Section 3E. Confidence Intervals

Statistics for Data Science

Victor M. Preciado, PhD MIT EECS
Dept of Electrical & Systems Engineering
University of Pennsylvania
preciado@seas.upenn.edu

Recap

- The PDF of $\widehat{\beta}_1$ is (approximately) a normal distribution $\mathcal{N}(\beta_1, \mathrm{SD}(\widehat{\beta}_1)^2)$

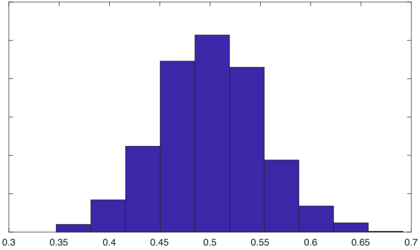

Figure: Histogram of the values of $\widehat{\beta}_1^k$

Confidence Intervals

- Even though the exact value of β_1 that Nature uses to generate the data can never be learnt *exactly* from data, we can build a confidence interval for its value:
  - 95% of the probability mass of a normal distribution $\mathcal{N}\left(\mu, \sigma^2\right)$ is in the interval $[\mu - 2\sigma, \mu + 2\sigma]$
  - Hence,
  $$\Pr\left\{\widehat{\beta}_1 \in \left[\beta_1 - 2\,\mathsf{SD}\left(\widehat{\beta}_1\right), \beta_1 + 2\,\mathsf{SD}\left(\widehat{\beta}_1\right)\right]\right\} = 95\%$$
  - Notice that
  $$\beta_1 - 2\,\mathsf{SD}\left(\widehat{\beta}_1\right) \leq \widehat{\beta}_1 \leq \beta_1 + 2\,\mathsf{SD}\left(\widehat{\beta}_1\right)$$
  is equivalent to (try to prove this yourself)
  $$\widehat{\beta}_1 - 2\,\mathsf{SD}\left(\widehat{\beta}_1\right) \leq \beta_1 \leq \widehat{\beta}_1 + 2\,\mathsf{SD}\left(\widehat{\beta}_1\right)$$

Confidence Intervals (cont.)

- **Confidence interval**:
  - Based on the above, we conclude that the real (but unknown) value of the parameter β_1 satisfies
$$\Pr\left\{\beta_1 \in \left[\widehat{\beta}_1 - 2\,\mathrm{SD}\left(\widehat{\beta}_1\right), \widehat{\beta}_1 + 2\,\mathrm{SD}\left(\widehat{\beta}_1\right)\right]\right\} = 95\%$$
  - Assuming we have access to a single training dataset $\mathcal{D}_{\mathrm{Tr}}\left\{(\mathbf{x}_1, y_N), \ldots, (\mathbf{x}_1, y_N)\right\}$, we can compute
$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \text{ and } \mathrm{SD}\left(\widehat{\beta}_1\right) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2}}$$
  - Therefore, we can explicitly compute the 95% confidence interval (CI) for β_1

Confidence Intervals (cont.)

- Identical logic can be used to compute the 95% CI for β_0
- ***Warnings***: Remember the assumptions we have made above:
  - The estimation $\widehat{\beta}_1$ is a *normally distributed* random variable. In certain situations, this is not the case
  - We have access to σ^2, the variance of the measurement noise. In practice, this may be unknown