Definition. *The goal of* ***regression*** *is to summarize observed data as simply, usefully, and elegantly as possible.* ***Simple Linear Regression*** *has a summarizing model*

$$\mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x_i$$

$$V(Y|X = x) = \sigma^2$$

while making some error assumptions.

Definition. ***parameter*** *a population quantity*
statistic *a quantity based on a sample drawn from the population*

Definition. ***Central limit theorem*** *if* $X_1, X_2, \cdots$ *is an independent sequence of identically distributed random variables with mean* $\mu = \mathbb{E}(X_i)$ *and variance* $\sigma^2 = V(X_i)$ *then*

$$\lim_{n\to\infty} \P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq x\right) = \phi(x)$$

where $\overline{x} = \sum_i X_i/n$ *and* $\phi(x)$ *is standard normal CDF*

Definition. ***Relationship between normal and*** χ^2 ***distribution*** *Let* $X_1, \cdots \sim \mathcal{N}(\mu, \sigma^2)$ *be independent, then distribution of sample variance* $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$

$$S^2 \sim \frac{\sigma^2}{(n-1)}\chi^2_{n-1}$$

Definition. ***t distribution***

Definition. ***F distribution***

Definition. *A linear regression model of mortality versus temperature is by estimating intercept* β_0 *and slope* β_1

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

for $i \in \{1, \cdots, n\}$ *and* $\epsilon \sim \mathbb{N}(0, \sigma^2)$. *Try to find least-square estimators* β_0 *and* β_1 *tha minimize the sum of squares*

$$\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$$

The solution is given by

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = r\frac{S_Y}{S_X}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$