Due: Monday December 6, 2004

READING

1. Text (Rosner, Fundamentals of Biostatistics, 5th Edition) Chapter 6. Note – I think it would be fine to just skim 157-168, with the possible exception of the section 6.2 which reviews the relationship between populations and samples. – cb.


EXERCISES:

1. The results of IQ tests are known to be normally distributed. Suppose that in 2002, the distribution of IQ test scores for persons aged 18-35 years has a variance $\sigma^2 = 225$. A random sample of 9 persons take the IQ test. The sample mean score is 115.

   a. Calculate the 50%, 75%, 90% and 95% confidence interval estimates of the unknown population mean IQ score.
   b. What trade-offs are involved in reporting one interval estimate over another?
   c. If it is known that the population mean IQ score is $\mu = 105$, what proportion of samples of size 6 will result in sample mean values in the interval $[135,150]$?

2. A sample of 100 apparently normal adult males, 25 years old, had a mean systolic blood pressure of 125. If it is believed that the population standard deviation is $\sigma = 15$, find:

   a. the 90 percent confidence interval for $\mu$.
   b. the 95 percent confidence interval for $\mu$. 

Web540w\docu\extopic6 (Estimation)
3. Some studies of Alzheimer's disease (AD) have shown an increase in $^{14}$CO$_2$ production in patients with the disease. In one such study, the following $^{14}$CO$_2$ values were obtained from 16 neocortical biopsy samples from AD patients.

$$
1009 \quad 1280 \quad 1180 \quad 1255 \quad 1547 \quad 2352 \quad 1956 \quad 1080 \\
1776 \quad 1767 \quad 1680 \quad 2050 \quad 1452 \quad 2857 \quad 3100 \quad 1621
$$

Assume that the population of such values is normally distributed with a standard deviation of $\sigma = 350$.

a. Construct a 95 percent confidence interval for $\mu$.
b. If the true population mean is $\mu = 1800$ with $\sigma = 350$, what proportion of patient values would be greater than 1900?
c. If the true population mean is $\mu = 1800$ with $\sigma = 350$, what proportion of means of size 16 would be greater than 1900? What proportion of means from samples of size 25 would be greater than 1900?
d. Considering the derivation of confidence interval estimates, comment on the role of sample size in the estimation of the unknown population mean parameter.

4. Using the AD sample data from exercise 3, assume that the population of such values is normally distributed with unknown mean and unknown variance. Construct a 95% confidence interval for the population mean. You might try this in three ways:

a. Do all computations by hand calculator.
b. (OPTIONAL) In STATISTIX (or some other software) from a print out of descriptive statistics (in STATISTIX, select the DESCRIPTIVE STATISTICS menu), obtain the CI.
c. Compare your results with your answers to exercise #3. Comment.
Solutions

1a. Let the random variable $X = \text{IQ test result}$ assumed normal with:

$\mu$ unknown
$\sigma^2 = 225$, known
$\sigma = 15$, known

Confidence interval estimate of the unknown mean is given by:

$\text{estimate} \pm \{ \text{critical value} \} \{ \text{se of estimate} \}$

where,

$\text{estimate} = \text{observed sample mean} = 115$
$\text{critical value} = (1 - \alpha / 2)100\text{th percentile Normal}(0,1)$
$\text{se of estimate} = \text{standard error of sample mean}$

$= \frac{\sqrt{225}}{9}$
$= \frac{15}{3}$
$= 5$

For 50% confidence interval estimate:

$1 - \alpha = (1 - 0.50) = 0.50$
$\alpha/2 = 0.50 / 2 = 0.25$

Therefore want $(1 - .25)100\text{th}$ or 75th percentile $= 0.6745$
The required confidence interval estimate is thus,

$\text{estimate} \pm \{ \text{critical value} \} \{ \text{se of estimate} \}$
$= 115 + \{ 0.6745 \} \{ 5 \}$
$= (111.6, 118.4)$
For 75% confidence interval estimate:

\[ 1 - \alpha = (1 - 0.25) = 0.75 \]
\[ \alpha/2 = 0.25 / 2 = 0.125 \]

Therefore want \((1 - .125)100\)th or 87.5th percentile = 1.1505
The required confidence interval estimate is thus,

\[ \text{estimate} \pm \{ \text{critical value} \} \{ \text{se of estimate} \} \]
\[ = 115 \pm \{ 1.1505 \} \{ 5 \} \]
\[ = (109.2, 120.8) \]

For 90% confidence interval estimate:

\[ 1 - \alpha = (1 - 0.10) = 0.90 \]
\[ \alpha/2 = 0.10 / 2 = 0.05 \]

Therefore want \((1 - .05)100\)th or 95th percentile = 1.645
The required confidence interval estimate is thus,

\[ \text{estimate} \pm \{ \text{critical value} \} \{ \text{se of estimate} \} \]
\[ = 115 \pm \{ 1.645 \} \{ 5 \} \]
\[ = (106.8, 123.2) \]

For 95% confidence interval estimate:

\[ 1 - \alpha = (1 - 0.05) = 0.95 \]
\[ \alpha/2 = 0.05 / 2 = 0.025 \]

Therefore want \((1 - .025)100\)th or 97.5th percentile = 1.96
The required confidence interval estimate is thus,

\[ \text{estimate} \pm \{ \text{critical value} \} \{ \text{se of estimate} \} \]
\[ = 115 \pm \{ 1.96 \} \{ 5 \} \]
\[ = (105.2, 124.8) \]
1b. For a given probability distribution with a known variance and a fixed sample size,

(i) Increasing the confidence coefficient is at the price of a wider confidence interval.

(ii) Decreasing the width of a confidence interval estimate is at the price of a lower confidence coefficient.

This is apparent in the following summary of the solution to Exercise #1a:

<table>
<thead>
<tr>
<th>Confidence Coefficient</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50</td>
<td>111.6</td>
<td>118.4</td>
<td>6.8</td>
</tr>
<tr>
<td>.75</td>
<td>109.2</td>
<td>120.8</td>
<td>11.6</td>
</tr>
<tr>
<td>.90</td>
<td>106.8</td>
<td>123.2</td>
<td>16.4</td>
</tr>
<tr>
<td>.95</td>
<td>105.2</td>
<td>124.8</td>
<td>19.6</td>
</tr>
</tbody>
</table>

1c. Solve for \( \text{Prob} \left[ 135 < \bar{X}_{n=6} < 150 \right] \) using the standardization formula.

Note that \( \bar{X}_{n=6} \) is normally distributed with:

\[
\begin{align*}
\mu &= 105 \\
\sigma^2 &= 225 / 6 = 37.5 \\
\sigma &= \sqrt{37.5} = 6.1238 \\
\text{Thus,}
\end{align*}
\]

\[
\text{Prob} \left[ 135 < \bar{X}_{n=6} < 150 \right] = \text{Prob} \left[ \{ (135-105)/6.1238 \} < Z < \{ (135-105)/6.1238 \} \right] = \text{Prob} \left[ 4.89 < Z < 7.34 \right] < < 0.0001
\]

2a. 

estimate = observed sample mean = 125 

critical value = \((1 - \alpha / 2)\)100th percentile Normal(0,1) 

se of estimate = standard error of sample mean 

= 15 / 10 

= 1.5
For 90% confidence interval estimate:

\[ 1 - \alpha = (1 - 0.10) = 0.90 \]
\[ \alpha/2 = 0.10 / 2 = 0.05 \]

Therefore want \((1 - 0.05)\)100th or 95th percentile = 1.645

The required confidence interval estimate is thus,

\[ \text{estimate} \pm \{ \text{critical value}\} \{ \text{se of estimate}\} \]
\[ = 125 \pm \{1.645\} \{1.5\} \]
\[ = (122.53, 127.47) \]

2b.

\[ \text{estimate} = \text{observed sample mean} = 125 \]
\[ \text{critical value} = (1 - \alpha /2)100\text{th percentile Normal}(0,1) \]
\[ \text{se of estimate} = \text{standard error of sample mean} \]
\[ = 15 / 10 \]
\[ = 1.5 \]

For 95% confidence interval estimate:

\[ 1 - \alpha = (1 - 0.05) = 0.95 \]
\[ \alpha/2 = 0.05 / 2 = 0.025 \]

Therefore want \((1 - 0.025)\)100th or 97.5th percentile = 1.96

The required confidence interval estimate is thus,

\[ \text{estimate} \pm \{ \text{critical value}\} \{ \text{se of estimate}\} \]
\[ = 125 \pm \{1.96\} \{1.5\} \]
\[ = (122.06, 127.94) \]

3a.

(i) point estimate is \(\bar{X} = 1747.6\)
(ii) standard error of point estimate is \(\text{SE}(X) = \sigma / \sqrt{n} = 350/4 = 87.5\)
(iii) solution for confidence coefficient: Since \((1 - \alpha) = 0.95, \ (1-\alpha/2) = 0.975\). Get 97.5th percentile of Normal(0,1) = 1.96. Thus the required confidence interval is

\[ \text{estimate} \pm \{ \text{confidence coefficient}\} \{ \text{se of estimate}\} \]
\[ = 1747.6 \pm \{1.96\} \{87.5\} \]
\[ = (1576.1, 1919.1) \]
3b. Since $\mu = 1800$ and $\sigma = 350$, solve as

$$\text{Prob ( } X \geq 1900 \text{ ) } = \text{Prob ( } Z \geq \{ (1900 - 1800)/350 \} \text{ )}$$
$$= \text{Prob ( } Z \geq 0.2857 \text{ )}$$
$$= 0.3876$$

Conclude that about 38.8% of the patient values would be greater than 1900.

3c.

$$\text{Prob ( } \bar{X}_{n=16} \geq 1900 \text{ ) } = \text{Prob ( } Z \geq \{ (1900 - 1800)/(350/4) \} \text{ )}$$
$$= \text{Prob ( } Z \geq 1.143 \text{ )}$$
$$= 0.1265. \text{ Conclude that about 12.7% of sample means of size 16 would be greater than 1900.}$$

$$\text{Prob ( } \bar{X}_{n=25} \geq 1900 \text{ ) } = \text{Prob ( } Z \geq \{ (1900 - 1800)/(350/5) \} \text{ )}$$
$$= \text{Prob ( } Z \geq 1.429 \text{ )}$$
$$= 0.0765. \text{ Conclude that about 7.7% of sample means of size 25 would be greater than 1900.}$$

3d.

As indicated in problem #1, the width of a confidence interval for the mean parameter of a normal probability distribution with known variance is:

$$\text{(2) \{ critical z \} \{ standard deviation / \sqrt{n} \}}$$

where $n$ is the number of observations in the sample.

This means that, for fixed critical $z$ and for fixed standard deviation, as $n$ increases, the quantity $1/\sqrt{n}$ decreases. Consequently, the width of the confidence interval also decreases. Narrower, or more precise, confidence interval estimates are obtained when obtained from data sets of larger sample sizes.
4a. Calculate sample variance $s^2$ and from this obtain $s = 604.65$

(i) point estimate is $\bar{X} = 1747.6$

(iv) **sample** standard error of point estimate is $s/\sqrt{n} = 604.65/4 = 151.16$

(v) solution for confidence coefficient: Since $(1 - \alpha) = 0.95$, $(1-\alpha/2)=0.975$. Get 97.5th percentile of student’s t distribution with 15 degrees of freedom = 2.13. Thus the required confidence interval is

\[
\text{estimate} \pm \left\{ \text{confidence coefficient from student’s } t \right\} \left\{ \text{standard error of estimate} \right\} \\
= 1747.6 \pm \left\{ 2.13 \right\} \left\{ 151.16 \right\} \\
= (1425.63, 2069.57)
\]

4b. If you used STATISTIX, check that your answer matches.

4c. The confidence interval in exercise #4a is much wider than that obtained in #3a. This reflects the added uncertainty in using the estimate $s$ of $\sigma$, rather than $\sigma$ itself. It also reflects the comparatively larger value of $s$ ($s=604.65$) compared to $\sigma$ ($\sigma = 350$).