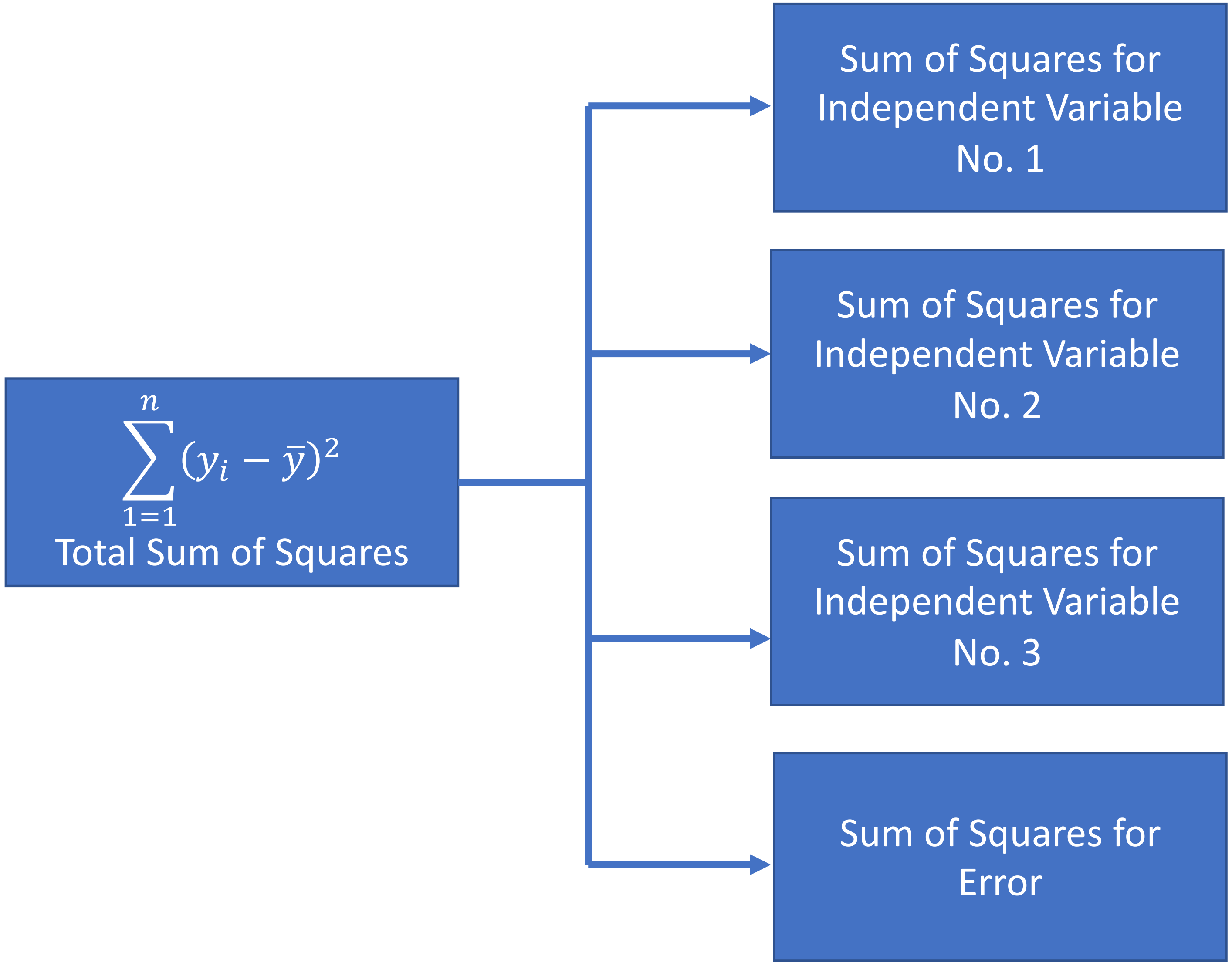
Analysis of Variance (ANOVA)

PSTAT 120C

Outline

- **Introduction**
- One-way ANOVA
- Two-way ANOVA
- Sample size
- ANOVA and linear models

Introduction

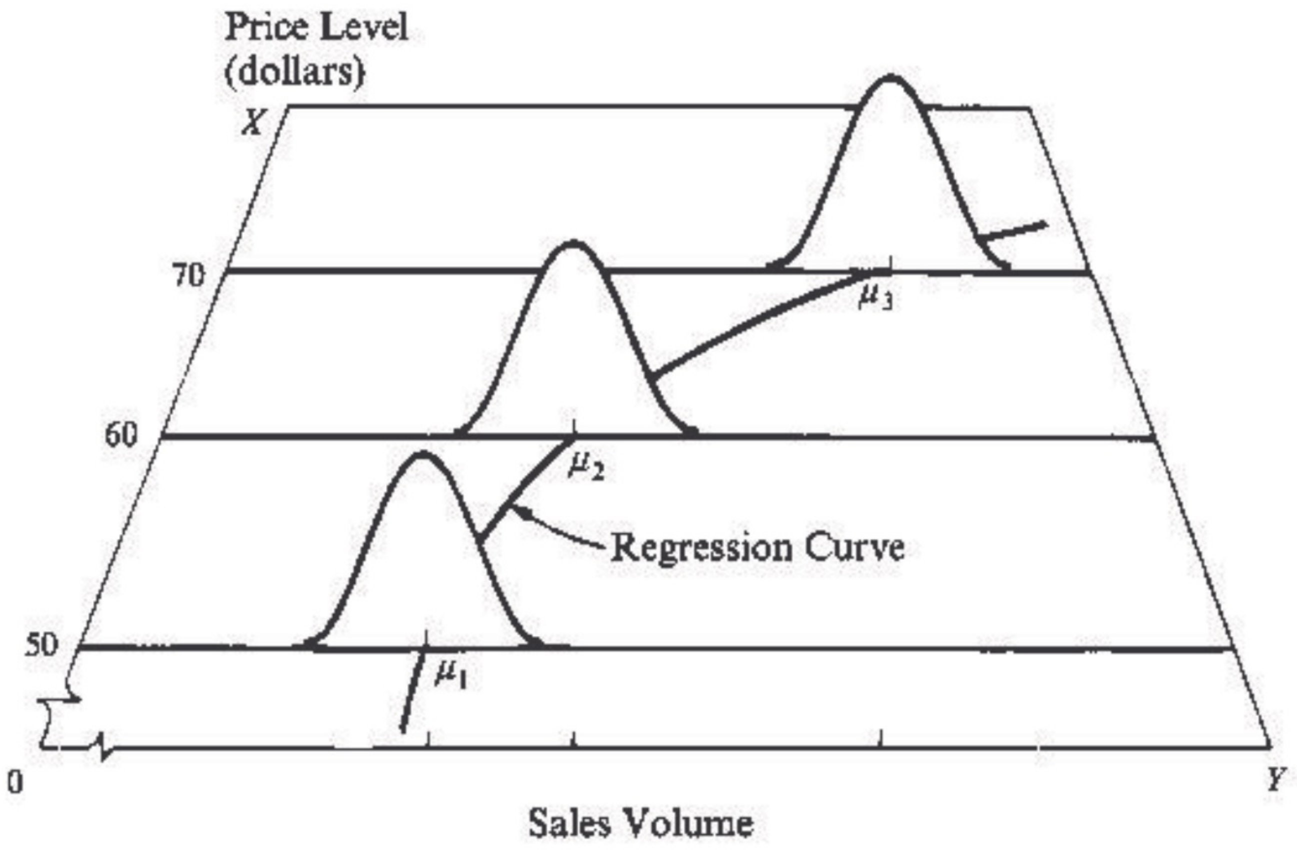


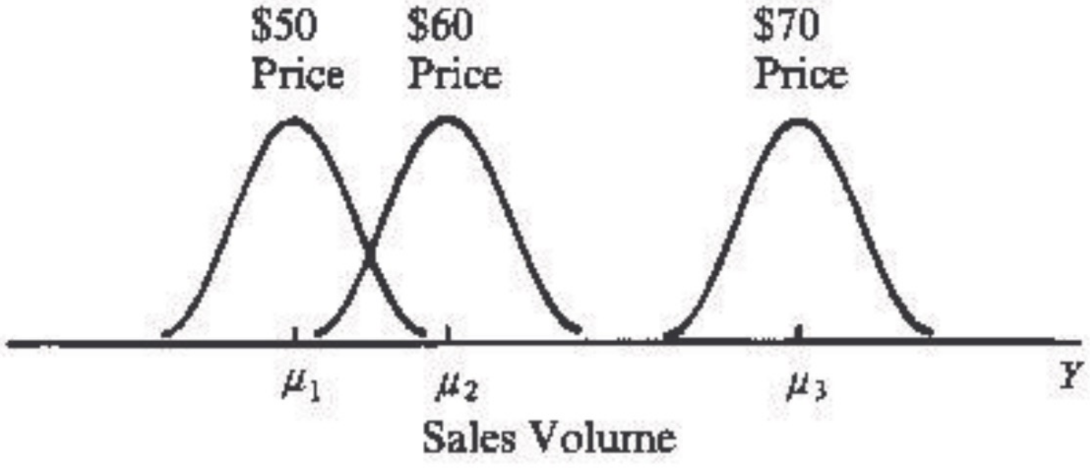
- An **an**alysis **o**f the **va**riation in a set of responses (y_i)
- Each independent variable explains a portion of the total variation in y_i

Definitions

- **Factor:**
- **Levels:**
- **Fixed factor:**
- **Random factor:**
- **Qualitative:**
- **Quantitative:**
- **Balanced:**

Regression vs. ANOVA





Example:

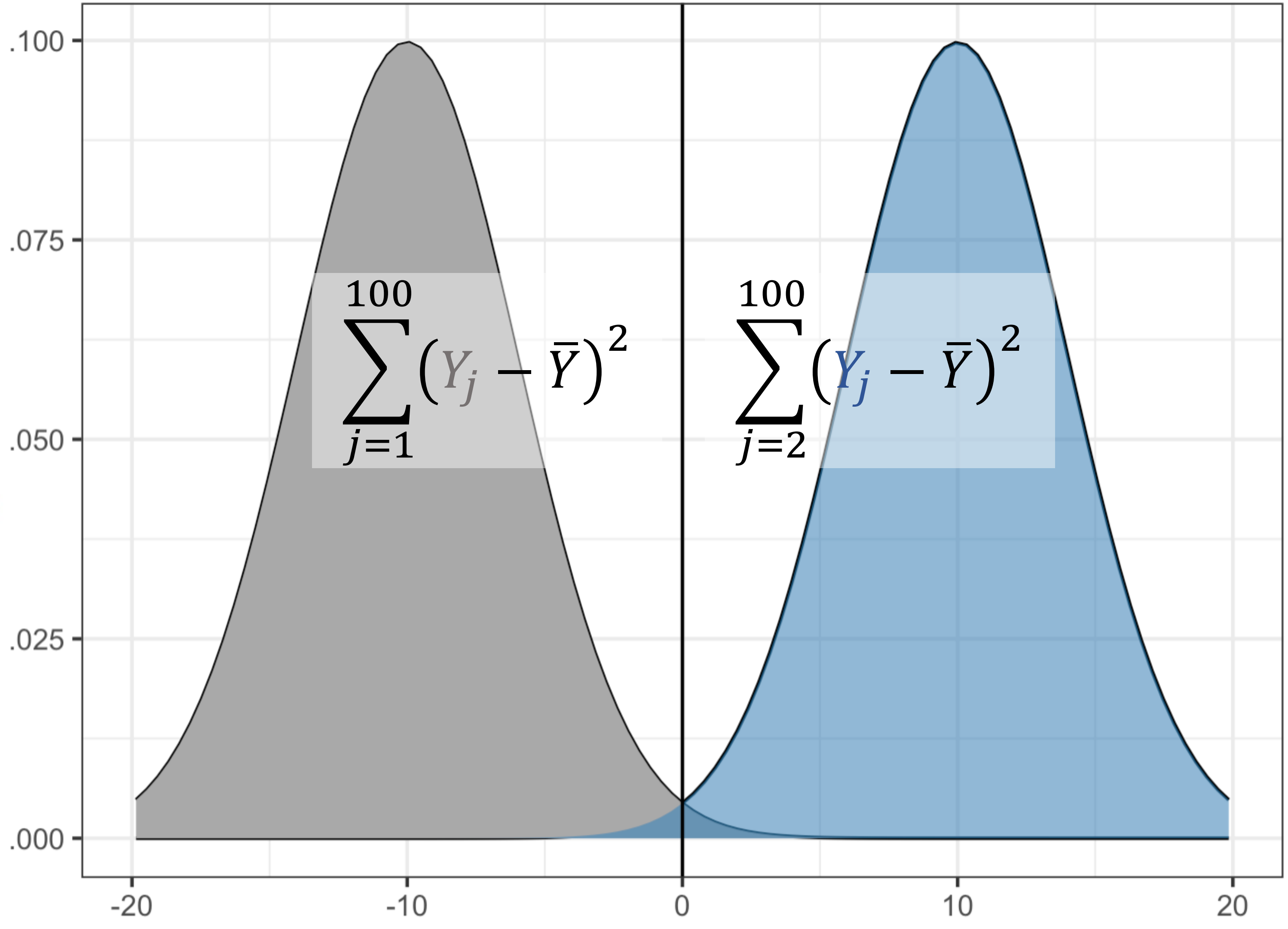
- $N(-10, 4); n_1 = 100$
- $N(10, 4); n_2 = 100$

Total SS = $\sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$

```
total_ss <- sum(((y[1:100] - mean(y))^2),
               ((y[101:200] - mean(y))^2))
```

n_1
Sample from population 1:
$N(\mu_1, \sigma)$

n_2
Sample from population 2:
$N(\mu_2, \sigma)$



$$\sum_{j=1}^{100} (Y_j - \bar{Y})^2$$

$$\sum_{j=2}^{100} (Y_j - \bar{Y})^2$$

Example:

- $N(-10, 4); n_1 = 100$
- $N(10, 4); n_2 = 100$

Total SS = $\sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}\right)^2$

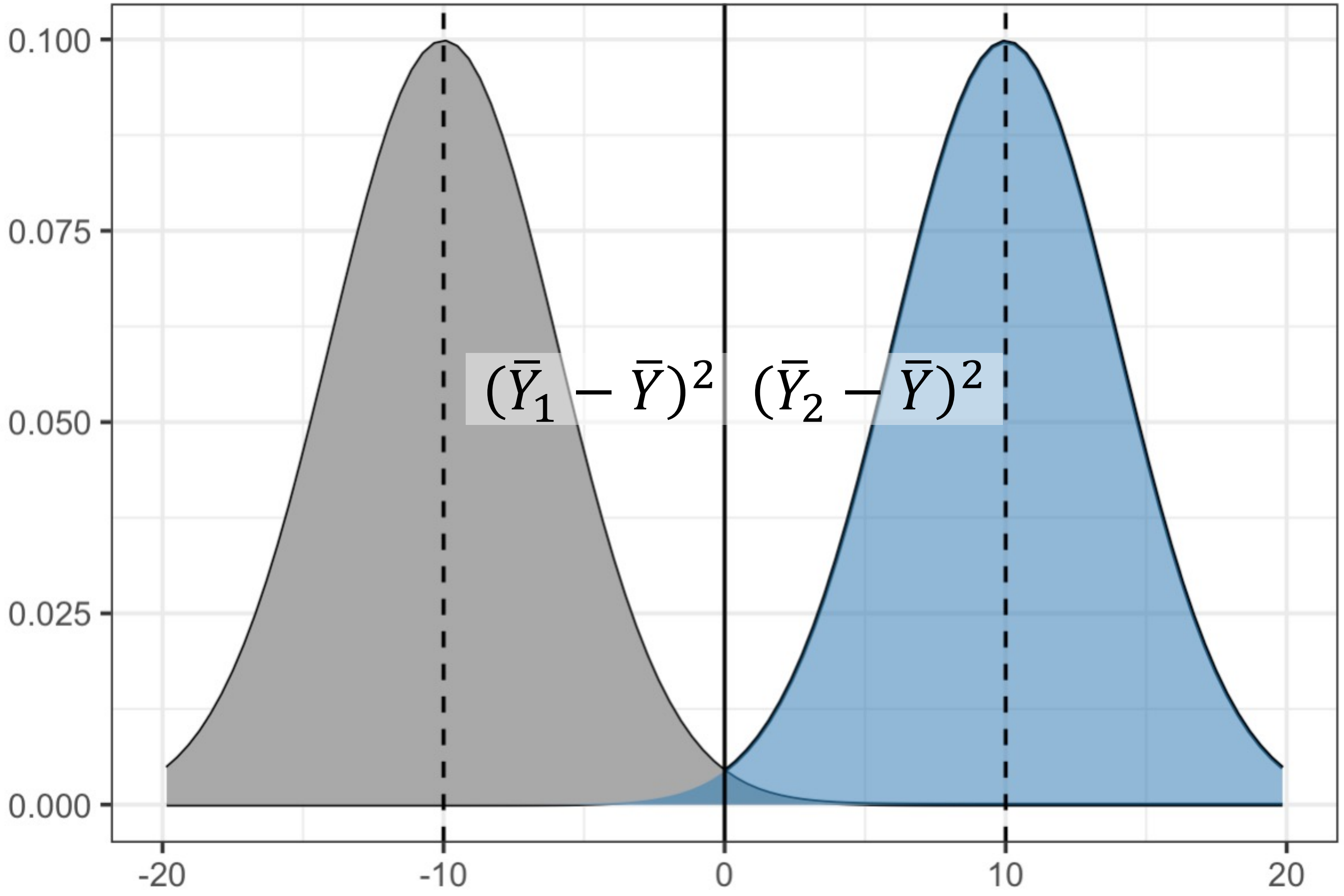
```
total_ss <- sum(((y[1:100] - mean(y))^2),
                ((y[101:200] - mean(y))^2))
```

SST = $n_1 \sum_{i=1}^{2} (\bar{Y}_i - \bar{Y})^2$

```
ss_t <- 100 * sum(c((mean(y[1:100]) - mean(y))^2,
                    (mean(y[101:200]) - mean(y))^2))
```

n_1
Sample from population 1:
$N(\mu_1, \sigma)$

n_2
Sample from population 2:
$N(\mu_2, \sigma)$



Example:

- $N(-10, 4); n_1 = 100$
- $N(10, 4); n_2 = 100$

Total SS = $\sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}\right)^2$

```r
total_ss <- sum(((y[1:100] - mean(y))^2),
                ((y[101:200] - mean(y))^2))
```

SST = $n_1 \sum_{i=1}^{2} (\bar{Y}_i - \bar{Y})^2$

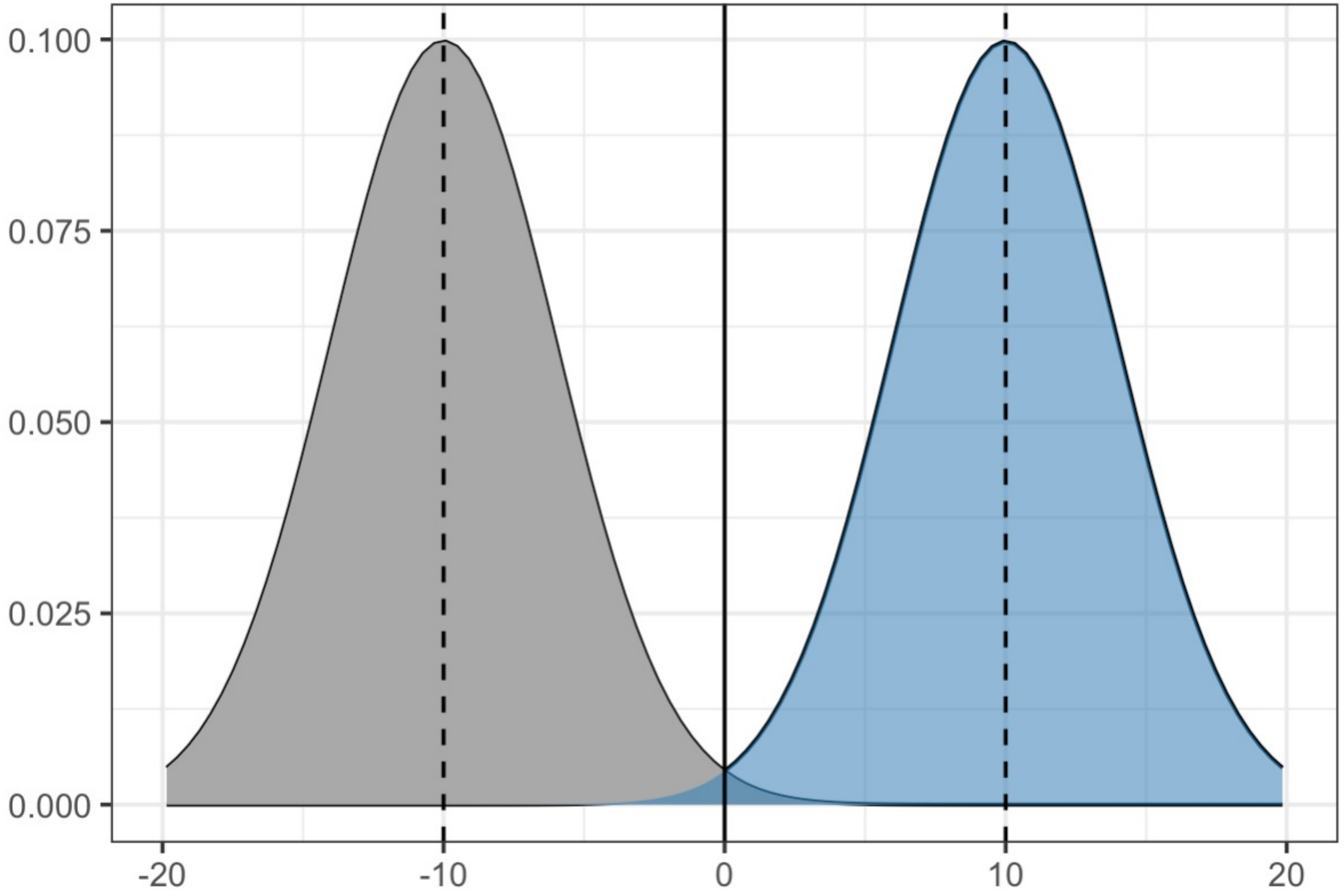
```r
ss_t <- 100 * sum(c((mean(y[1:100]) - mean(y))^2,
                    (mean(y[101:200]) - mean(y))^2))
```

SSE = $\sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2$

```r
ss_e <- sum(((y[1:100] - mean(y[1:100]))^2),
            ((y[101:200] - mean(y[101:200]))^2))
```

n_1
Sample from population 1: $N(\mu_1, \sigma)$

n_2
Sample from population 2: $N(\mu_2, \sigma)$



Example, continued:

- SSE can also be written as:

$$SSE = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2$$

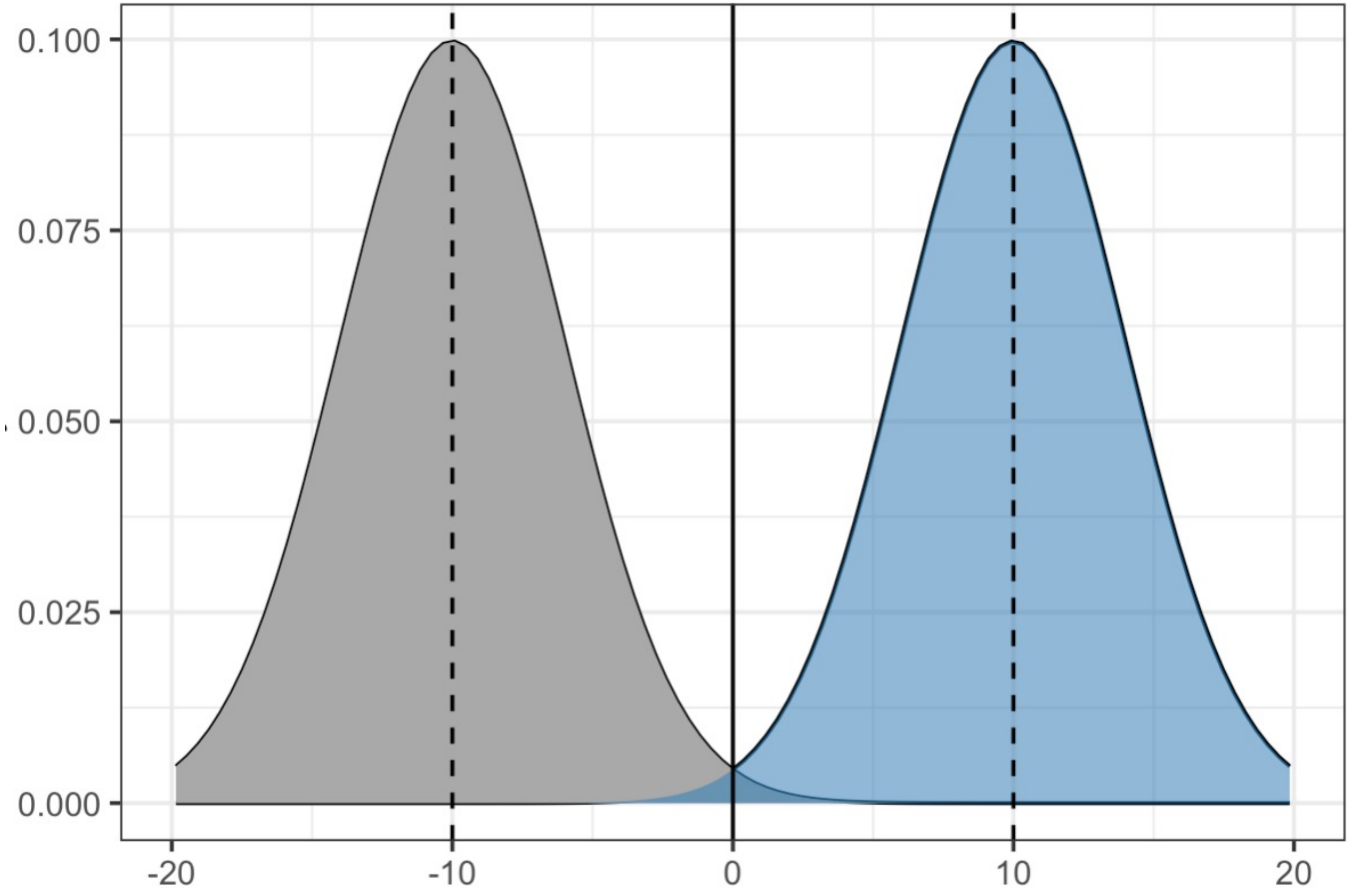
$$= (n_1 - 1)S_1^2 + (n_1 - 1)S_2^2$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_1 - 1)S_2^2}{2n_1 - 2} = \frac{SSE}{2n_1 - 2}$$

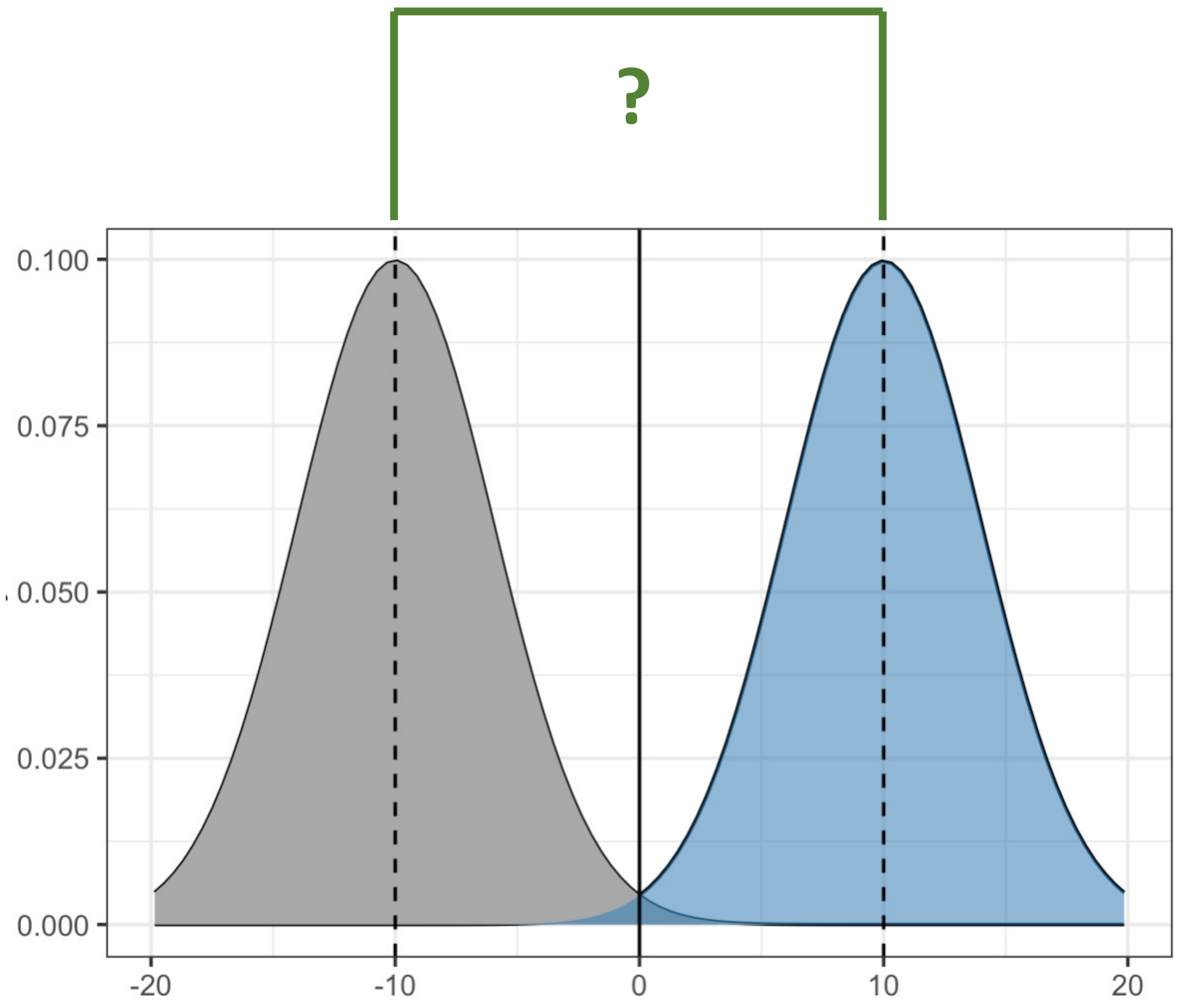
- And here, SST is:

$$SST = n_1 \sum_{i=1}^{2} (\bar{Y}_i - \bar{Y})^2$$

$$= \frac{n_1}{2} (\bar{Y}_1 - \bar{Y}_2)^2$$



How large does SST need to be?



- Since Y_{ij} is normally distributed with $E\left[Y_{ij}\right] = \mu_i$ for $i = 1, 2$, and $V\left(Y_{ij}\right) = \sigma^2$
- And because $E\left[\frac{SSE}{2n_1 - 2}\right] = \sigma^2$:

$$\frac{SSE}{\sigma^2} = \sum_{j=1}^{n_1} \frac{\left(Y_{1j} - \bar{Y}_1\right)^2}{\sigma^2} + \sum_{j=1}^{n_1} \frac{\left(Y_{2j} - \bar{Y}_2\right)^2}{\sigma^2}$$

$$\frac{SSE}{\sigma^2} \sim \chi^2_{(2n_1 - 2)}$$

- We'll demonstrate later: $\frac{SST}{\sigma^2} \sim \chi^2_{(1)}$