# Chapter 7: Intervals From a Single Sample

# Point estimates

*Point Estimate* A point estimate of a population parameter is a single number calculated from the sample. For example

- Mean: Usually we estimate $\mu$ by the sample mean, $\bar{x}$. That is

$$\widehat{\mu} = \bar{x}$$

- Proportion: We estimate the population proportion, $p$, by the sample proportion $\widehat{p} = X/n$.

*Estimator* The estimator is the random variable whose value will be the point estimate. For example $\bar{X}$ is the estimator of the population mean.

*Confidence Interval* A confidence interval is an interval calculated from the data that has a given probability of covering the true parameter value.

# Confidence interval when $\sigma$ known

- This is an artificial case – for illustration only.
- We assume that the population is normal and $\sigma$ is known. Then

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

  will have a standard normal distribution.
- Hence

$$\mathrm{P}\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

- This implies that

$$\mathrm{P}\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- We construct the *95% confidence interval on $\mu$* as

$$\bar{x} \pm 1.96 \cdot \sigma/\sqrt{n}$$

# Interpretation

If we were repeatedly to sample from the distribution, about 95% of the intervals calculated in this way would cover the true mean $\mu$.

The calculation of the interval from one sample would be like

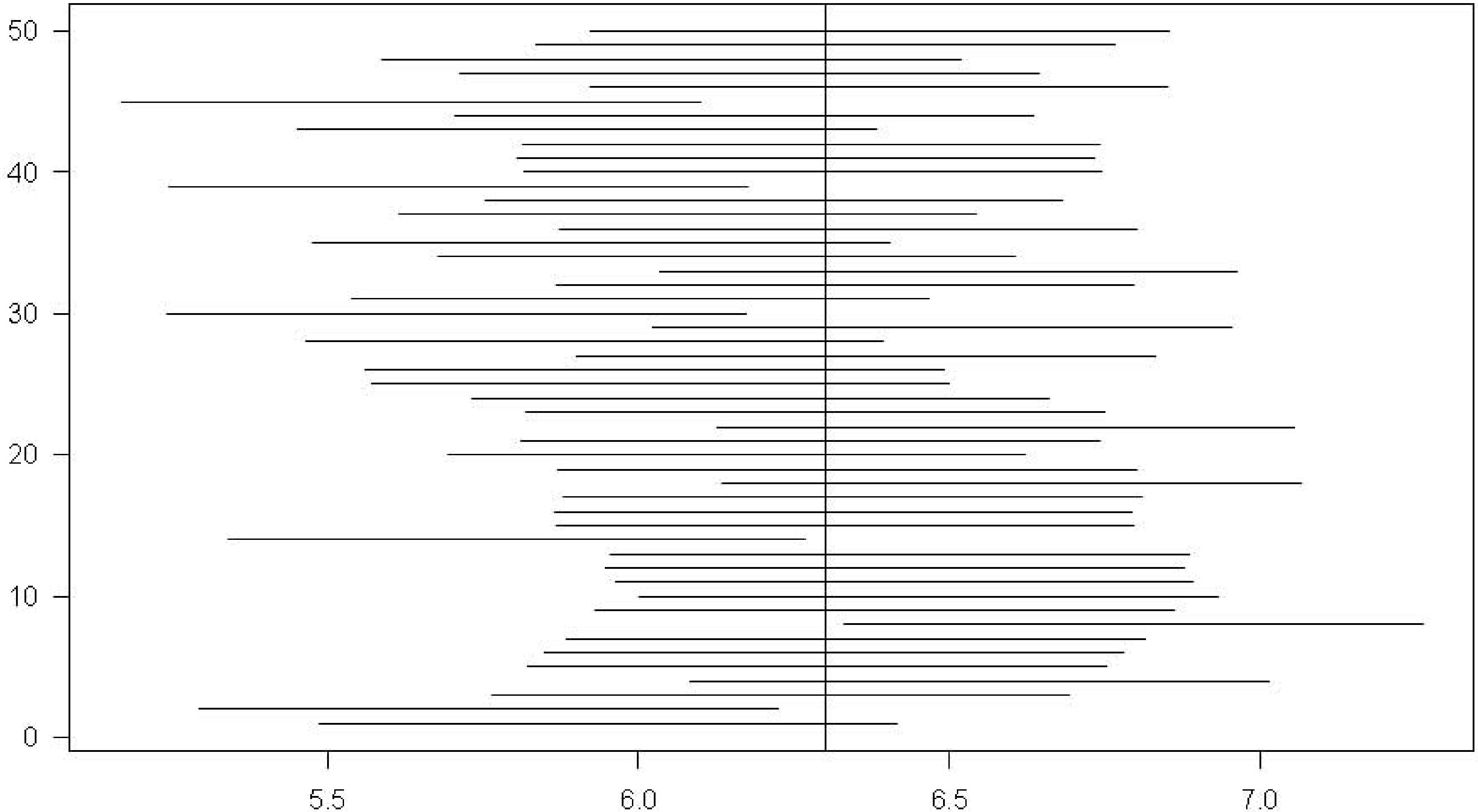
```
> str(samp <- rnorm(10, mean = 6.3, sd = 0.75))

 num [1:10] 4.84 7.37 5.62 5.82 7.60 ...

> mean(samp)

[1] 6.218253

> mean(samp) + c(lower = -1, upper = +1) * 1.96 * 0.75/sqrt(10)

   lower    upper
5.753398 6.683108
```

# Confidence intervals from 50 simulations

```
> (samp <- matrix(rnorm(50 * 10, m = 6.3, sd = 0.75), nr = 10))[,
+     1:5]

          [,1]     [,2]     [,3]     [,4]     [,5]
 [1,] 5.431663 4.977463 7.130403 7.471731 5.939213
 [2,] 6.614273 6.273929 6.215217 6.061846 5.819205
 [3,] 5.304584 6.816851 5.184538 6.290099 5.848198
 [4,] 5.700838 5.435696 6.071947 5.787558 6.436641
 [5,] 5.607297 5.242968 6.026160 7.334839 7.118434
 [6,] 5.876926 5.335455 6.756167 6.043545 5.799262
 [7,] 5.727950 5.930014 6.222570 6.831208 6.616852
 [8,] 6.484057 5.526202 5.870854 6.432547 7.038318
 [9,] 6.906185 6.048388 6.896734 7.179988 5.579032
[10,] 5.862187 6.004191 5.911062 6.054618 6.684354

> (lims <- outer(c(lower = -1, upper = 1) * 1.96 * 0.75/sqrt(10),
+     colMeans(samp), "+"))[, 1:5]

          [,1]     [,2]    [,3]     [,4]     [,5]
lower 5.486741 5.294261 5.76371 6.083943 5.823096
upper 6.416451 6.223970 6.69342 7.013653 6.752806
```

# Plots of simulated confidence intervals

In this case only 44/50 or 88% of the intervals cover $\mu = 6.3$.



# Other levels of confidence

The multiplier 1.960 is determined from probabilities of the standard normal curve. In general a $100(1-\alpha)\%$ confidence interval on $\mu$ (when $\sigma$ is known) is

$$\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

Some common values of the multiplier are

| $100(1-\alpha)\%$ | $\alpha$ | $\alpha/2$ | $z_{\alpha/2}$ |
|---|---|---|---|
| 80% | 0.200 | 0.100 | 1.282 |
| 90% | 0.100 | 0.050 | 1.645 |
| 95% | 0.050 | 0.025 | 1.960 |
| 99% | 0.010 | 0.005 | 2.576 |

The numerical values come from the last row of table A.5 (p. 725), reproduced on the inside back cover of the text.

# Graph showing critical values

For a 90% confidence interval we use $z_{\alpha/2} = z_{0.05} = 1.645$. When the standard normal curve is divided at -1.645 and +1.645 it has 5% of the area in the left tail, 90% in the middle, and 5% in the right tail.