第八章

8.6 Summary statistics

	n	mean	sum	stdev
m	25	27.12	678	7.4516
y_i	25	1398.3	34957	394.52
$y_i - \bar{y}m_i$	25	0	0	103.97

$N = 108$

By Equation (8.1), the estimate of the population mean μ is,

$$\bar{y} = \frac{\sum y_i}{\sum m_i} = \frac{34957}{678} = 51.56$$

Since M is not known, the $\overline{M}$ appearing in Equation (8.2) must be estimated by $\overline{m}$, where

$$\overline{m} = \frac{\sum m_i}{n} = \frac{678}{25} = 27.12$$

The from Equation (8.2), the estimated variance of population mean is,

$$\hat{V}(\bar{y}) = \left(\frac{N-n}{Nn\overline{M}^2}\right)\frac{1}{n-1}\sum(y_i - \bar{y}m_i)^2 = \left(\frac{N-n}{Nn\overline{M}^2}\right)s_r^2$$

$$= \left(\frac{108-25}{108(25)(27.12)^2}\right)(103.97)^2$$

with

$$B = 2\sqrt{\hat{V}(\bar{y})} = 1.344$$

8.7 From equation (8.2),

$$n = \frac{N\sigma_r^2}{ND + \sigma_r^2} \quad \text{where} \quad D = \frac{B^2\overline{M}^2}{4}$$

Here $N = 100$, $B = 2$, $\overline{m}$ is used to estimate $\overline{M}$ for D and s_r^2 is used to estimate σ_r^2. These values are obtained from Exercise (8.6).

Then

$$n = \frac{100(103.97)^2}{100(27.12)^2 + (103.97)^2} = 12.8 \approx 13$$

8.8 The following table is the data for Exercise 8.8 and the basic summary statistics. The $(a_i - \bar{a}m_i)$ column is established since the basic component of the estimated variance is the sample variance of these differences.

Cluster	m_i	a_i	$a_i - \bar{a}m_i$
1	51	42	5.83535
2	62	53	9.03513
3	49	40	5.25357
4	73	45	-6.76509
5	101	63	-8.62019
6	48	31	-3.03732
7	65	38	-8.09221
8	49	30	-4.74643
9	73	54	2.23491
10	61	45	1.74424
11	58	51	9.87157
12	52	29	-7.87376
13	65	46	-0.09221
14	49	37	2.25357
15	55	42	2.99890

Summary statistics

	n	mean	sum	stdev
m	15	60.73	911	14.007
a_i	15	43.07	646	9.5728
$a_i - \bar{a}m_i$	15	0	0	6.2230

$N = 87$

By Equation (8.16), the estimate of the population proportion p is,

$$\hat{p} = \frac{\sum a_i}{\sum m_i} = \frac{646}{911} = .709$$

Since M is not known, the $\overline{M}$ appearing in Equation (8.17) must be estimated by $\overline{m}$, where

$$\overline{m} = \frac{\sum m_i}{n} = \frac{911}{15} = 60.73$$

The from Equation (8.17), the estimated variance of population proportion is,

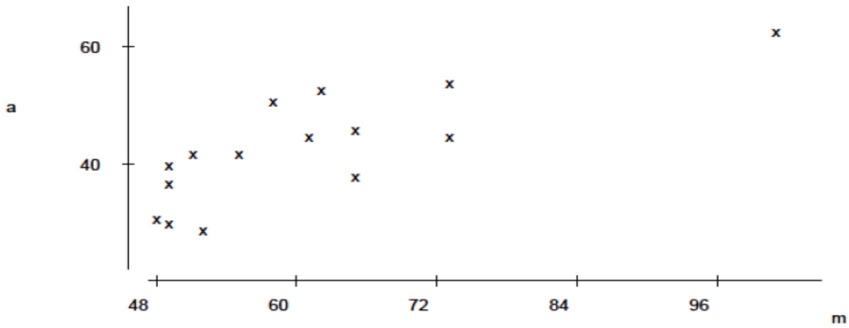
$$\hat{V}(\hat{p}) = \left(\frac{N-n}{Nn\overline{M}^2}\right)\frac{1}{n-1}\sum(a_i - \bar{a}m_i)^2 = \left(\frac{N-n}{Nn\overline{M}^2}\right)s_p^2$$

$$= \left(\frac{87-15}{87(15)(60.73)^2}\right)(6.223)^2$$

with

$$B = 2\sqrt{\hat{V}(\hat{p})} = 0.048$$

A plot of a versus m for the data is provided. Note that the linearity is fairly strong, so the estimate will have low bias potential. There is, however, one highly influential data point.



8.9 From section 8.7,

$$n = \frac{N\sigma_p^2}{ND + \sigma_p^2} \qquad \text{where} \qquad D = \frac{B^2\overline{M}^2}{4}$$

Here $N = 87$, $B = .08$, $\overline{m}$ is used to estimate $\overline{M}$ for D and s_p^2 is used to estimate σ_p^2. These values are obtained from Exercise 8.8.

Then

$$n = \frac{87(6.223)^2}{87\left[(.08)^2(60.73)^2/4\right] + (6.223)^2} = 6.1 \approx 7$$

8.16 $N = 48,\ n = 10$

$$\sum y_i = 736$$
$$\sum m_i = 365$$

$$\overline{y}_t = \frac{\sum y_i}{n} = \frac{736}{10} = 73.6$$
$$\hat{\tau} = N\overline{y}_t = 48(73.6) = 3532.8$$
$$\hat{V}(\hat{\tau}) = N^2\left(\frac{N-n}{Nn}\right)s_t^2$$
$$= 48^2\left(\frac{48-10}{48(10)}\right)(398.93)$$
$$B = 2\sqrt{\hat{V}(\hat{\tau})} = 539.50$$

8.17 $N = 175,\ n = 25$

m_i = number of elements in cluster i = 4 tires per cab

$\overline{M}$ = average cluster size = 4

$$\sum a_i = 40$$
$$\sum m_i = 100$$
$$\hat{p} = \frac{\sum a_i}{\sum m_i} = \frac{40}{100} = .4$$
$$\hat{V}(\hat{p}) = \left(\frac{N-n}{Nn\overline{M}^2}\right)s_p^2$$
$$= \left(\frac{175-25}{175(25)4^2}\right)(1.583)$$
$$B = 2\sqrt{\hat{V}(\hat{p})} = .116$$

8.19 Summary statistics

	Florida			California		
	n	mean	stdev	n	mean	stdev
m_i	8	13.625	6.05	10	14.4	8.83
y_{ti}	8	42.125	15.17	10	35.5	15.72
$y_{ti} - \overline{y}^* m_i$	8	5.54	4.8858	10	−3.17	16.946
	$N_1 = 80$			$N_2 = 140$		

An estimate of the average of sick leave per employee is then

$$\overline{y}^* = \frac{N_1\overline{y}_{t1} + N_2\overline{y}_{t2}}{N_1\overline{m}_1 + N_2\overline{m}_2} = \frac{80(42.125) + 140(35.5)}{80(13.625) + 140(14.4)} = 2.685$$

The variance of $\overline{y}^*$ is

$$\begin{aligned}\hat{V}(\overline{y}^*) &= \frac{1}{M^2}\left[\frac{N_1(N_1 - n_1)}{n_1}s_{r1}^2 + \frac{N_2(N_2 - n_2)}{n_2}s_{r2}^2\right] \\ &= \frac{1}{3106^2}\left(\frac{80(80-8)}{8}(4.8858)^2 + \frac{140(140-10)}{10}(16.946)^2\right) \\ &= .056\end{aligned}$$

where

$$M = M_1 + M_2 = N_1\overline{M}_1 + N_2\overline{M}_2 = 80(13.625) + 140(14.4) = 3106$$

$\overline{M}_1, \overline{M}_2$ are estimated by $\overline{m}_1, \overline{m}_2$ and

$$s_{ri}^2 = \frac{1}{n_i - 1}\sum\left(y_{ti} - \overline{y}^* m_i\right)^2$$

8.20 $N = 386$, $n = 20$

$y_i =$ total height of all trees in plot i.

$y_i = m_i\overline{y}_i$

$$\sum y_i = 6180.8$$

$$\sum m_i = 1046$$

$$\overline{y} = \frac{\sum y_i}{\sum m_i} = \frac{6180.8}{1046} = 5.91$$

$$\overline{M} = \frac{\sum m_i}{n} = \frac{1046}{20} = 52.3$$

$$\begin{aligned}\hat{V}(\overline{y}) &= \left(\frac{1}{n\overline{M}^2}\right)s_p^2 \\ &= \left(\frac{386 - 20}{386(20)(52.3)^2}\right)(1499.55)\end{aligned}$$

(Use $\overline{m}$ to estimate $\overline{M}$)

$$B = 2\sqrt{\hat{V}(\overline{y})} = .322$$

8.21 a_i = number of defective microchips on board i
m_i = number of microchips on board i (12 per board)

$$n = 10, \quad \overline{M} = 12$$
$$\sum a_i = 16$$
$$\sum m_i = 120$$
$$\hat{p} = \frac{\sum a_i}{\sum m_i} = \frac{16}{120} = .1333$$

By Equation (8.17), ignoring the fpc,

$$\hat{V}(\hat{p}) = \left(\frac{1}{n\overline{M}^2}\right)s_p^2$$
$$= \frac{1}{10(12^2)}(2.046)$$
$$B = 2\sqrt{\hat{V}(\hat{p})} = .075$$

8.22 $N = 50, \quad n = 10$

$$\hat{\tau} = N\frac{\sum a_i}{n} = 50\frac{16}{10} = 80$$
$$\hat{V}(\hat{\tau}) = N^2\left(\frac{N-n}{nN}\right)s_p^2$$
$$= (50)^2\left(\frac{50-10}{10(50)}\right)(2.046)$$
$$B = 2\sqrt{\hat{V}(\hat{\tau})} = 40.46$$

8.23 m_i = number of equipment items
a_i = number of items not properly identified

$$N = 15, \quad n = 5$$
$$\sum a_i = 9$$
$$\sum m_i = 98$$
$$\hat{p} = \frac{\sum a_i}{\sum m_i} = \frac{9}{98} = .0918 \qquad \overline{m} = \frac{\sum m_i}{n} = \frac{98}{5} = 19.6$$
$$\hat{V}(\hat{p}) = \left(\frac{N-n}{Nn\overline{M}^2}\right)s_p^2$$
$$= \left(\frac{15-5}{15(5)(19.6)^2}\right)(1.095)$$

$\overline{M}$ is estimated by $\overline{m}$)

$$B = 2\sqrt{\hat{V}(\hat{p})} = .039$$