Probabilistic modeling

To model processes that have any kind of uncertainty, we use probabilistic or random modeling.

The idea is to make the response variable a random variable with distribution parametrized by the predictor variables.

A very common choice is $Y \sim \mathcal{N}(f(x), \sigma^2)$, for some function f taken from some specified family.

Equivalently, this can be expressed as

$$\begin{cases} Y = f(X) + \varepsilon \\ \varepsilon \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

As usual, once a random model is established, we can use probability theory to make predictions or to analyze the resulting distribution.

Linear modeling

Our focus here will be on linear models:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon$$
$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Linearity here refers to the coefficients β_i — the predictors can involve more complicated functions of variables — eg:

$$\beta_0 + \beta_1 t + \beta_2 e^t + \varepsilon \quad \text{or} \quad \beta_0 + \beta_1 x + \beta_2 \sin x \quad \text{etc.}$$

Throughout, we will assume $\mathbb{E}\varepsilon = 0$. If this were not the case, we could replace β_0 with $\beta_0 - \mathbb{E}\varepsilon$ to change to a model where it is true.

Simple vs. Multiple

This week the focus is on simple linear regression where the model is

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

and next week we will look closer at multiple regression

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon.$$

Least Squares Estimation

Given observations $(x_1, y_1), \dots, (x_n, y_n)$, our goal is to estimate β_0 and β_1 such that $\hat\beta_0 + \hat\beta_1 x_i \approx y_i$ with as small an error as possible.

If we choose to measure error by sum of square deviations, we can produce explicit and easy to use estimators!

Write $\hat y_i = \hat\beta_0 + \hat\beta_1 x_i$ for the predicted value.
Then $y_i - \hat y_i$ is the deviation (or error, or residual) and our error measure is

$$SSE = \sum (y_i - \hat y_i)^2 = \sum \left(y_i - (\hat\beta_0 + \hat\beta_1 x_i)\right)^2.$$

To find a minimum, we look to

$$\frac{\partial SSE}{\partial \hat\beta_0} = \sum 2(y_i - (\hat\beta_0 + \hat\beta_1 x_i)) \cdot (-1)$$

$$= -2\left(\sum y_i - n\hat\beta_0 - \hat\beta_1 \sum x_i\right).$$

$= 0$ when $\hat\beta_0 = \bar y - \hat\beta_1 \bar x$.

$$\frac{\partial SSE}{\partial \hat\beta_1} = \sum 2(y_i - (\hat\beta_0 + \hat\beta_1 x_i)) \cdot (-x_i) = -2\left(\sum x_i y_i - \hat\beta_0 \sum x_i - \hat\beta_1 \sum x_i^2\right)$$

$= 0$ when $\hat\beta_1 \sum x_i^2 = \sum x_i y_i - \hat\beta_0 \sum x_i = \sum x_i y_i - n\bar y \bar x + \hat\beta_1 \cdot n \cdot \bar x^2$

So $\hat\beta_1 (\sum x_i^2 - n\bar x^2) = \sum x_i y_i - n \bar x \bar y$

So $\hat\beta_1 = (\sum x_i y_i - n\bar x \bar y) / (\sum x_i^2 - n\bar x^2) = \sum (x_i - \bar x)(y_i - \bar y) / \sum (x_i - \bar x)^2.$

To simplify all formulas, we will adopt the convention

$$S_{ab} = \sum (a_i - \bar{a})(b_i - \bar{b})$$

so that

$$\hat{\beta}_1 = S_{xy} / S_{xx}$$

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

These are the <u>least</u> squares <u>estimators</u> for the simple linear model.

No-intercept model

Consider the model $Y_i = \beta_1 x_i + \varepsilon_i$, where ε_i are iid $E=0$.

$$SSE = \sum \left[y_i - \hat{\beta}_1 x_i \right]^2$$

So $\dfrac{\partial SSE}{\partial \hat{\beta}_1} = \sum 2 \cdot (y_i - \hat{\beta}_1 x_i) \cdot (-x_i) = -2\left(\sum x_i y_i - \hat{\beta}_1 \sum x_i^2 \right)$

$= 0$ when $\hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$

or $\hat{\beta}_1 = \sum x_i y_i / \sum x_i^2$

Properties of $\hat\beta_0$ and $\hat\beta_1$.

$\hat\beta_0$ and $\hat\beta_1$ are unbiased. When $\varepsilon \sim N(0, \sigma^2)$, then their sampling distributions are also normal.

We will now prove these assertions and calculate $\mathbb{V}\hat\beta_0$ and $\mathbb{V}\hat\beta_1$.

Claim $\hat\beta_1$ is unbiased.

We assume $\mathbb{V}\varepsilon$ does not depend on x (homoscedasticity). Assume we make n observations, ie we have samples

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

$$\hat\beta_1 = S_{xy}/S_{xx} = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) Y_i - \bar{Y} \sum (x_i - \bar{x})}{S_{xx}}$$

$$= \frac{\sum (x_i - \bar{x}) Y_i}{S_{xx}} \quad \text{since } \sum (x_i - \bar{x}) = 0.$$

Now, $$\mathbb{E}\hat\beta_1 = \frac{\sum (x_i - \bar{x}) \mathbb{E} Y_i}{S_{xx}} = \frac{\sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{S_{xx}} =$$

$$= \beta_0 \frac{\sum (x_i - \bar{x})}{S_{xx}} + \beta_1 \frac{\sum (x_i - \bar{x}) \cdot x_i}{S_{xx}}$$

By the same argument as above, $\sum (x_i - \bar{x}) x_i = \sum (x_i - \bar{x})^2$. and $\sum (x_i - \bar{x}) = 0$. Hence $\mathbb{E}\hat\beta_1 = \beta_1$. ■

$$\mathbb{V}\hat{\beta}_1 = \mathbb{V}\left[\frac{\sum(x_i-\bar{x})Y_i}{S_{xx}}\right] = \frac{1}{S_{xx}^2}\sum \mathbb{V}[(x_i-\bar{x})Y_i] =$$

$$= \frac{1}{S_{xx}^2}\sum(x_i-\bar{x})^2\,\mathbb{V}Y_i = \frac{S_{xx}}{S_{xx}^2}\cdot\sigma^2 = \sigma^2/S_{xx}.$$

because Y_i independent

<u>Claim</u> $\hat{\beta}_0$ unbiased

$$\mathbb{E}\hat{\beta}_0 = \mathbb{E}[\bar{Y}] - \mathbb{E}[\hat{\beta}_1]\bar{x} = \beta_0 + \overbrace{\beta_1\bar{x} + \underbrace{\mathbb{E}[\bar{\varepsilon}]}_{=0} - \beta_1\bar{x}}^{\text{cancel}} =$$

$$= \beta_0$$

$$\mathbb{V}\hat{\beta}_0 = \mathbb{V}\bar{Y} + \bar{x}^2\,\mathbb{V}\hat{\beta}_1 - 2\bar{x}\,\mathrm{Cov}(\bar{Y},\hat{\beta}_1).$$

$$\mathbb{V}\bar{Y} = \mathbb{V}[\beta_0+\beta_1\bar{x}+\bar{\varepsilon}] = \mathbb{V}\bar{\varepsilon} = \frac{1}{n^2}\sum\mathbb{V}\varepsilon_i = \frac{n}{n^2}\sigma^2 = \sigma^2/n$$

Let's write $c_i = \frac{x_i-\bar{x}}{S_{xx}}$. Then $\hat{\beta}_1 = \sum c_i Y_i$ so

$$\mathrm{Cov}(\bar{Y},\hat{\beta}_1) = \mathrm{Cov}\left(\frac{1}{n}\sum Y_i, \sum c_i Y_i\right) = \sum\sum\frac{c_i}{n}\mathrm{Cov}(Y_i,Y_j) =$$

$=$ only $i=j$ terms survive because Y_i independent $= \sum\frac{c_i}{n}\mathbb{V}Y_i =$

$$= \sum\frac{c_i}{n}\sigma^2 = \frac{\sigma^2}{n}\frac{\sum x_i-\bar{x}}{S_{xx}} = 0. \qquad (!)$$

Hence $\mathbb{V}\hat{\beta}_0 = \mathbb{V}\bar{Y} + \bar{x}^2\,\mathbb{V}\hat{\beta}_1 = \frac{\sigma^2}{n} + \bar{x}^2\left(\frac{\sigma^2}{S_{xx}}\right) = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right) =$

$$= \sigma^2\left(\frac{S_{xx}-n\bar{x}^2}{nS_{xx}}\right) = \frac{\sigma^2\sum x_i^2}{nS_{xx}}.$$

$$Cov(\hat{\beta}_0, \tilde{\beta}_1) = Cov(\bar{Y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1) = \underbrace{Cov(\bar{Y}, \hat{\beta}_1)}_{=0} - \bar{x}\, \mathbb{V}\hat{\beta}_1 = \frac{-\bar{x}\sigma^2}{S_{xx}}$$

<u>Claim</u> We can use SSE to estimate σ^2!

$$\mathbb{E}[SSE] = \mathbb{E}\left[\sum (Y_i - \hat{Y}_i)^2\right] = \mathbb{E}\left[\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right] =$$

$$= \mathbb{E}\left[\sum (Y_i - \bar{Y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i)^2\right] = \mathbb{E}\left[\sum (Y_i - \bar{Y})^2 + \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum (x_i - \bar{x})(Y_i - \bar{Y})\right]$$

Now, $\sum (x_i - \bar{x})(Y_i - \bar{Y}) = \sum (x_i - \bar{x})^2 \hat{\beta}_1$ so the last two terms combine to

$$= \mathbb{E}\left[\sum (Y_i - \bar{Y})^2\right] - \mathbb{E}\left[\hat{\beta}_1 S_{xx}\right] = \mathbb{E}\left[\sum Y_i^2\right] - \mathbb{E}\left[n\bar{Y}^2\right] - \mathbb{E}\left[\hat{\beta}_1 S_{xx}\right]$$

$$= \sum \mathbb{E}[Y_i^2] - n\mathbb{E}[\bar{Y}^2] - S_{xx}\mathbb{E}[\hat{\beta}_1^2] = \text{recall } \mathbb{E}X^2 = \mathbb{V}X + (\mathbb{E}X)^2.$$

$$= \sum \left(\mathbb{V}Y_i + (\mathbb{E}Y_i)^2\right) - n\left[\mathbb{V}\bar{Y} + (\mathbb{E}\bar{Y})^2\right) - S_{xx}\left(\mathbb{V}\hat{\beta}_1 + (\mathbb{E}\hat{\beta}_1)^2\right) =$$

$$= n\sigma^2 + \sum (\beta_0 + \beta_1 x_i)^2 - n\left(\frac{\sigma^2}{n} + (\beta_0 + \beta_1 \bar{x})^2\right) - S_{xx}\left(\frac{\sigma^2}{S_{xx}} + \beta_1^2\right) =$$

$$= n\sigma^2 + \sum (\beta_0 + \beta_1 x_i)^2 - \sigma^2 - n(\beta_0 + \beta_1 \bar{x})^2 - \sigma^2 - \beta_1^2 S_{xx}$$

Consider $\sum (\beta_0 + \beta_1 x_i)^2 - n(\beta_0 + \beta_1 \bar{x})^2 - \beta_1^2 S_{xx} =$

$$= n\beta_0^2 + 2\beta_0\beta_1 \sum x_i + \beta_1^2 \sum x_i^2 - n\beta_0^2 - 2n\beta_0\beta_1\bar{x} - \beta_1^2 \bar{x}^2 - \beta_1^2 \sum (x_i - \bar{x})^2$$

$= 0.$

So $\mathbb{E}[SSE] = (n-2)\sigma^2$ so $S^2 = \frac{SSE}{n-2}$ is unbiased for σ^2.

If all the ε_i are normal, then

① $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear functions of the ε_i.
Hence they too are normally distributed.

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

Writing $c_{ab}\sigma^2 = \operatorname{cov}(\hat{\beta}_a, \hat{\beta}_b)$, we can summarise with

1. $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased
2. $\mathbb{V}\hat{\beta}_0 = c_{00}\sigma^2$ where $c_{00} = \Sigma x_i^2/(nS_{xx})$
3. $\mathbb{V}\hat{\beta}_1 = c_{11}\sigma^2$ where $c_{11} = 1/S_{xx}$
4. $\operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1) = c_{01}\sigma^2$ where $c_{01} = -\bar{x}/S_{xx}$
5. $S^2 = SSE/(n-2)$ is unbiased for σ^2, where $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$

If all ε_i are normal, then

6. $\hat{\beta}_0$ and $\hat{\beta}_1$ are normal
7. $(n-2)S^2/\sigma^2 = SSE/\sigma^2 \sim \chi^2(n-2)$
8. S^2 is independent of both $\hat{\beta}_0$ and $\hat{\beta}_1$.