Chapter 9. Inferences based on Two ~~Populations~~ "Samples"

Recall: Chapter 8 began with the problem
"A well-known population has mean μ_0.
Sample a 'similar' population X n times to get $\bar{x}$"

- §8.2 Unknown $E[X]$ and known $Var[X]$
  ↳ "Use z-Test against $H_0: E[X] = \mu_0$"
- §8.3 Unknown $E[X]$ and unknown $Var[X]$
  ↳ "Use t-Test against $H_0: E[X] = \mu_0$"

Chapter 9 removes another "known"
- Ch 9 Unknown μ_0 (mean of comparison population.)

§9.1: Two ~~Population~~ "Sample" z-Test

Testing difference between indep. population means

Setup: X & Y are independent (but similar) random variables with $\mu_X = E[X]$ and $\mu_Y = E[Y]$ unknown.
- Sample X n times to get $\bar{x}$
- Sample Y m times to get $\bar{y}$

Test $H_0: \mu_X - \mu_Y = \Delta$ or $H_0: \mu_X - \mu_Y < \Delta$
↑ Often $\Delta = 0$ so $H_0: \mu_X = \mu_Y$

Theory:
$\bar{X} \approx \text{Normal}\left(\mu_X, \sigma_X/\sqrt{n}\right)$
$\bar{Y} \approx \text{Normal}\left(\mu_Y, \sigma_Y/\sqrt{m}\right)$
(If $n, m > 30$ by "Central Limit Theorem")

$\Downarrow$

$$\bar{X} - \bar{Y} \approx \text{Normal}\left(\mu_X - \mu_Y, \sqrt{\sigma_X^2/n + \sigma_Y^2/m}\right)$$

$$\left[\begin{aligned} Var[\bar{X} - \bar{Y}] &= Var[\bar{X}] + Var[\bar{Y}] \\ &= \sigma_X^2/n + \sigma_Y^2/m \end{aligned}\right]$$
(because X & Y are indep.)

Hypothesis Test: $H_0: \mu_X - \mu_Y = \Delta$

Test Statistic:

$$\frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} \approx \text{Normal}(0, 1)$$

p-value:

Absolute value converts prob. to left tail for pnorm

$$\text{pnorm}\left(-\left|\frac{(\bar{x} - \bar{y}) - \Delta}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}}\right|\right)$$

↑ 2×this for Two-Tailed Test.

(If σ_X & σ_Y are unknown & n, m are big then we can use s_X & s_Y instead.)