- One Sample
  - Estimation of the population mean
  - Estimation of the population proportion
- Two Samples
  - Estimation of the difference between two population means

    (Independent samples, Paired samples)
  - Estimation of the difference between two population proportions

Sampling Distribution

A random sample $y_1, y_2, \ldots, y_n$ is drawn from a distribution with mean μ and variance σ^2

$$\text{sample mean } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\text{sample var } s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

- By Central Limiting Theorem,

$$\frac{\bar{y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- If σ is unknown,

$$\frac{\bar{y}-\mu}{s/\sqrt{n}} \sim \text{ Student's t distribution}$$

  with $(n-1)$ degree of freedom.

- When n is large,

$$\frac{\bar{y}-\mu}{s/\sqrt{n}} \sim N(0,1)$$

Estimation of a Population Mean

Textbook:	Section 8.5
Population Model :	mean μ and variance σ^2
Estimator:	$\hat{\mu} = \bar{y}$

Large Sample, $n \geq 30$

Pivotal stat : $z = \frac{\bar{y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Approximation: $\sigma \approx s$ when σ is unknown

CI : $\bar{y} \pm z_{\alpha/2}(\frac{s}{\sqrt{n}})$

Small Sample $n < 30$

Pivotal stat : $z = \frac{\bar{y}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

CI : $\bar{y} \pm t_{\alpha/2,n-1}(\frac{s}{\sqrt{n}})$

Example 8.10: Want to estimate the difference between the mean starting salaries for recent graduates with mechanical engineering and civil engineering degrees.

(1) Randomly choose 59 ME graduates and record their salaries

(2) Randomly choose 30 CE graduates and record their salaries

...

Generally Setting: Let $\bar{y_1}$ and s_1^2 be the sample mean and sample variance, respectively, of n_1 obs randomly selected from a population with mean μ_1 and variance σ^2. Similarly, define $\bar{y_2}$ and s_2^2 for an independent random sample of n_2 obs from another population with mean μ_2 and variance σ_2^2.

How to estimate $\mu_1 - \mu_2$*?*

Difference Between Two Means $\mu_1 - \mu_2$

Textbook:	Section 8.6
Population Models :	one with mean μ_1 and variance σ_1^2
	another with mean μ_2 and variance σ_2^2
Estimator :	$\bar{y}_1 - \bar{y}_2$

Independent Samples, Large Sample

Pivotal stat : $z = \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}} \sim N(0,1)$

Approximation: $\sigma_1 \approx s_1$ and $\sigma_2 \approx s_2$

CI : $(\bar{y}_1 - \bar{y}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Independent Samples, Small Sample, $\sigma_1^2 = \sigma_2^2$

Pivotal stat : $z = \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1)$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

CI : $(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2, n_1 + n_2 - 2}\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}$

Independent Samples, Small Sample, $\sigma_1^2 \neq \sigma_2^2$

Example: (Cement mixes example from section 8.7) Suppose we want to compare two methods for drying concrete.

Experiment 1: randomly choose 10 cement mixes from among all available mixes and then randomly assign 5 to method 1 and 5 to method 2, ...

Experiment 2 (matched pairs): randomly choose 5 mixes; for each mix, assign one specimen to method 1 and one for method 2, ...

Generally Setting: In the experiment, we have n individuals or experimental objects, and two observations are made on each individual or object. So we have n independent selected pairs $(x_1, y_1), \ldots, (x_n, y_n)$. Let $d_i = x_i - y_i$ record the difference within pairs. Assume that the paired difference d_i's are from some distribtion with mean μ_d and variance σ_d.

How to estimate the paired difference μ_d

Difference Between Two Means

Matched Pairs

Textbook:	Section 8.7
Population Model :	pairwise difference $d_1, \ldots d_n$ from a dist with mean μ_d and variance σ_d^2
Estimator:	$\bar{d}$

Matched Pairs, Large Sample

Pivotal stat : $z = \frac{\bar{d}-\mu_d}{\sigma_d} \sim N(0,1)$

Approximation: $\sigma_d \approx s_d$ when σ_d is unknown

CI : $\bar{d} \pm z_{\alpha/2}(\frac{s_d}{\sqrt{n}})$

Matched Pairs, Small Sample

Pivotal stat : $z = \frac{\bar{d}-\mu_d}{s_d} \sim t_{n-1}$

CI : $\bar{d} \pm t_{\alpha/2,n-1}(\frac{s_d}{\sqrt{n}})$

Estimation of a Population Proportion

Example: A quality control inspector may be interested in the proportion of defective items produced on an assembly line...

Textbook:	Section 8.8
Population Model :	Binomial(n,p)
Estimator :	$\hat{p} = y/n$

Large Sample

Pivotal stat : $z = \frac{\hat{p}-p}{\sqrt{pq/n}} \sim N(0,1)$

Approximation: $pq \approx \hat{p}\hat{q}$

CI : $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

Difference Between Two Proportions

Example: A quality control inspector may be interested in comparing the proportion p_1 of defective items produced by machine 1 to the proportion p_2 of defective items produced by machine 2...

Textbook:	Section 8.9
Population Models :	$y_1 \sim \text{Bino}(n_1, p_1)$
	$y_2 \sim \text{Bino}(n_2, p_2)$
Estimator:	$\widehat{p_1} - \widehat{p_2} = y_1/n_1 - y_2/n_2$

Large Sample

Pivotal stat :	$z = \frac{\widehat{p}_1 - \widehat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}}} \sim N(0, 1)$
Approximation:	$p_i q_i \approx \widehat{p}_i \widehat{q}_i$ for $i = 1, 2$
CI :	$(\widehat{p_1} - \widehat{p_2}) \pm z_{\alpha/2} \sqrt{\frac{\widehat{p}\widehat{q}}{n_1} + \frac{\widehat{p}\widehat{q}}{n_2}}$