Review: recall Table 4.1

- statement: \( s^2 = \frac{SS}{n-1} \)
- proof:
  - collect the set of all possible samples for \( n=2 \) selected from the population
  - compute \( s^2 \) for \( n \) and for \( n-1 \) and you’ll see for yourself: \( \mu = 4 \) = the average of SS using \( n-1 \)
Our proof begins with an analogy that has us knowing “reality” (population)

= “Reality” (population; “known unknown”)
e.g., \( P(\text{King}) \) given an old deck with some cards missing

The distribution of sample means
(dfn) the set of all possible random (w/ replacement) samples of size \( n \)

\( \mu = ?? \) \( \sigma = ?? \)

What is the probability of obtaining sample (2,4), or \( M > 7 \)?
Reminder

- We are not talking about the distribution of the sample but the distribution of means from the set of all possible samples.

\[
\begin{array}{cccccc}
\text{Sample 1} & X & X \\
1 & 2 & 3 & 4 & 5
\end{array}
\]

Sample n
.
.
.

The distribution of sample means for n=2

- But we can’t draw the complete set of samples, thus...
- Central Limit Theorem: \( M = \mu \) and standard deviation = \( \sigma/\sqrt{n} \) as \( n \to \infty \)

Fig. 7.3
Distribution of sample means (contd.)

The distribution of sample means will be normal if ... pick one:
- the population sample is normally distributed, or
- n is relatively large, i.e., >30

In most situations with n > 30 the distribution of means will be normal regardless of the distribution of the population. Note. M (of sample means) is an unbiased stat; cf. M in chapter 4 - had to correct s²

The expected value

Distribution of sample means (contd.)

• expected value of M will equal μ

• Thus far to describe samples we’ve used:
  - central tendency (mean, median, mode)
  - shape (or variability of the sample)

• Similarly of the distribution of sample means
  - mean
  - now add, shape or standard deviation, called standard error of M (i.e., a sample will not perfectly represent the population)
    = standard distance between M and μ

• Magnitude of the error is related to: size of sample and standard dev of the population
Distribution of sample means (contd.)

- The law of large numbers
  - $M = \mu$
  - standard error of $M$ decreases
- Formula for standard error: $\sigma_M = \frac{\sigma}{\sqrt{n}}$

<table>
<thead>
<tr>
<th>SAMPLE SIZE</th>
<th>STANDARD ERROR IN TERMS OF $\sigma$</th>
<th>STANDARD ERROR IN TERMS OF $\sigma^2$</th>
<th>STANDARD ERROR FOR $\sigma = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 1$</td>
<td>$\sigma\sqrt{1}$</td>
<td>$\sqrt{\sigma^2/1}$</td>
<td>10.00</td>
</tr>
<tr>
<td>$n = 2$</td>
<td>$\sigma\sqrt{2}$</td>
<td>$\sqrt{\sigma^2/2}$</td>
<td>7.07</td>
</tr>
<tr>
<td>$n = 3$</td>
<td>$\sigma\sqrt{3}$</td>
<td>$\sqrt{\sigma^2/3}$</td>
<td>5.77</td>
</tr>
<tr>
<td>$n = 4$</td>
<td>$\sigma\sqrt{4}$</td>
<td>$\sqrt{\sigma^2/4}$</td>
<td>5.00</td>
</tr>
<tr>
<td>$n = 5$</td>
<td>$\sigma\sqrt{5}$</td>
<td>$\sqrt{\sigma^2/5}$</td>
<td>4.47</td>
</tr>
<tr>
<td>$n = 6$</td>
<td>$\sigma\sqrt{6}$</td>
<td>$\sqrt{\sigma^2/6}$</td>
<td>4.08</td>
</tr>
<tr>
<td>$n = 7$</td>
<td>$\sigma\sqrt{7}$</td>
<td>$\sqrt{\sigma^2/7}$</td>
<td>3.78</td>
</tr>
<tr>
<td>$n = 8$</td>
<td>$\sigma\sqrt{8}$</td>
<td>$\sqrt{\sigma^2/8}$</td>
<td>3.54</td>
</tr>
</tbody>
</table>

More about standard error

- standard error vs. standard deviation?????
  - whenever you are talking about a sample use standard error, i.e., always
  - always some error: 50% of $M$s < $\mu$
When \( n = 1 \), then \( \sigma_M = \sigma \) (recall: for the full set of samples \( n = 1 \))

Looking ahead to inferential statistics

- Rat pups are treated with a growth hormone: \( \mu = 400 \) but not all same size \( \sigma = 20 \), treat a sample of 25

\[
\sigma_M = \frac{\sigma}{\sqrt{n}} \\
Z = \frac{M - \mu}{\sigma_M}
\]
Standard error as an estimate of reliability

- The problem:
  - I must estimate the population from a sample
  - but if I had a different sample, I would obtain a different result
  - the question becomes: would the first sample be similar to the second sample (or 3rd or 4th, etc.)
  - in the previous example it was easy to find two samples with very different means
  - by increasing n, you increase your confidence that your sample is a measure of your population

Summary (p. 222)

1) What is the distribution of sample means? the set of all Ms for all possible random samples for sample size n for a given population.
   1) shape: population must be N or n>30
   2) central tendency: $M = \mu$
   3) variability: $\sigma_M = \frac{\sigma}{\sqrt{n}}$

2) Standard error: the standard deviation of (1) - tells us how much error to expect if using a sample to estimate a population

3) Location of M in the distribution
Learning check (p. 209 Q#1)

- Population of scores is normal $\mu=80$ $\sigma=20$
  - describe the distribution of sample means of size $n = 16$: shape? central tendency? variability?
  - What if $n = 100$?