Simple Linear Regression

- Or, linear regression with one predictor variable.
- All the big questions about relationships!
- Setup:
  - Y: Response variable
    (a.k.a., outcome variable, dependent variable).
  - X: Predictor variable
    (a.k.a., covariates, independent variable).
- Simple.
- Linear.

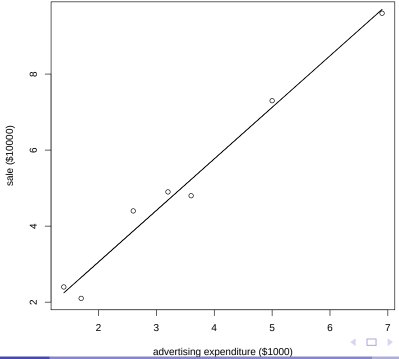
Simple Linear Regression

- Observed data: $(X_i, Y_i), i = 1, \dots, n$.
  - Observational data.
  - Experimental data.
  - Caution: ____________________
- Uses:
  - Description.
  - Control (e.g., effective dose level).
  - Prediction (e.g., disease risks).
- Scope of the model:
  - By design.
  - By data range.
  - Caution: ____________________

Data Example: **advertising**

From a random sample of 7 small businesses, the following data were collected on X, advertising expenditures (\$1000) and Y, sales (\$10000).

Company	A	B	C	D	E	F	G
X	3.2	1.4	2.6	6.9	3.6	1.7	5.0
Y	4.9	2.4	4.4	9.6	4.8	2.1	7.3



Objectives

- The question of interest is: What is the relationship between the amount of advertising expenditures (X) and the sales (Y)?
  - The relationship is apparently positive.
  - The relationship appears to be linear.
  - How to quantify or assess linear relationship?
- In general, the objectives are to:
  - Describe the relationship between X and Y.
  - Estimate the expected Y for a given value of X.
  - Predict a new Y for a given value of X.

Linear Line Fit

- The main idea behind simple linear regression is to fit data with a straight line:
- Recall equation for a straight line $y = mx + b$.
  - Here β_0 is an intercept and β_1 is a slope (rise/run).
- We will discuss the statistical model later.
- An immediate goal is to find β_0, β_1 for the best fitting line.
  - The approach is least squares.

Method of Least Squares

- Find β_0, β_1 that minimize the sum of squares:

  where Y_i is the observed value.

- We will show the best fitting line has slope and intercept:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
&= \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}(\sum_{i=1}^{n} X_i)^2}, \\
\hat{\beta}_0 &=
\end{aligned}$$

Method of Least Squares

- In the advertising example,

$$\sum_{i=1}^{n} X_i = 24.40, \sum_{i=1}^{n} X_i^2 = 107.42, \sum_{i=1}^{n} X_i Y_i = 154.07$$
$$\sum_{i=1}^{n} Y_i = 35.50, \sum_{i=1}^{n} Y_i^2 = 222.03.$$

- Thus

$$\hat{\beta}_1 = \frac{154.07 - 24.40 \times 35.50/7}{107.42 - 24.4 \times 24.40/7} = \frac{30.33}{22.37} = 1.356$$
$$\hat{\beta}_0 = 35.50/7 - 1.356 \times 24.40/7 = 0.345$$

- We may also predict Y at say $X = x_0$ for $x_0 \in \mathbb{R}^+$,

$$\hat{Y} = 0.345 + 1.356 x_0$$

Simple Linear Regression Model

- How to account for uncertainty in the fitted line and variation?
  - Model Y as a random variable.
  - Regard X as fixed, although X could be random.
  - Consider the model of Y conditional on X such that

    where β_0 and β_1 are fixed unknown parameters representing intercept and slope, characterizing the relationship between X and Y.
- The model is called **"linear"** because it is linear in the *parameters*, **not because it is linear in X**

Formal Statement of Model

The formal simple linear regression model is:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

for $i = 1, \ldots, n$.

- Y_i is the value of the ith response variable.
- X_i is the known value of the ith predictor variable.
- ϵ_i is the ith random error component such that

$$\epsilon_i \sim \text{i.i.d. } N(0, \sigma^2).$$

- i.i.d. means "independent, identically distributed"

Assumptions of Random Error $\{\epsilon_i\}$

1. Mean $E(\epsilon_i) = 0$.
2. Variance $Var(\epsilon_i) = \sigma^2$.
3. ϵ_i and ϵ_j are uncorrelated with $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$.
4. ϵ_i follows a normal distribution.
   - Actually, normality is not required to justify least squares. Independence with mean zero suffices. However, normality is needed to permit maximum likelihood estimation of β_0, β_1

Implications of Assumptions on $\{Y_i\}$

- Y_i is random variable with (conditional on X):
  1. mean of Y

     $E(Y_i) \quad =$

  2. variance of Y:

     $Var(Y_i) \quad =$

  3. correlation of Y_i and Y_j

     $Cov(Y_i, Y_j) =$

  4. $Y_i \sim$

Model Parameters

- The model parameters are β_0, β_1, and σ^2.
- β_0 and β_1: Regression coefficients (population parameters).
- β_0: Intercept. When the model scope includes $X = 0$, β_0 can be interpreted as the mean of Y at $X = 0$.
- β_1: Slope. Interpreted as the change in the mean of Y per unit increase in X.
- σ^2: Variance of the error term.

Estimation of Model Parameters

- Our goal is to estimate these model parameters by estimators $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\sigma}^2$, based on data.
- Two methods:
  - Least squares (LS).
  - Maximum likelihood (ML).
- Additional notation:
  - Let $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ denote the ith fitted value.
  - Let $e_i = Y_i - \hat{Y}_i$ denote the ith residual.

Estimation of β_0 and β_1

- Both LS and ML give the same estimator for β_0 and β_1:
  - Why?

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

- Note that $\hat{\beta}_0$ and $\hat{\beta}_1$ are denoted as b_0 and b_1 in KNNL.

Least Squares (LS) Estimation

- Consider the criterion:

- The LS estimators of β_0 and β_1 are those values, $\hat{\beta}_0$ and $\hat{\beta}_1$, that minimize Q, for the given observed data $(X_1, Y_1), \ldots, (X_n, Y_n)$.

LS Derivation

- Differentiate Q with respect to β_0 and β_1:

- Set above gradient equal to 0 and let the solutions to these two equations be $\hat{\beta}_0$ and $\hat{\beta}_1$.
- Let $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Since $\frac{\partial^2 Q}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ is positive definite, $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize Q.
  - Why?

LS Derivation

- Let the gradient equal to zero, we obtain a set of **normal equations**:

- That is,

LS Derivation

- From (1),

$$\hat{\beta}_0 =$$

- Plugging $\hat{\beta}_0$ into (2), we obtain

- Thus,

$$\hat{\beta}_1 =$$

Gauss-Markov Theorem

Under the simple linear regression model (**without having to assume normality**), the LS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are **unbiased** and have **minimum variance among all linear unbiased estimators** of β_0 and β_1.

This result will be proven later.

Properties of $\hat{\beta}_0$ and $\hat{\beta}_1$ by Gauss-Markov Theorem

1. $E(\hat{\beta}_0) =$

   $E(\hat{\beta}_1) =$

2. $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear estimators because

3. The variances are the minimal

Estimation of σ^2

- LS and ML give **different estimators** for σ^2.
- Define an error sum of squares (SSE):
- The LS estimate of σ^2 is the error mean square (MSE):
- The ML estimate of σ^2 is:
- Which estimator is better?

Maximum Likelihood (ML) Estimation

- In a general setting, let $Y_1, \ldots, Y_n \sim$ iid with probability density function $f(y; \boldsymbol{\theta})$.
- With $\boldsymbol{y} = (y_1, \ldots, y_n)'$, the likelihood function for $\boldsymbol{\theta}$ is

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}).$$

- Find the value of $\boldsymbol{\theta}$ that maximizes $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$.
- Intuitively, one would want to find the value of $\boldsymbol{\theta}$ that makes the data most likely.
- Equivalently, find the value of $\boldsymbol{\theta}$ that maximizes the log-likelihood

$$\ell(\boldsymbol{\theta}; \boldsymbol{y}) = \log \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \log \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\theta}).$$

ML Derivation

- Let $\boldsymbol{\theta} =$
- Since $Y_i \sim$, we have $f(y_i; \boldsymbol{\theta})$
- Thus the likelihood function is $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$

  and the log-likelihood function $\ell(\boldsymbol{\theta}; \boldsymbol{y})$ is

ML Derivation

- Set the first-order partial derivatives equal to 0:

$$
\begin{aligned}
0 &= \frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \beta_0} = \frac{2}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 X_i) \\
0 &= \frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \beta_1} = \frac{2}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 X_i) X_i \\
0 &= \frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 X_i)^2.
\end{aligned}
$$

- Solve for the parameters and obtain the ML estimates:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \\
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
\hat{\sigma}^2 &= \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n}
\end{aligned}
$$

- Note that $E(\hat{\sigma}^2) = \frac{n-2}{n} \sigma^2$.

Properties of Fitted Regression Line

- $\sum_{i=1}^{n} e_i = 0$.
- $\sum_{i=1}^{n} e_i^2$ is a minimum.
- $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$.
- $\sum_{i=1}^{n} X_i e_i = 0$.
- $\sum_{i=1}^{n} \hat{Y}_i e_i = 0$.
- The regression line always goes through $(\bar{X}, \bar{Y})$.