Regression:
Least Squares
+
Statistical Inference



Estimation
from a Normal Population
$\mathcal{N}(\mu, \sigma^2)$



$\sigma^2$ known

Consider

$$X_1, X_2, \ldots, X_n \quad \text{i.i.d.} \sim \mathcal{N}(\overbrace{\mu}^{\text{unknown: wish to estimate}}, \overbrace{\sigma^2}^{\text{known}}) .$$

Define sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \bar{X}_n \text{ is estimator for } \mu \ (\mathbb{E}(\bar{X}_n) = \mu)$$

and recall

$$\frac{(\bar{X}_n - \mu)}{\underbrace{\sigma/\sqrt{n}}_{\sigma_{\bar{X}_n}}} \sim \mathcal{N}(0,1) . \qquad \bar{X}_n \text{ is a good estimator for } \mu$$

intuition: As n increases, $P(|\bar{X}_n - \mu| > 2\sigma)$ (say) is increasingly unlikely:

(i) $P(|X_i - \mu| > 2\sigma) = 0.046$; AND
(ii) need many $X_i > 2\sigma$.



Hence

$$P\left(-z \le \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le z\right) = \Phi(z) - \Phi(-z)$$
$$= \Phi(z) - (1 - \Phi(z))$$
$$= 2\Phi(z) - 1$$

Set

$$2\Phi(z_\gamma) - 1 = \gamma \quad \text{(confidence level)}$$
$$\Phi(z_\gamma) = (1+\gamma)/2$$

or

$$z_\gamma = \tilde{z}_{\left(\frac{1+\gamma}{2}\right)} \quad (\text{e.g., } \gamma = 0.95,\ z_\gamma = 1.96) .$$

Thus,

$$P\Big(\underset{-1.96}{-z_\gamma} \le \frac{\bar X_n - \mu}{\sigma/\sqrt{n}} \le \underset{1.96}{z_\gamma}\Big) = \underset{0.95}{\gamma}$$

But

$$-z_\gamma \le \frac{\bar X_n - \mu}{\sigma/\sqrt{n}} \Rightarrow \mu \le \bar X_n + z_\gamma \, \sigma/\sqrt{n}$$

$$\frac{\bar X_n - \mu}{\sigma/\sqrt{n}} \le z_\gamma \Rightarrow \mu \ge \bar X_n - z_\gamma \, \sigma/\sqrt{n}$$

or

$$P(\mu \text{ is in } CI_{\sigma^2 \text{known}}) = \gamma$$

$$CI_{\sigma^2 \text{known}} = [\bar X_n - z_\gamma \, \sigma/\sqrt{n}, \ \bar X_n + z_\gamma \, \sigma/\sqrt{n}] .$$



## $\sigma^2$ unknown

Consider

$$X_1, X_2, \ldots, X_n \text{ i.i.d.} \sim \mathcal{N}(\overset{\text{unknown}}{\mu}, \overset{\text{unknown}}{\sigma^2}) .$$

Now, let

$$\bar X_n = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ be estimator for } \mu \text{, and}$$

$$\hat\sigma_n^2 = \underbrace{\frac{1}{n-1}} \sum_{i=1}^{n} (X_i - \underset{\approx \mu}{\bar X_n})^2 \text{ be estimator for } \sigma^2 .\dagger$$

dof after approximate $\mu$ by $\bar X_n$ (e.g., $n=1$)

† Note $\mathbb{E}(\bar X_n) = \mu$ and $\mathbb{E}(\hat\sigma_n^2) = \sigma^2$.



Then, as $n \to \infty$,

$$P(\mu \text{ is in } CI^0) = \gamma$$

$$CI^0 = [\bar X_n - z_\gamma \, \hat\sigma_n/\sqrt{n}, \ \bar X_n + z_\gamma \, \hat\sigma_n/\sqrt{n}] .$$

More precisely, for any $n$,

$$P(\mu \text{ is in } CI) = \gamma$$

$$CI = [\bar X_n - \rho_{\gamma,1,n} \, \hat\sigma_n/\sqrt{n}, \ \bar X_n + \rho_{\gamma,1,n} \, \hat\sigma_n/\sqrt{n}]$$

where

$$\rho_{\gamma,k,q} \equiv (k \cdot \text{finv}(\gamma, k, q-k))^{1/2} \quad \dagger$$

(related to quantile of F distribution/Student's t) .

† Note $(\text{finv}(0.95, 1, n-1))^{1/2} \to 1.96..$ as $n \to \infty$.



## Regression: Simple$_0$ Model

We will connect Regression for Simple$_0$ Model
to
Estimation from a Normal Population

But note:

| Estimation from Normal | Regression for Simple$_0$ |
|---|---|
| $n$: sample size | $n$: 1 ($\beta \equiv (\beta_0 \ldots \beta_{n-1})^T$) |
| | $m$: sample size $Y_1, \ldots, Y_m$ |
| | (since $X$ is $m \times n$) |

## review of Simple$_0$

$p = 1$, $x_{(1)} = x$ (say)

$n = 1$: $h_0(x) = 1$

$\beta_0$: unknown coefficient

$$Y_{model}(x;\beta) = \sum_{j=1}^{n-1} \beta_j h_j(x) = \beta_0 \quad \text{(fit data to constant)}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} \quad X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad Y_{model}(x_i;\beta) = (X\beta)_i = \beta_0$$

$$\bar{Y} = \bar{Y}_m = \frac{1}{m}\sum_{i=1}^{m} Y_i \quad \text{(sample mean)}$$

$$Y^TY = \|Y\|^2, \quad X^TY = m\bar{Y}, \quad X^TX = m, \quad \hat{\beta}_0 = \bar{Y}_m$$



## application: IR Rangefinder

$$D(\text{istance}) = \frac{\text{Constant}^{?}}{V(\text{oltage})} \Rightarrow \underbrace{\underbrace{DV}_{y} = \text{Constant } \text{``}\underbrace{(V)}_{x}\text{''}}_{Y_{model}(x;\beta) = \beta_0^{\text{true}}}$$

## application: friction coefficient

$$F_{f,\text{static}}^{max} = \mu_s^{?} F_{\text{normal,applied}} \Rightarrow \underbrace{\frac{F_{f,\text{static}}^{max}}{F_{\text{normal,applied}}} = \mu_s \quad \text{Constant ``}(F_{\text{normal,applied}}, A_{\text{surface}}, \ldots)\text{''}}_{Y_{model}(x;\beta) = \beta_0^{\text{true}}}$$

## application: spring constant

$$F_{\text{spring}} = k^{?}\delta \Rightarrow \underbrace{\frac{F_{\text{spring}}}{\delta} = k \quad \text{Constant ``}(\delta, \ldots)\text{''}}_{Y_{model}(x;\beta) = \beta_0^{\text{true}}}$$



## hypotheses on noise

Assume that

$$\underset{\substack{\text{measurement}\\ \text{at } x_i}}{Y_i} = \underbrace{Y_{model}(x;\beta^{\text{true}})}_{\beta_0} + \overset{\text{noise}}{\epsilon_i}, \quad 1 \le i \le m,$$

where

$\epsilon_1, \epsilon_2, \ldots, \epsilon_m$ are i.i.d. $\sim \mathcal{N}(0,\sigma^2)$.

no bias

N1: $\epsilon_i$ normal with zero mean
$\Rightarrow \mathbb{E}(Y_i) = Y_{model}(x_i;\beta^{\text{true}})$;

N2: $\epsilon_i$ homoscedastic — $\sigma_i^2 = \sigma^2$

N3: $\epsilon_i, \epsilon_j$ independent



But then

shift

$$Y_i = \beta_0^{\text{true}} + \overset{\sim \mathcal{N}(0,\sigma^2)}{\epsilon_i} \sim \mathcal{N}(\underbrace{\beta_0^{\text{true}}}_{\mu}, \sigma^2)$$

and hence

$$Y_1, Y_2, \ldots, Y_m \text{ are i.i.d.} \sim \mathcal{N}(\beta_0^{\text{true}}, \sigma^2).$$

Thus

$$\overbrace{\bar{Y}_m}^{\hat{\beta}_0 \text{ (which minimizes } \|r(\beta)\|^2)} = \frac{1}{m}\sum_{i=1}^{m} Y_i \text{ is estimator for } \mu \equiv \beta_0^{\text{true}}; \text{ and}$$

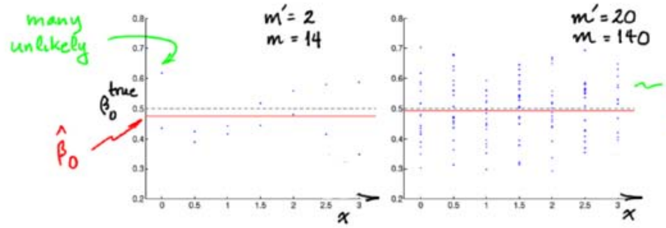
$$\hat{\sigma}_m^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2 \text{ is estimator for } \sigma^2; \text{ and}$$

$$CI = \left[\bar{Y}_m - \beta_{\gamma,1,m}\hat{\sigma}_m/\sqrt{m},\ \bar{Y}_m + \beta_{\gamma,1,m}\hat{\sigma}_m/\sqrt{m}\right]$$

is confidence interval for $\beta_0^{\text{true}}$

$\frac{1}{2} + .1 * \text{randn}$

numerical experiment : (synthetic noise) $\quad Y_i \sim \mathcal{N}(\beta_0^{\text{true}} = \tfrac{1}{2}, \sigma^2 = .01)$ "unknown"



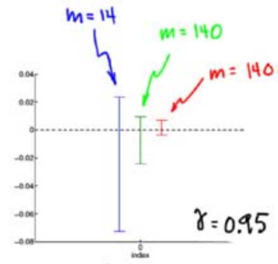




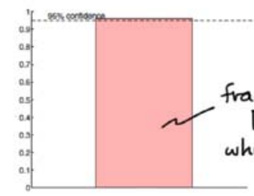
$\bar{x} = 0, .5, 1, 1.5, 2., 2.5, 3.0$ (7 sites)
$m'$ = measurements/site
$\Rightarrow x_i,\ 1 \le i \le m = 7 \cdot m'$



[ci] shifted by $-\beta_0^{\text{true}}$




(three) particular realizations




100 realizations of experiment, each with $m = 140$ measurements



in least-squares lexicon : $\quad m, n = 1$

$\hat\beta_0$ satisfies $X^TX\hat\beta = X^TY \qquad (\hat\beta_0 = \bar{Y}_m)$

$\hat{Y} = X\hat\beta$ model predictions for $\beta = \hat\beta \qquad (\hat Y = \hat\beta_0)$

$r(\hat\beta) = Y - \hat{Y}$

$\|r(\hat\beta)\|^2 = \|Y - \hat Y\|^2 \qquad (< \|Y - X\beta\|^2$ for any $\beta \neq \hat\beta)$

$\hat\sigma_m^2 = \frac{1}{m-n}\|Y - \hat Y\|^2 \qquad X^TX = m$ "design"

$\gamma$ CI (for $\beta_0^{\text{true}}$) $= \left[\hat\beta_0 - \rho_{\gamma,n,m}\,\hat\sigma_m\sqrt{(X^TX)^{-1}},\ \hat\beta_0 + \rho_{\gamma,n,m}\,\hat\sigma_m\sqrt{(X^TX)^{-1}}\right]$

$\rho_{\gamma,k,q} = (k\ \text{finv}(\gamma, k, q-k))^{1/2}$



Regression
General Case

a tiny change to notation

## Regression General Case

a tiny change to notation

but a

# HUGE LEAP of FAITH



## summary of result

Given $m, n, Y_{model}(x;\beta), Y, X$,

$\hat{\beta}$ satisfies $(X^TX)\hat{\beta} = X^TY$ (least squares)

$\hat{Y} = X\hat{\beta}$

$\hat{\sigma}_m^2 = \dfrac{1}{m-n} \|Y - \hat{Y}\|^2$

and (joint confidence intervals for $\beta_j^{true}$, $0 \le j \le n-1$)

$$I_j^{joint} = \left[\hat{\beta}_j - \rho_{\gamma,n,m}\,\hat{\sigma}_m \sqrt{(X^TX)^{-1}_{j+1,j+1}},\ \hat{\beta}_j + \rho_{\gamma,n,m}\,\hat{\sigma}_m \sqrt{(X^TX)^{-1}_{j+1,j+1}}\right]$$

$0 \le j \le n-1$.

---

Note $\rho_{\gamma,n,m} = s_{\gamma,n,m-n}$ of textbook.



Then, joint confidence intervals

$$P\left(\begin{array}{l}\beta_0^{true} \text{ is in } I_0^{joint}, \text{ AND} \\ \beta_1^{true} \text{ is in } I_1^{joint}, \text{ AND} \\ \vdots \\ \beta_{n-1}^{true} \text{ is in } I_{n-1}^{joint}\end{array}\right) = \gamma \qquad (\ge \gamma)$$

Hence can draw conclusions or test hypotheses which involve (simultaneously) all coefficients $\beta_0^{true}, \dots, \beta_{n-1}^{true}$.

$$F_{f,static}^{max} = \beta_0^{true} + \underbrace{\beta_1^{true}}_{\mu_s} F_{normal,applied} + \beta_2^{true} A_{surface}$$

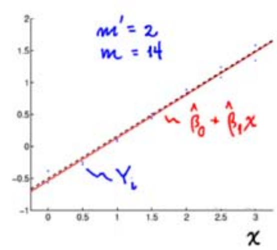
Does $I_0^{joint}$ include 0 ? AND $I_2^{joint}$ include 0 ?

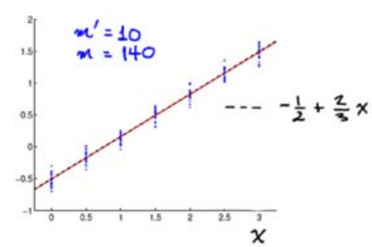


numerical experiment: Simple$_1$ (synthetic data)

$$Y_i = \underbrace{-\tfrac{1}{2}}_{\beta_0^{true}} + \underbrace{\tfrac{2}{3}}_{\beta_1^{true}} x + \mathcal{N}(0,\sigma^2) \qquad \sigma^2 = .01$$

$(Y_{model}(x;\beta) = \beta_0 + \beta_1 x)$




$\bar{x}$ = 0, .5, 1, 1.5, 2, 2.5, 3.0 (7 sites)
m' = measurements/site

$\Rightarrow x_i,\ 1 \le i \le m = 7 \cdot m'$

2.086 Numerical Computation for Mechanical Engineers
Spring 2013