Review: Central Limit Theorem

What is the sampling distribution of the sample average?

Central Limit Theorem (CLT): For population with mean $\mu$ and standard deviation $\sigma$,

1. the mean value of the collection of all possible sample means will equal the mean of the original population;
2. the standard deviation of the collection of all possible means of samples of a given size is $\frac{\sigma}{\sqrt{n}}$;
3. the distribution of sample means will be approximately normal regardless of the distribution of values in the original population from which the samples were taken.

$\bar{Y}$ is approximately $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

**Review: t-distribution**

- In most applications, $\sigma$ is unknown, and is estimated by $s$. This approximation is usually satisfactory for $n \geq 30$.
- When we have to estimate $s$, the quantity $\frac{\bar{Y} - \mu}{s/\sqrt{n}}$ is not normally distributed.
- The quantity $\frac{\bar{Y} - \mu}{s/\sqrt{n}}$ has a $t$ distribution with $n - 1$ degrees of freedom.

$$\bar{Y} - \mu \approx t_{n-1}$$

$$SE(\bar{Y}) = s/\sqrt{n}$$ with (d.f.=$n$-1) measures:
- the accuracy of our estimate of $\bar{Y}$
- the variability of possible values of means of samples of $n$ from $N$

- For $n$ large, the $t$ distribution with $n$ degrees of freedom approaches the standard normal distribution.
- Percentiles of the $t$ distribution: Table A.2, p. 710, Sleuth
- Application: 100(1-$\alpha$) CI for the mean $\mu$ when $\sigma$ unknown: $\bar{Y} \pm t_{1-\alpha/2, n-1} \left(\frac{s}{\sqrt{n}}\right)$

**1/9 Lecture: Review of 1-sample t-test**

- Observational study of pollutant levels during the day and at night.
- Paired study: rationale?
- Calculate $\bar{D}$ and $SE(\bar{D})$
- Hypothesis test: $H_0 : \mu_D = 0$ vs. $H_a : \mu_D \neq 0$
- Assumptions of $t$-test
- Rejection region, significance level, $p$-value, 95% confidence interval for $\mu_D$.
- Duality of confidence intervals and hypothesis tests

Today: Further review of Statistical Sleuth, Ch. 2, 3, 5

**Review of two-sample t-test: Mormon Cricket**

In a field study of the mating behavior of the Mormon cricket (Anabrus simplex), a biologist noted that some females mated successfully while others were rejected by the males before coupling was complete. The question arose whether some aspect of body size might play a role in mating success. Summary of head width (mm) in two groups of females.

<table>
<thead>
<tr>
<th></th>
<th>Successful</th>
<th>Unsuccessful</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{Y}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Give a 95% CI for the difference in mean head widths for successful and unsuccessful groups.
- Use the interval in (a) to make a conclusion about the test of

$$H_0 : \mu_s - \mu_u = 0$$ vs. $$H_A : \mu_s - \mu_u \neq 0$$
Review: ANOVA models

- Chapter 5 of Sleuth
- As in 2-sample problems, one-way classification setting
- K populations: \( n_j \) observations from population \( j \), \( j = 1, \cdots, K \)
- \( \sum_{j=1}^{K} n_j = N \).
- ANOVA model:
  \[
  y_{ij} = \mu_j + \epsilon_{ij}, \quad i = 1, \cdots, n_j; \quad j = 1, \cdots, K
  \]
  - \( y_{ij} \)'s varying about their population mean \( \mu_j \).
  - \( \epsilon_{ij} \sim N(0, \sigma) \) where \( \epsilon_{ij} \) represent random variation from fit of ANOVA model.

- 4 Key Assumptions? (top p. 114, Sleuth)
- Types of inference
  - single comparison of two means
  - multiple, simultaneous comparisons (all possible pairwise differences, linear combinations of means)

Review: ANOVA hypotheses

\( H_0 : \quad \mu_1 = \mu_2 = \cdots = \mu_K \)
(reduced model; equal means)

\( H_A : \quad \) not all \( \mu_i \)'s equal
(full model; separate means)
ANOVA example: Fertilizer and Plant Growth

- What is the optimum level of fertilization (using MgNH₄PO₄) to maximize vertical growth of a plant?
- 100 seedlings were divided into 4 groups, 25 plants each.
- Each was planted in a similar pot containing a uniform growth medium, under uniform conditions for several weeks in a greenhouse.
- Increasing concentrations of MgNH₄PO₄, measured in grams per bushel, for each group.

Data:
- Col 1: height in cm
- Col 2: code for level. 1=50 grams/bushel, 2=100 g/bu, 3=200 g/bu, 4=400 g/bu.