Math Stats Recitation 11

TA: Sonia Reilly

April 17, 2026

Lecture Review

Lecture 20: Simple Linear Regression

Standard statistical model:

$$y_i = \beta_o + \beta_1 x_i + e_i, \quad \mathbb{E}[e_i] = 0, \quad \mathrm{Var}(e_i) = \sigma^2$$

Task: Given fixed x_i's and the corresponding y_i's, find best fit β_o, β_1 (intercept + slope).
Standard method: minimize squared sum of residuals:

$$\hat{\beta}_o, \hat{\beta}_1 = \arg\min_{\beta_o, \beta_1} \sum_i (y_i - \beta_o - \beta_1 x_i)^2$$

Can be solved explicitly by setting derivative to 0:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \qquad \hat{\beta}_o = \bar{y} - \hat{\beta}_1 \bar{x}$$

$\hat{\beta}_o$ and $\hat{\beta}_1$ found by least squares are unbiased, and we can calculate their estimated variances:

$$s^2_{\hat{\beta}_1} = \frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \qquad s^2_{\hat{\beta}_o} = \frac{s^2}{n} \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

where $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\beta}_o - \hat{\beta}_1 x_i)^2$ (estimator for σ^2).
Can show for $i = 0, 1$, $\frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} \to N(0,1)$, or $\sim t_{n-2}$ if $e_i \sim N(0, \sigma^2)$. Lets us form CI's and tests.
Linear model is very generalizable. E.g., to find c_o, c_1, c_2 in

$$\varphi = c_o + c_1 x + c_2 x^2$$

let $\vec{x}_i = \begin{bmatrix} x_i \\ x_i^2 \end{bmatrix}$ and $\hat{\beta} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$.

Lecture 21

Model accuracy Plot residuals $e_i = y_i - \hat{\beta}_o - \hat{\beta}_1 x_i$ to check whether your model is good. Any dependence of e_i on x (e.g., heteroskedasticity = variance depends on x) suggests the model is insufficient.

Correlation and regression slope If you normalize x_i's and y_i's (subtract out mean, divide by std dev), the correlation is the slope of the normalized x_i's and y_i's. More mathematically, approximate the variances and covariance by

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

Then the best fit line is

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \Longleftrightarrow \qquad \frac{\hat{y}_i - \bar{y}}{S_{yy}} = r \cdot \frac{x_i - \bar{x}}{S_{xx}}$$

where r is the correlation

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \in [-1, 1].$$

Linear algebra view Extension to x_i's that are vectors.

$$\begin{array}{ccccc} y & = & X & \beta & + \quad \epsilon \end{array}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots \\ 1 & x_{21} & x_{22} & \dots \\ \vdots & \vdots & \vdots & \dots \\ 1 & x_{n1} & x_{n2} & \dots \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

Each row of X is a data point, and each column is a feature.
Least squares:

$$\min_{\beta \in \mathbb{R}^p} ||y - X\beta||_2^2.$$

Taking derivative in every β_i and setting to 0,

$$-2X^T(y - X\beta) = 0 \quad \Longrightarrow \quad \underbrace{X^T X\hat{\beta} = X^T y}_{\text{“normal equations”}} \quad \Longrightarrow \quad \hat{\beta} = (X^T X)^{-1}(X^T y).$$

Assumes $X^T X$ is invertible, which assumes $n \geq p$ (overdetermined, more data points than features).

Random vectors Covariance of a random vector z: $\Sigma = \Sigma_{zz} = \mathbb{E}[(z - \mathbb{E}[z])(z - \mathbb{E}[z])^T]$.
$\Sigma_{ij} = \text{Cov}(z_i, z_j)$, $\Sigma_{ii} = \text{Var}(z_i)$, Σ is symmetric.
Linearity of expectation: $\mathbb{E}[Az] = A\mathbb{E}[z]$.
Covariance of linear transformation: $\Sigma_{Az,Az} = A\Sigma_{zz}A^T$.
Expectation of quadratic form: $\mathbb{E}(z^T Az) = \text{Tr}(A\Sigma) + \mu^T A\mu \quad (\mu = \mathbb{E}z)$.

Worksheet Problems

1. Which of the following are random variables, under the standard model of linear least squares regression?

   (✓ = random, × = not random)

   (a) x_i ... ×

   (b) y_i ... ✓

(c) e_i ... ✓
(d) σ^2 ... ×
(e) (β_o, β_1) ... ×
(f) $(\hat{\beta}_o, \hat{\beta}_1)$ ... ✓
(g) $\mathrm{Cov}(\hat{\beta}_o, \hat{\beta}_1)$ ... ×
(h) $(s^2_{\hat{\beta}_o}, s^2_{\hat{\beta}_1})$ ... ✓

2. The Arrhenius equation in chemistry, given by

$$k = Ae^{\frac{-E_a}{k_B T}},$$

relates the rate of a chemical reaction k to the temperature T. For a particular reaction, we are given pairs (T_i, k_i) of rates and temperatures, and we want to find the coefficients A and E_a (k_B is a known constant). Set this up as a linear least squares problem, identifying y_i, x_i, β_0, and β_1.

Solution:

$$k = Ae^{-\frac{E_a}{k_B T}} \Rightarrow \log k = \log A - \frac{E_a}{k_B T}$$
$$\Rightarrow \underbrace{\log k_i}_{y_i} = \underbrace{\log A}_{\beta_o} + \underbrace{\left(-\frac{E_a}{k_B}\right)}_{\beta_1} \underbrace{\frac{1}{T_i}}_{x_i}$$

3. [14.5] Three objects are located on a line at points $0 < p_1 < p_2 < p_3$. These locations are not precisely known. A surveyor makes the following measurements:

(a) He stands at the origin and measures the three distances from there to p_1, p_2 and p_3. Let these measurements be denoted by Y_1, Y_2, Y_3.
(b) He goes to p_1 and measures the distances from there to p_2 and p_3. Let these measurements be denoted by Y_4 and Y_5.
(c) He goes to p_2 and measures the distance from there to p_3. Denote this measurement by Y_6.

Using matrix notation, explain how to set up a least squares problem to find p_1, p_2, p_3.

Solution: We have the following equations:

$$Y_1 = p_1 + e_1, \quad Y_2 = p_2 + e_2, \quad Y_3 = p_3 + e_3,$$
$$Y_4 = -p_1 + p_2 + e_4, \quad Y_5 = -p_1 + p_3 + e_5, \quad Y_6 = -p_2 + p_3 + e_6$$

(Distance is a positive number, so I would interpret this as e.g., $Y_4 = -p_1 + p_2$, not $p_1 - p_2$ though the solutions at the back of the book disagree with my interpretation.)

In matrix form:

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{pmatrix}}_{e}$$

Then solve for $\min_{\beta\in\mathbb{R}^p} ||Y - X\beta||_2^2$.

Do we need a column of 1's on the X matrix? No! There is no β corresponding to an intercept term that we are trying to solve for.

4. [14.28b] Let $X_1, \ldots, X_n$ be independent random variables with means $\mu_1, \ldots, \mu_n$ and variances σ^2 for all $i = 1, \ldots, n$. Find the expectation of

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j$$

using one of the rules we learned for random vectors. Verify your answer in a different (simpler?) way.

Solution:

Define $X = [X_1, \ldots, X_n]^T$ with mean μ and covariance $\Sigma = \sigma^2 I$. The sum is the quadratic form $X^T A X$ where A is a matrix of all 1's, which we can write as $A = \mathbf{1}\mathbf{1}^T$, where $\mathbf{1} = [1, \ldots, 1]^T$. Using the identity $\mathbb{E}[X^T A X] = \text{tr}(A\Sigma) + \mu^T A \mu$:

$$\text{tr}(\sigma^2 \mathbf{1}\mathbf{1}^T) = \sigma^2 \text{tr}(\mathbf{1}\mathbf{1}^T) = n\sigma^2$$

$$\mu^T \mathbf{1}\mathbf{1}^T \mu = \left(\sum \mu_i\right)^2$$

Summing these gives $n\sigma^2 + (\sum \mu_i)^2$.

Alternative solution, not using the rules from lecture: Observe that $\sum_i \sum_j X_i X_j = (\sum X_i)^2$. Let $S = \sum X_i$. Then $\mathbb{E}[S^2] = \text{Var}(S) + (\mathbb{E}[S])^2$. Since the variables are independent, $\text{Var}(S) = n\sigma^2$ and $\mathbb{E}[S] = \sum \mu_i$. Substituting yields:

$$\mathbb{E}[Q] = n\sigma^2 + \left(\sum \mu_i\right)^2$$