Chapter 7

7-1. a. Sample mean = 11.94; sample standard error = 1.75; the 95% confidence interval = (8.42, 15.46).

c. Sample median = 8; the 95% confidence interval = {3 (19th observation), 12 (32nd observation)}.

d. The 95% tolerance interval to cover 90% of observation, based on normal distribution = 11.94 ± 1.992(12.5) = (0, 36.84; based on distribution-free method, the interval (0, 39) covers 89.6% of observations; the latter method is more appropriate, since the data not distributed normally (it is skewed to the right).

7-2. The 95% confidence interval = {6 days (19th observation), 13 days (31st observation)}.

7-3. Examples can be found in American Journal of Public Health, Journal of American Medical Association, American Journal of Epidemiology, New England Journal of Medicine and many other journals of medicine, public health, nursing, and health services research.

7-4. Would expect a negative correlation because those states that have the higher workplace safety score should have the lower fatality rates. r = -0.435; Since the data are based on population values, there is no need to calculate a confidence interval. However, if we viewed these data as a sample in time, then the formation of a confidence interval is appropriate. The 95% confidence interval = (-0.636, -0.178); a significant negative correlation exists, since the confidence interval does not include zero.

7-5. The prediction interval is appropriate because we are concerned about the tablet that we are about to take.

7-6. variance = 309.31; the 95% confidence interval = (165.80, 709.34); the laboratory may not meet the measurement standard, since the confidence interval does not include 100. The pattern of being close to the true value initially and then larger

and larger deviations from the true value suggests that the process is calibrated on Mondays and then the measuring process goes out of control as the week progresses.

7-7. correlation = 0.145; these data may be viewed as a sample in time; the 95% confidence interval = (-0.136, 0.404); no significant linear relation exists, since the confidence interval includes zero; region of the country, perhaps reflecting the unemployment levels, may play a role.

```
Stata v11 commands:
Note n=51 (50 states plus Washington DC)

input obs pop pct_hlth_ins
 1  1.23 11.1
 2  1.11 10.1
 3  0.56 12.7
 4  6.02 10.9
 5  1.00 10.2
 6  3.29  7.5
 7 17.99 12.3
 8  7.73 10.8
 9 11.88  7.8
10 10.85 10.3
11  5.54 13.0
12 11.43 11.5
13  9.30  9.0
14  4.89  8.0
15  0.64  7.6
16  0.70  9.9
17  1.58  8.3
18  2.48 11.4
19  4.38  9.3
20  2.78  8.8
21  5.12 12.2
22  0.67 13.2
23  4.78 13.1
24  6.19 16.3
25  1.79 15.7
26 12.94 18.6
27  6.63 14.9
28  3.49 13.2
29  6.48 14.1
30  0.61 25.7
31  3.69 13.1
32  4.88 13.4
33  4.04 17.9
```

```
34  2.57 18.9
35  2.35 15.7
36  4.22 20.7
37  3.15 18.2
38 16.99 22.1
39  0.80 12.7
40  1.01 17.8
41  0.45 11.3
42  3.29 10.1
43  1.52 21.5
44  3.67 16.9
45  1.72 13.8
46  1.20 18.7
47  4.87 10.4
48  2.84 14.2
49 29.76 18.7
50  0.55 13.2
51  1.11  7.0
end

correlate pop pct_hlth_ins
(obs=51)

             |      pop pct_hl~s
-------------+------------------
         pop |   1.0000
pct_hlth_ins |   0.1450   1.0000
```

To find the confidence interval refer to Paper 170-31 by Shen and Lu: Computation of Correlation Coefficient and Its Confidence Interval in SAS.

```
gen corr = 0.145
gen n = 51

gen fishersz = 0.5*(ln(1+corr) - ln(1-corr))
gen sigmaz = 1/sqrt(n-3)
gen low95_un = fishersz - 1.96*sigmaz
gen up95_un = fishersz + 1.96*sigmaz

gen low95 = (exp(2*low95_un)-1)/(exp(2*low95_un)+1)
gen up95  = (exp(2*up95_un)-1)/(exp(2*up95_un)+1)
```

7-8. mean percentage = 13.4% which is different from the overall U.S. percentage; the overall percentage can be obtained by calculating a weighted percentage using the population as the weight.

7-9. a. n = 271

7-11. difference = -5.66; 95% confidence interval for the difference = (-15.59, 4.27); no difference, since the confidence interval includes zero.

7-13. difference = -0.261; 99% confidence interval = (-0.532, 0.010); no difference, since the confidence interval includes zero, although a 95% confidence interval would not include zero.

7-14. the indirect age-adjusted fertility rate for Japanese American women = 60.7 per 1,000 with an approximate standard error of 0.797 per 1,000.