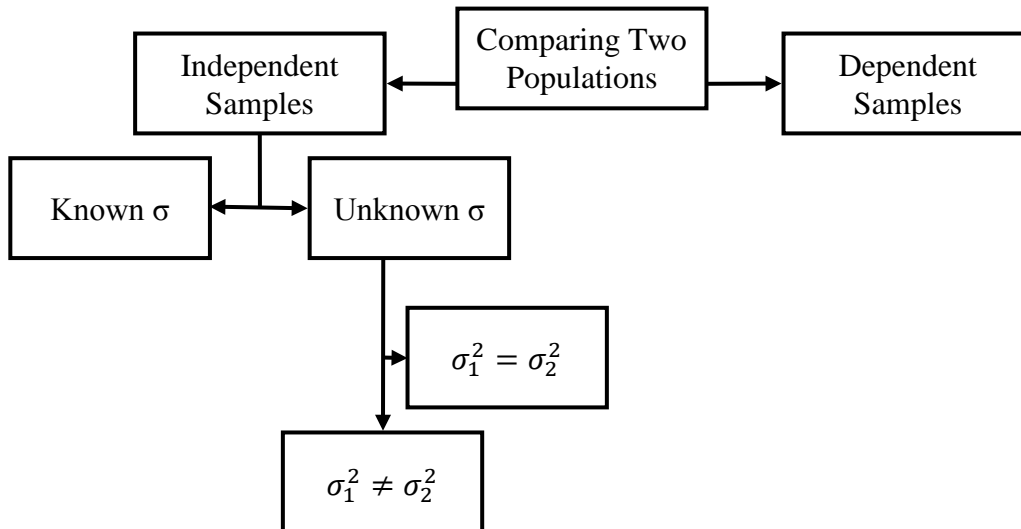David Youngberg
BSAD 210—Montgomery College

LECTURE 24: COMPARING TWO POPULATIONS I

I. On Comparing Populations
   a. We don't really compare two different populations directly; as usual, we take a sample from each population.
   b. But we call it comparing populations because these samples represent that bigger picture. There's really no point to limiting our analysis to just the sample. We want to know what's going on for the population.
   c. There's a lot of complexity that seeps into this sort of analysis. Two summarize what we'll be working with, here's a diagram:

```
            ┌─────────────┐
            │ Comparing Two│
┌──────────┐│  Populations │┌──────────┐
│Independent│◄────────────►│ Dependent │
│ Samples   │└─────────────┘│  Samples  │
└──────────┘               └──────────┘
   │    │
┌──────┐ ┌──────────┐
│Known σ│◄►│Unknown σ │
└──────┘ └──────────┘
              │
         ┌──────────┐
         │ σ₁² = σ₂² │
         └──────────┘
    ┌──────────┐
    │ σ₁² ≠ σ₂² │
    └──────────┘
```

Known σ   Unknown σ

$$\sigma_1^2 = \sigma_2^2$$

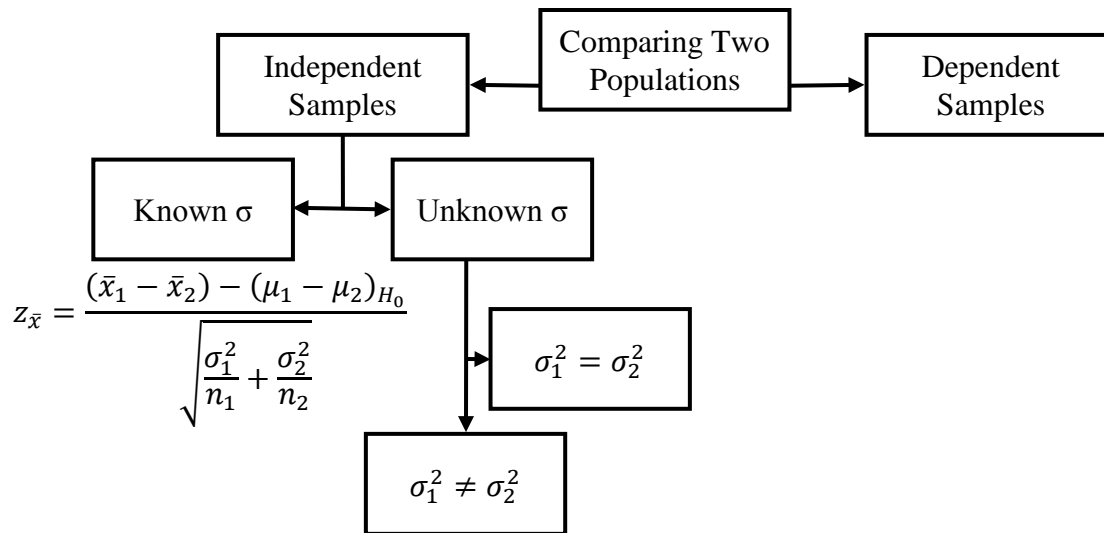$$\sigma_1^2 \neq \sigma_2^2$$

   i. When samples are *independent*, each observation for one sample does not change with the observations for the other sample.
      1. Example: the safety ratings of two different kinds of cars. If one trial for a Honda is good, that wouldn't alter the trial of a Ford good or bad. Ford's safety is independent from Honda's.
   ii. *Dependent* samples are when the observations of one sample change based on the observations of the other sample.
      1. Example: comparing the effectiveness of a diet. The more you weigh before a diet, the more you should weigh after a diet, even if the diet is effective. In other words,

what you're really interested in is the change in weight between each observation.

    iii. The population standard error can either be known or unknown. Known σ typically relate to samples drawn from a population with a pre-existing body of empirical data.

II. Known σ

    a. Let's begin with the simplest case: the standard deviation of both populations is known.

    b. This occurs when there's a large body of pre-existing data on the population in question. Here's the equation:

```
                    ┌──────────────┐
  ┌──────────────┐  │ Comparing Two│  ┌──────────────┐
  │ Independent  │◄─┤ Populations  ├─►│  Dependent   │
  │   Samples    │  └──────────────┘  │   Samples    │
  └──────┬───────┘                    └──────────────┘
         │
  ┌──────────┐       ┌──────────────┐
  │ Known σ  │◄─────►│  Unknown σ   │
  └──────────┘       └──────┬───────┘
                            │      ┌──────────────┐
                            ├─────►│ σ₁² = σ₂²    │
                            │      └──────────────┘
                            ▼
                     ┌──────────────┐
                     │  σ₁² ≠ σ₂²   │
                     └──────────────┘
```

$$z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

$$\sigma_1^2 = \sigma_2^2$$

$$\sigma_1^2 \neq \sigma_2^2$$

III. Example

    a. Suppose our company has a new candy bar and marketing came up with two different displays. Which display works better? We put each display in different stores to find out.

    b. Our null hypothesis is that the difference doesn't matter: $\mu_1=\mu_2$; our alternative is that it does: $\mu_1\neq\mu_2$.

        i. We used Display 1 in 50 stores and got $400K in sales.

        ii. We used Display 2 in 25 stores and got $410K in sales.

        iii. We know from previous new candy bars, the population standard deviation for each is $20K.

$$z_{\bar{x}} = \frac{(400 - 410) - 0}{\sqrt{\dfrac{20^2}{50} + \dfrac{20^2}{25}}} = \frac{-10}{\sqrt{\dfrac{400}{50} + \dfrac{400}{25}}} = \frac{-10}{\sqrt{8 + 16}} = \frac{-10}{4.899} = -2.04$$

        iv. At 95% confidence (z=1.96), the results are statistically significant.