_gt\_1k: f, t.

mod\_sn\_gt\_2k: f, t.

mod\_sn\_gt\_3k: f, t.

mod\_sn\_gt\_4k: f, t.

mod\_sn\_gt\_500: f, t.

notch\_4k: f, t.

notch\_at\_4k: f, t.

o\_ar\_c(): normal,elevated,absent.

o\_ar\_u(): normal,absent,elevated.

s\_sn\_gt\_1k: f, t.

s\_sn\_gt\_2k: f, t.

s\_sn\_gt\_4k: f, t.

speech(): normal,good,very\_good,very\_poor,poor,unmeasured.

static\_normal: f, t.

tymp(): a,as,b,ad,c.

viith\_nerve\_signs: f, t.

wave\_V\_delayed: f, t.

waveform\_ItoV\_prolonged: f, t.

indentifier (unique for each instance)

class: cochlear\_unknown,mixed\_cochlear\_age\_fixation,poss\_central

mixed\_cochlear\_age\_otitis\_media,mixed\_poss\_noise\_om,

cochlear\_age,normal\_ear,cochlear\_poss\_noise,cochlear\_age\_and\_noise,

acoustic\_neuroma,mixed\_cochlear\_unk\_ser\_om,conductive\_discontinuity,

retrocochlear\_unknown,conductive\_fixation,bells\_palsy,

cochlear\_noise\_and\_heredity,mixed\_cochlear\_unk\_fixation,

otitis\_media,possible\_menieres,possible\_brainstem\_disorder,

cochlear\_age\_plus\_poss\_menieres,mixed\_cochlear\_age\_s\_om,

mixed\_cochlear\_unk\_discontinuity,mixed\_poss\_central\_om

8. Missing attributes: Yes

9. Class Distribution: (in the training set)

1. acoustic\_neuroma: 1

2. bells\_palsy: 1

3. cochlear\_age: 46

4. cochlear\_age\_and\_noise: 18

5. cochlear\_age\_plus\_poss\_menieres: 1

6. cochlear\_noise\_and\_heredity: 2

7. cochlear\_poss\_noise: 16

8. cochlear\_unknown: 48

9. conductive\_discontinuity: 2

10. conductive\_fixation: 6

11. mixed\_cochlear\_age\_fixation: 1

12. mixed\_cochlear\_age\_otitis\_media: 4

13. mixed\_cochlear\_age\_s\_om: 2

14. mixed\_cochlear\_unk\_discontinuity: 2

15. mixed\_cochlear\_unk\_fixation: 5

16. mixed\_cochlear\_unk\_ser\_om: 3

17. mixed\_poss\_central\_om: 1

18. mixed\_poss\_noise\_om: 2

19. normal\_ear: 20

20. otitis\_media: 4

21. poss\_central: 1

22. possible\_brainstem\_disorder: 4

23. possible\_menieres: 8

24. retrocochlear\_unknown: 2

Soybean Large

**Attribute Information:**

-- 19 Classes  
diaporthe-stem-canker, charcoal-rot, rhizoctonia-root-rot,  
phytophthora-rot, brown-stem-rot, powdery-mildew,  
downy-mildew, brown-spot, bacterial-blight,  
bacterial-pustule, purple-seed-stain, anthracnose,  
phyllosticta-leaf-spot, alternarialeaf-spot,  
frog-eye-leaf-spot, diaporthe-pod-&-stem-blight,  
cyst-nematode, 2-4-d-injury, herbicide-injury.  
  
1. date: april,may,june,july,august,september,october,?.  
2. plant-stand: normal,lt-normal,?.  
3. precip: lt-norm,norm,gt-norm,?.  
4. temp: lt-norm,norm,gt-norm,?.  
5. hail: yes,no,?.  
6. crop-hist: diff-lst-year,same-lst-yr,same-lst-two-yrs,  
same-lst-sev-yrs,?.  
7. area-damaged: scattered,low-areas,upper-areas,whole-field,?.  
8. severity: minor,pot-severe,severe,?.  
9. seed-tmt: none,fungicide,other,?.  
10. germination: 90-100%,80-89%,lt-80%,?.  
11. plant-growth: norm,abnorm,?.  
12. leaves: norm,abnorm.  
13. leafspots-halo: absent,yellow-halos,no-yellow-halos,?.  
14. leafspots-marg: w-s-marg,no-w-s-marg,dna,?.  
15. leafspot-size: lt-1/8,gt-1/8,dna,?.  
16. leaf-shread: absent,present,?.  
17. leaf-malf: absent,present,?.  
18. leaf-mild: absent,upper-surf,lower-surf,?.  
19. stem: norm,abnorm,?.  
20. lodging: yes,no,?.  
21. stem-cankers: absent,below-soil,above-soil,above-sec-nde,?.  
22. canker-lesion: dna,brown,dk-brown-blk,tan,?.  
23. fruiting-bodies: absent,present,?.  
24. external decay: absent,firm-and-dry,watery,?.  
25. mycelium: absent,present,?.  
26. int-discolor: none,brown,black,?.  
27. sclerotia: absent,present,?.  
28. fruit-pods: norm,diseased,few-present,dna,?.  
29. fruit spots: absent,colored,brown-w/blk-specks,distort,dna,?.  
30. seed: norm,abnorm,?.  
31. mold-growth: absent,present,?.  
32. seed-discolor: absent,present,?.  
33. seed-size: norm,lt-norm,?.  
34. shriveling: absent,present,?.  
35. roots: norm,rotted,galls-cysts,?.

Breast Cancer Wisconsin

**Attribute Information:**

1) ID number  
2) Diagnosis (M = malignant, B = benign)  
3-32)  
  
Ten real-valued features are computed for each cell nucleus:  
  
a) radius (mean of distances from center to points on the perimeter)  
b) texture (standard deviation of gray-scale values)  
c) perimeter  
d) area  
e) smoothness (local variation in radius lengths)  
f) compactness (perimeter^2 / area - 1.0)  
g) concavity (severity of concave portions of the contour)  
h) concave points (number of concave portions of the contour)  
i) symmetry  
j) fractal dimension ("coastline approximation" - 1)

Hayes Roth

**Attribute Information:**

-- 1. name: distinct for each instance and represented numerically  
-- 2. hobby: nominal values ranging between 1 and 3  
-- 3. age: nominal values ranging between 1 and 4  
-- 4. educational level: nominal values ranging between 1 and 4  
-- 5. marital status: nominal values ranging between 1 and 4  
-- 6. class: nominal value between 1 and 3

Nursery

Number of Instances: 12960

(instances completely cover the attribute space)

1. Number of Attributes: 8

2. Attribute Values:

parents usual, pretentious, great\_pret

has\_nurs proper, less\_proper, improper, critical, very\_crit

form complete, completed, incomplete, foster

children 1, 2, 3, more

housing convenient, less\_conv, critical

finance convenient, inconv

social non-prob, slightly\_prob, problematic

health recommended, priority, not\_recom

3. Missing Attribute Values: none

4. Class Distribution (number of instances per class)

class N N[%]

------------------------------

not\_recom 4320 (33.333 %)

recommend 2 ( 0.015 %)

very\_recom 328 ( 2.531 %)

priority 4266 (32.917 %)

spec\_prior 4044 (31.204 %)

Mushroom

**Attribute Information:**

Number of Instances: 8124

6. Number of Attributes: 22 (all nominally valued)

7. Attribute Information: (classes: edible=e, poisonous=p)

1. cap-shape: bell=b,conical=c,convex=x,flat=f,

knobbed=k,sunken=s

2. cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s

3. cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r,

pink=p,purple=u,red=e,white=w,yellow=y

4. bruises?: bruises=t,no=f

5. odor: almond=a,anise=l,creosote=c,fishy=y,foul=f,

musty=m,none=n,pungent=p,spicy=s

6. gill-attachment: attached=a,descending=d,free=f,notched=n

7. gill-spacing: close=c,crowded=w,distant=d

8. gill-size: broad=b,narrow=n

9. gill-color: black=k,brown=n,buff=b,chocolate=h,gray=g,

green=r,orange=o,pink=p,purple=u,red=e,

white=w,yellow=y

10. stalk-shape: enlarging=e,tapering=t

11. stalk-root: bulbous=b,club=c,cup=u,equal=e,

rhizomorphs=z,rooted=r,missing=?

12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s

13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s

14. stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,

pink=p,red=e,white=w,yellow=y

15. stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,

pink=p,red=e,white=w,yellow=y

16. veil-type: partial=p,universal=u

17. veil-color: brown=n,orange=o,white=w,yellow=y

18. ring-number: none=n,one=o,two=t

19. ring-type: cobwebby=c,evanescent=e,flaring=f,large=l,

none=n,pendant=p,sheathing=s,zone=z

20. spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r,

orange=o,purple=u,white=w,yellow=y

21. population: abundant=a,clustered=c,numerous=n,

scattered=s,several=v,solitary=y

22. habitat: grasses=g,leaves=l,meadows=m,paths=p,

urban=u,waste=w,woods=d

8. Missing Attribute Values: 2480 of them (denoted by "?"), all for

attribute #11.

9. Class Distribution:

-- edible: 4208 (51.8%)

-- poisonous: 3916 (48.2%)

-- total: 8124 instances

**Cancer**

Number of Instances: 699 (as of 15 July 1992)

1. Number of Attributes: 10 plus the class attribute

2. Attribute Information: (class attribute has been moved to last column)

# Attribute Domain

-- -----------------------------------------

1. Sample code number id number

2. Clump Thickness 1 - 10

3. Uniformity of Cell Size 1 - 10

4. Uniformity of Cell Shape 1 - 10

5. Marginal Adhesion 1 - 10

6. Single Epithelial Cell Size 1 - 10

7. Bare Nuclei 1 - 10

8. Bland Chromatin 1 - 10

9. Normal Nucleoli 1 - 10

10. Mitoses 1 - 10

11. Class: (2 for benign, 4 for malignant)

3. Missing attribute values: 16

There are 16 instances in Groups 1 to 6 that contain a single missing

(i.e., unavailable) attribute value, now denoted by "?".

4. Class distribution:

Benign: 458 (65.5%)

**Glass**

1. Number of Instances: 214

2. Number of Attributes: 10 (including an Id#) plus the class attribute

-- all attributes are continuously valued

3. Attribute Information:

1. Id number: 1 to 214

2. RI: refractive index

3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as

are attributes 4-10)

4. Mg: Magnesium

5. Al: Aluminum

6. Si: Silicon

7. K: Potassium

8. Ca: Calcium

9. Ba: Barium

10. Fe: Iron

4. Type of glass: (class attribute)

-- 1 building\_windows\_float\_processed

-- 2 building\_windows\_non\_float\_processed

-- 3 vehicle\_windows\_float\_processed

-- 4 vehicle\_windows\_non\_float\_processed (none in this database)

-- 5 containers

-- 6 tableware

-- 7 headlamps