Modal Summaries for Mapping

Table 5.1 Typical Frequency Distributions in Nominal, Ordinal, and Interval Scales

Nominal		Ordinal		Interval	
Crop	**Frequency**	**Hardiness**	**Frequency**	**Yields**	**Frequency**
Wheat	10	Most	12	400	18
Barley	8	Intermediate	15	300	22
Corn	6	Least	8	250	23
				175	15
				125	10

Table 5.2 Classed Frequency Distribution and Relative Frequency Table

Class Interval	Frequency	Relative Frequency
0–10,000	39	.204
10,000–20,000	51	.267
20,000–30,000	33	.172
30,000–40,000	25	.130
40,000–50,000	20	.104
50,000–60,000	18	.094
60,000–70,000	5	0.26
Total	191	1.000

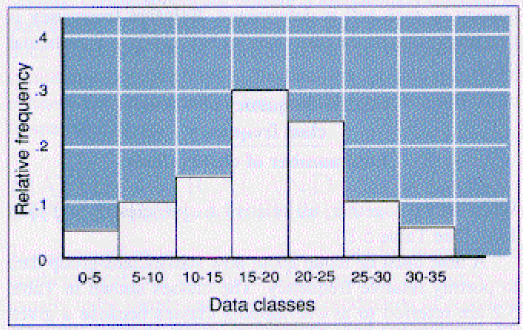


Figure 5.1 Relative frequency histogram.
Heights of bars are drawn proportional to their relative frequencies. The sum of the bar heights equals one.

(m = Modal Class)

Zip code area 1	Frequency	ZIP code area 6	Frequency
A	10	A	0
B	147 *m*	B	172 *m*
C	32	C	26
D	11	D	2
ZIP code area 2	Frequency	ZIP code are 7	Frequency
A	125 *m*	A	38
B	12	B	70 *m*
C	33	C	42
D	30	D	50
ZIP code area 3	Frequency	ZIP code area 8	Frequency
A	137 *m*	A	24
B	10	B	20
C	25	C	24
D	28	D	132 *m*
ZIP code area 4	Frequency	ZIP code area 9	Frequency
A	153 *m*	A	36
B	5	B	40
C	5	C	46
D	37	D	78 *m*
ZIP code area 5	Frequency	ZIP code area 10	Frequency
A	5	A	32
B	135 *m*	B	17
C	50	C	9
D	10	D	142 *m*

ZIP Code Area	Variation Ratio
1	.265
2	.375
3	.315
4	.235
5	.325
6	.140
7	.650
8	.340
9	.610
10	.290

Median-Based Mapping

Table 5.6 Two Hypothetical Ranked Distributions with Identical Medians

Distribution A		Distribution B
69		722
67		341
66		250
52		99
49		98
37	Median	37
36		30
32		28
32		14
32		1
30		1

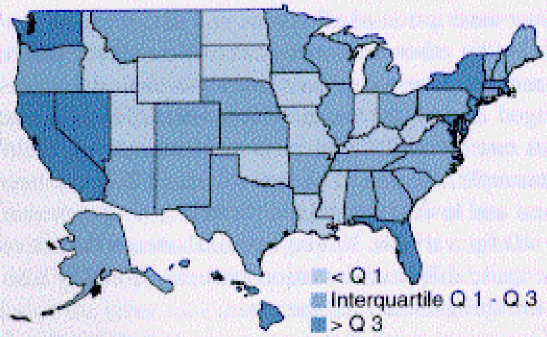


Figure 5.3 Mapping the interquartile range.
States whose values are in the interquartile range are mapped, as are those above Q_3 and below Q_1, with distinct area patterns. There appears to be an absence of spatial clustering of states in the quartile ranges. This map presents the data of Table 5.7.

Table 5.7 Abortion Rates in the United States, 1993

Alabama	18.2	Montana	18.2
Alaska	16.5	Nebraska	15.7
Arizona	24.1	Nevada	44.2
Arkansas	13.5	New Hampshire	14.6
California	42.1	New Jersey	31.0
Colorado	23.6	New Mexico	17.7
Connecticut	26.2	New York	46.2
Delaware	35.2	North Carolina	22.4
District of Columbia	138.4	North Dakota	10.7
Florida	30.0	Ohio	19.2
Georgia	24.0	Oklahoma	12.5
Hawaii	46.0	Oregon	23.9
Idaho	7.2	Pennsylvania	18.6
Illinois	25.4	Rhode Island	30.0
Indiana	12.0	South Carolina	14.2
Iowa	11.4	South Dakota	6.8
Kansas	22.4	Tennessee	16.2
Kentucky	11.4	Texas	23.1
Louisiana	13.4	Utah	9.3
Maine	14.7	Vermont	21.2
Maryland	26.4	Virginia	22.7
Massachusetts	28.4	Washington	27.7
Michigan	25.2	West Virginia	7.7
Minnesota	15.6	Wisconsin	13.6
Mississippi	12.4	Wyoming	4.3
Missouri	11.6		

Median ($Q2$) = 18.9; $Q1$ = 13.45; $Q3$ = 26.3;
Interquartile range = 26.3 – 13.45 = 12.85
Quartile deviation = 12.85/2 = 6.425
Note: Abortion rates per 1,000 women aged 15–44.

Source: Stanley K. Henshaw and Jennifer Van Vort, "Abortion Services in the United States," Family Planning Perspectives *26 (1994). pp. 100–106.*

Mean-Based Mapping

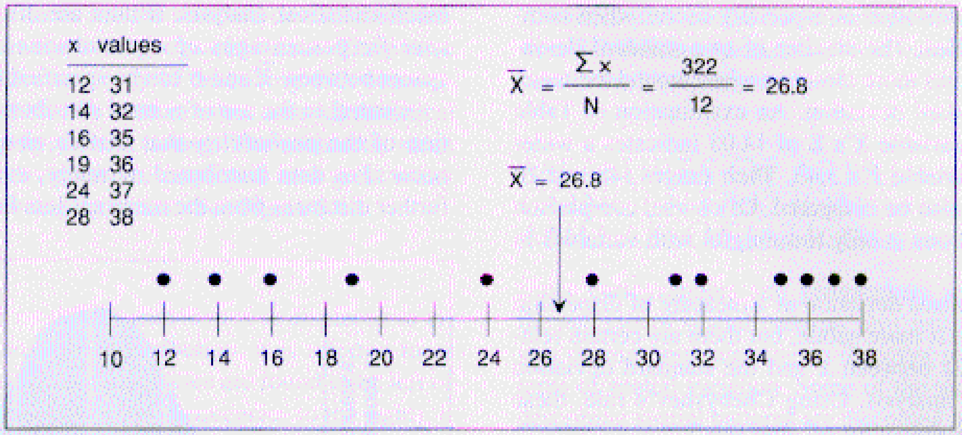


Figure 5.4 Plotting the arithmetic mean on a number line.
The arithmetic mean can be thought of as a balance point for the values on the line.

Table 5.8 Two Data Sets with Identical Means (Units = Tons)

Variable *X*	Variable *Y*
6	28
39	24
11	26
22	29
23	23
44	27
49	21
1	20
31	21
19	29
28	30
27	22
$\Sigma x = 300$	$\Sigma y = 300$
$N = 12$	$N = 12$
$\overline{X} = 25$	$\overline{Y} = 25$
Range = $\lvert 49 - 1 \rvert = 48$	Range = $\lvert 30 - 20 \rvert = 10$
$\overline{X}^2 = 625$	$\overline{Y}^2 = 625$
$\Sigma x^2 = 9{,}864$	$\Sigma y^2 = 7{,}642$
σx^2 (variance) = 196.84	σy^2 (variance) = 11.76
σ (standard deviation) = 14.03 tons	σ (standard deviation) = 3.43 tons

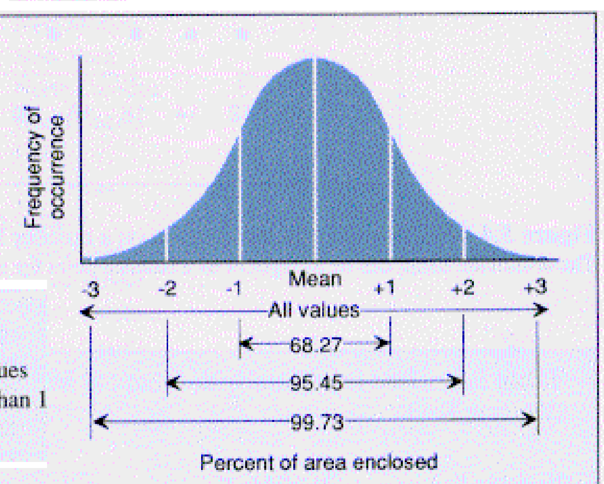


Figure 5.5 The normal distribution.
This figure illustrates the percentage of observations that fall within the standard deviations ±1, ± 2, and ± 3. Note that values above or below three standard deviations are very rare (less than 1 percent).

Table 5.11 Residuals from Regression Table: Variables, Values, and Residuals

Map Region	Independent Variable (*X*)	Dependent Variable (*Y*)	Predicted *Y*(*Ye*)	*Y* Residuals (*Yo-Ye*)
A	5	5	21.25	16.25
B	10	22	25.30	3.30
C	35	55	45.55	9.45
D	30	40	41.55	1.50
E	45	35	53.65	18.65
F	50	50	57.70	7.70
G	20	100	33.40	66.60
H	50	10	57.70	47.70
I	50	95	57.70	37.30
J	15	10	29.35	19.35
K	15	30	29.35	.65
L	25	25	37.45	12.45
M	25	15	37.45	22.45
N	25	55	37.45	17.55
O	15	20	29.35	9.35
P	35	30	45.55	15.55
Q	45	65	53.65	11.35
R	55	50	61.75	11.75
S	40	50	49.60	.40
T	55	75	61.75	13.25
U	35	75	45.55	29.45

Table 5.12 Formula Values for Data in Table 5.11

X Variable	*Y* Variable
$N = 21$	$N = 21$
$\bar{X} = 32.38$	$\bar{Y} = 43.42$
$\Sigma x = 680$	$\Sigma y = 912$
$\Sigma x^2 = 26{,}850$	$\Sigma y^2 = 54{,}634$
$\sigma_x = 15.1$	$\sigma_y = 26.75$
$\Sigma xy = 33{,}445$	

$\Sigma y = a \cdot N + b \cdot \Sigma x$

$\Sigma xy = a \cdot \Sigma x + b \cdot \Sigma x^2$

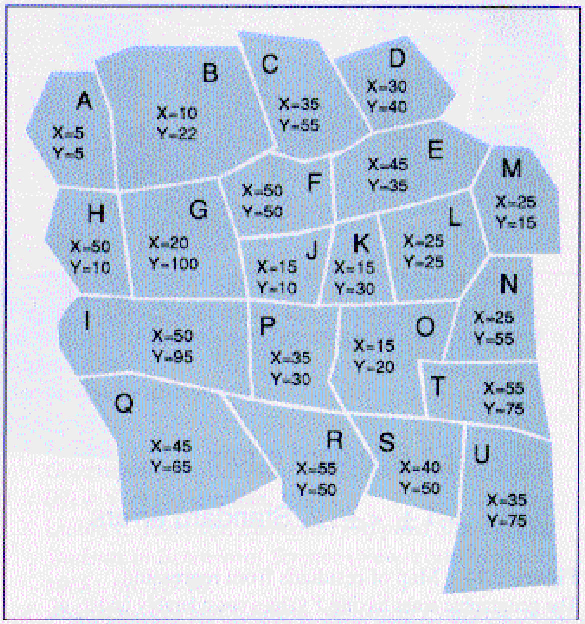
$b = .810$

$a = 17.2$

$\Sigma(Y_o - Y_E) = 11{,}857.012$

r (*correlation coefficient*) $= .4593$

S_E (*standard error of estimate*) $= 24.98$



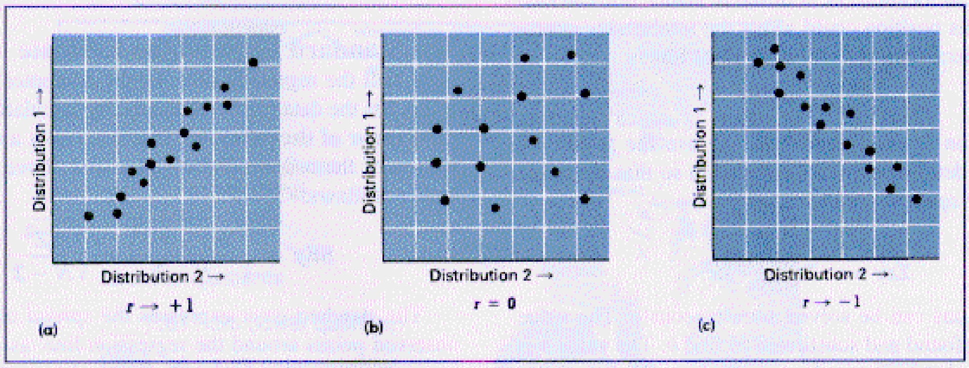


Figure 5.11 Scattergram patterns for different correlation (*r*) values.
In (a), the correlation approaches a strong positive relationship (that points would all fall on a line if $r = 1.0$). In (c), the relationship is strongly negative (inverse). No apparent linear correlation exists in (b), which yields an *r* value of 0.

Solving Simultaneous Eqns.
for Dent Table 5.13

$$\sum y = aN + b\sum x$$
$$\sum xy = a\sum x + b\sum x^2$$

$912 =$

$912 = a21 + b680$
$33{,}445 = a680 + b\,26850$

$912 - a21 - b680 = 33{,}445 - a680$

$21a + 680b - 912 = 680a + 26850b - 33445$

$21a - 680a + 680b - 26850b = -33445 + 912$
$-659a - 26170b = -32533$

$32533 = 659a + 26{,}170b$

$$\frac{32{,}533 - 26170b}{659} = a$$

$$912 = 21\left[\frac{32{,}533 - 26{,}170b}{659}\right] + 680b$$

$$912 = \frac{683{,}193}{659} - \frac{549{,}570b}{659} + 680b$$

$$912 = 1036.7116 - \frac{549{,}570b}{659} + 680b$$

$912 = 1036.7116 - 833.945b + 680b$
$124.7116 = 153.945b$
$.8101 = b$

$912 = 21a + 680(.8101)$
$912 = 21a + 550.868$
$361.132 = 21a$
$17.196 = a$

Regression Residuals and Standard Error

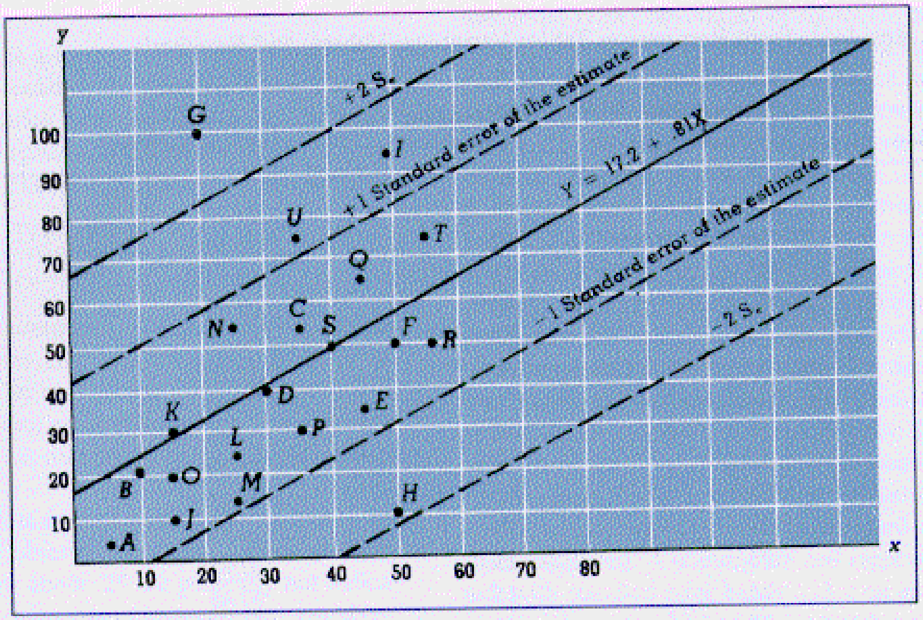


Figure 5.13 Scattergram, regression line, and standard error of the estimate sample plots for the sample data and problem from Table 5.12.

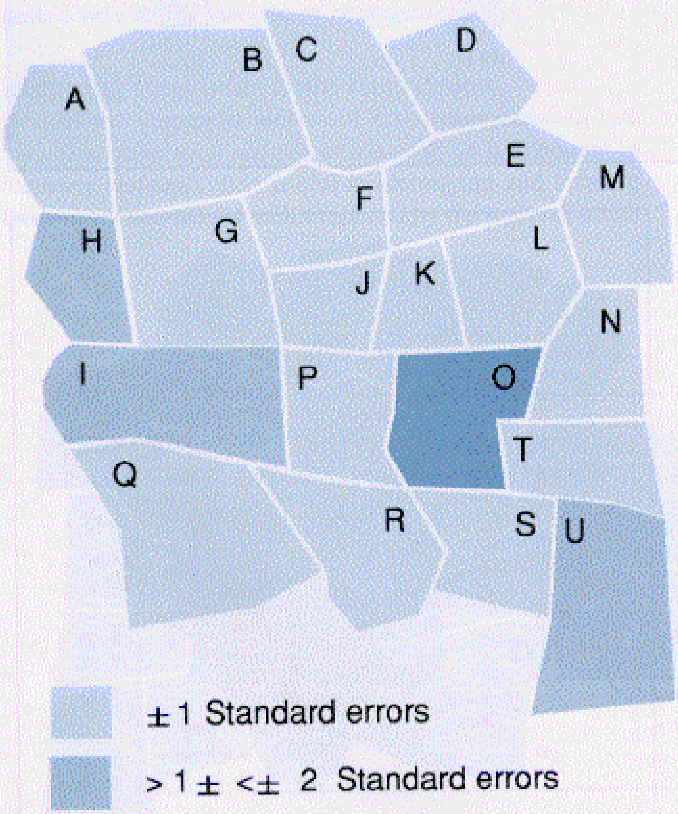


Figure 5.14 Map of residuals from regression. The geographic areas having *Y* values considerably under- or overpredicted relative to the regression line are mapped with identifiable area symbols. These sections of the study area need to be investigated more closely. Further data may be needed to determine why the dependent variables in these sections behave as they do. Identifying deviate areas is a major application of residual mapping.

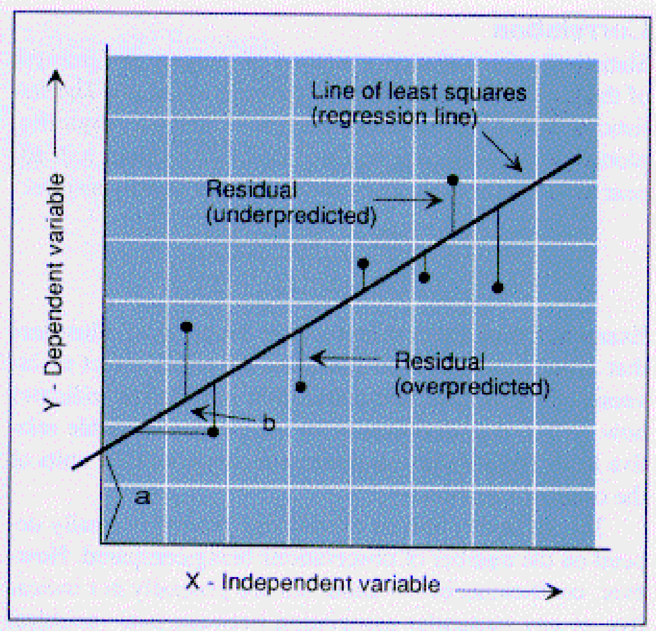


Figure 5.10 Components of the linear regression line. The *a* in the equation of a straight line ($y = ax + bx$) is the intercept of the line with the *y*-axis. The *b* refers to the slope of the line. Notice that both underpredicted and overpredicted residuals may be recorded. When a computed (expected) value of *y* is greater than the observed value, it is overpredicting.