1 W1

1.1 Lecture 1.

1.1.1 Review

Mean/Expectation

The **mean** or **expectation** of a continuous random variable Y is given by

$$\mathbb{E}[Y] = \int y f(y)\, dy$$

For random variables $Y_1, \ldots, Y_m$ and constants a_i, b_i for $i = 1, \ldots, m$,

$$\mathbb{E}\left[\sum_{i=1}^{m}(a_i Y_i + b_i)\right] = \sum_{i=1}^{m} a_i \mathbb{E}[Y_i] + \sum_{i=1}^{m} b_i.$$

This is called the **linearity of expectation**. For observations $y_1, \ldots, y_n$, the **sample mean** is

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Variance

The **variance** of a continuous random variable Y is given by

$$\mathrm{Var}[Y] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}^2[Y]$$

- For constants $a, b \in \mathbb{R}$, $\mathrm{Var}[aY + b] = a^2 \mathrm{Var}[Y]$.
- If X and Y are *independent*, then $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

For observations $y_1, \ldots, y_n$, the **sample variance** is

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

Covariance

The **covariance** of two continuous random variables X, Y is given by

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
- $\mathrm{Cov}(aY + c, bX + d) = ab \cdot \mathrm{Cov}(X, Y)$.
- $\mathrm{Cov}(U + V, X + Y) = \mathrm{Cov}(U, X) + \mathrm{Cov}(U, Y) + \mathrm{Cov}(V, X) + \mathrm{Cov}(V, Y)$.
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.

For observations $(y_1, x_1), \ldots, (y_n, x_n)$, the **sample covariance** is

$$\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}).$$

Correlation Coefficient

Correlation coefficient:

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \in [-1,1].$$

Sample correlation:

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^n (y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (y_i-\bar{y})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}.$$

Correlation measures the strength of linear relationship.

Normal Distribution

Suppose $Z \sim N(\mu, \sigma^2)$. Then

$$\mathbb{E}[Z] = \mu \qquad \mathrm{Var}(Z) = \sigma^2$$

For independent $Z_i \sim N(\mu_i, \sigma^2)$, $U = \sum_{i=1}^n (a_i Z_i + b_i)$ is normally distributed, i.e.,

$$U \sim N\left(\sum_{i=1}^n (a_i\mu_i + b_i), \sum_{i=1}^n a_i^2\sigma_i^2\right).$$

Chi-Square Distribution

Let ν denote the degrees of freedom. Suppose $X \sim \chi^2_\nu$. Then

$$\mathbb{E}[X] = \nu \qquad \mathrm{Var}(X) = 2\nu.$$

For standard normal random variables $Z_i \overset{\text{iid}}{\sim} N(0,1)$,

$$X = \sum_{i=1}^n Z_i^2 \sim \chi^2_n.$$

t-Distribution

Let ν denote the degrees of freedom. Suppose $Y \sim t_\nu$. Then

$$\mathbb{E}[Y] = 0 \quad \text{if } \nu > 1, \text{otherwise NA} \qquad \mathrm{Var}[Y] = \frac{\nu}{\nu - 2} \quad \text{if } \nu > 2, \text{otherwise } \infty$$

For *independent* $Z \sim N(0,1)$ and $X \sim \chi^2_\nu$,

$$\frac{Z}{\sqrt{X/\nu}} \sim t_\nu.$$

1.1.2 Motivation: Toward Linear Regression

Simple Linear Regression

- How do we characterize the relationship between x and y?
- How do we predict y given x?
- How does the mean of y change when x increases by a?

We can answer questions like these with **simple linear regression**.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

Intuitively, we are assuming that there exists some underlying linear relationship between the covariates x and the observations y, where β_0 and β_1 are unknown:

$$y \approx \beta_0 + \beta_1 x.$$

The error term ε captures the difference between the actual y and the predicted $\beta_0 + \beta_1 x$.

Multiple Linear Regression

What if we have multiple covariates? Suppose each sample x_i has three covariates x_{i1}, x_{i2}, x_{i3}. We can generalize the simple linear regression to **multiple linear regression**:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i,$$

Note that each covariate x_{ij} has a corresponding β_j parameter.

Course Outlook

This course will focus on developing multiple linear regression:

- Theoretically/mathematically: derive estimators.
- Practically: how to fit these models in R.
- How to choose and compare a model, i.e., which x_{ij} to include;
- How to evaluate the appropriateness of the model and assumptions

1.2 Lecture 2. Simple Linear Regression: Estimation.

In this lecture, we formally define the simple linear regression model and derive estimators for the regression parameters.

1.2.1 Simple Linear Regression

The general form of simple linear regression is given below:

$$\boxed{y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2).}$$

- $\beta_0, \beta_1, \sigma^2$: fixed, unknown parameters.
- ε_i: unobserved random error term.
- y_i, x_i are observed data.
  - Treat x_i as fixed in this course.

Equivalently, we can write

$$\boxed{y_i \overset{\text{indep}}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2).}$$

Note here y_i's are independent but no longer have the same distribution because they have different means (depending on x_i).

Observations

- $\mathbb{E}[y_i \mid x_i] = \beta_0 + \beta_1 x_i$.
- $\mathbb{E}[y_i \mid x_i = 0] = \beta_0$.
- $\mathbb{E}[y_i \mid x_i = x^*] = \beta_0 + \beta_1 x^*$.
- $\mathbb{E}[y_i \mid x_i = x^* + 1] = \beta_0 + \beta_1(x^* + 1) = \beta_0 + \beta_1 x^* + \beta_1$.

Assumptions: LINE

- **L**inearity: there is a linear relationship between x and y.
- **I**ndependence: the error terms are independent.
- **N**ormality: the error terms have mean 0.
- **E**qual variance (homoskedasticity): all error terms share the same variance.

1.2.2 Estimating Parameters with Least Squares

We wish to minimize the sum of least squares. We will later show that the resulting estimators match the ones derived using maximum likelihood.

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

For β_0:

$$\begin{aligned}
0 = \frac{\partial S(\delta_0, \delta_1)}{\partial \beta_0} &= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-1) \\
0 &= \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \\
0 &= \sum_{i=1} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i \\
\beta_0 &= \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - \beta_1 \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \bar{y} - \beta_1 \bar{x}.
\end{aligned}$$

Now for β_1:

$$\begin{aligned}
0 = \frac{\partial S(\delta_0, \delta_1)}{\partial \beta_1} &= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-x_i) \\
0 &= \sum_{i=1} y_i x_i - \beta_0 x\bar{x} - \beta_1 \sum_{i=1}^{n} x_i^2 \\
0 &= \sum_{i=1}^{n} y_i x_i - (\bar{y} - \beta_1 \bar{x}) n\bar{x} - \beta_1 \sum_{i=1}^{n} x_i^2 \\
0 &= \sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x} + \beta_1 n\bar{x}^2 - \beta_1 \sum_{i=1}^{n} x_i^2 \\
0 &= \sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x} + \beta_1 \left(n\bar{x}^2 - \sum_{i=1}^{n} x_i^2\right) \\
\beta_1 &= \frac{\sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}.
\end{aligned}$$

Rewrite the numerator:

$$\sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x} = \sum_{i=1}^{n} y_i x_i - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\bar{x} = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i \bar{x} = \sum_{i=1}^{n} y_i(x_i - \bar{x}).$$

Now observe

$$
\begin{aligned}
\sum_{i=1}^{n} \bar{y}(x_i - \bar{x}) &= \bar{y} \sum_{i=1}^{n} (x_i - \bar{x}) \\
&= \bar{y} \left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \right) \\
&= \bar{y} \left(\sum_{i=1}^{n} x_i - n\bar{x} \right) \\
&= \bar{y} \left(\sum_{i=1}^{n} x_i - \cancel{n} \frac{1}{\cancel{n}} \sum_{i=1}^{n} x_i \right) = 0.
\end{aligned}
$$

Thus, we have

$$
\sum_{i=1}^{n} y_i(x_i - \bar{x}) = \sum_{i=1}^{n} y_i(x_i - \bar{x}) - \underbrace{\sum_{i=1}^{n} \bar{y}(x_i - \bar{x})}_{0} = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}).
$$

A similar derivation (for the denominator) gives

$$
\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2.
$$

Therefore, we get

$$
\beta_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.
$$

To summarize, the **least squares estimators** are

$$
\boxed{
\begin{aligned}
\hat{\beta}_1^{\mathrm{LS}} &= \frac{S_{xy}}{S_{xx}} \\
\hat{\beta}_0^{\mathrm{LS}} &= \hat{y} - \hat{\beta}_1^{\mathrm{LS}} \bar{x}
\end{aligned}
}
$$

where

$$
\begin{aligned}
S_{xy} &= \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) \\
S_{xx} &= \sum_{i=1}^{n} (x_i - \bar{x})^2.
\end{aligned}
$$

1.2.3 Estimating Parameters with Least Squares

Starting from

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

the log-likelihood is given by (steps omitted)

$$\ell(\beta_0, \beta_1, \sigma^2 \mid y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_i])^2.$$

Maximizing the log-likelihood is equivalent to solving the following system of equations:

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta_0} &= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_i])\right) = 0 \\
\frac{\partial \ell}{\partial \beta_1} &= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_i])\right) x_i = 0 \\
\frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\left(\sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_i])\right)^2 = 0
\end{aligned}$$

Observe solving the first two equations is equivalent to minimizing the sum of squares! In other words, the estimators from MLE are equal to the those from LS. Thus, we will call the LS/MLE estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ (i.e., without the superscripts).

Solving the third equation, we get

$$\hat{\sigma}^2_{\text{ML}} = \frac{\sum_{i=1}^{n} e_i^2}{n}.$$

However we typically adopt a different estimator:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}.$$

Note this difference often doesn't matter when $n \geq 50$.

1.2.4 Fitted Values and Residuals

The **fitted values** $\hat{y}_i$ and the **residuals** e_i are given by

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.
- $e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$.

Note the residuals e_i and the errors $\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$ are not the same thing.