Lesson Plan 4/10

Tuesday, April 10, 2018 9:07 AM

Admin:
- HW4 due tonight (unless...) 1111 Due 4/12
HDSC:
- David Donoho (Stanford Prof. of Statistics): 50 years of Data Science: https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1384734
HW4 questions
Linear Regression
Logistic Regression

(1) $\bar{X}_A$ estimate of p_A

$\bar{X}_B$ estimate of p_B

$Y = \bar{X}_B - \bar{X}_A$ estimate $p_B - p_A$

4a) $VAR(Y) = VAR(\bar{X}_B - \bar{X}_A)$

$= VAR(\bar{X}_B) + VAR(-\bar{X}_A)$

$$= \frac{1}{n_B} p_B(1-p_B) + \frac{1}{n_A} p_A(1-p_A)$$

$$= \left(\frac{1}{n_B} + \frac{1}{n_A}\right) p(1-p)$$

4b) $$p = \frac{s_A + s_B}{n_A + n_B}$$

1a) $\mathrm{var}(X) = E\,(X - EX)^2$

$EX = \sum_x x\,p(X=x)$

Bernoulli: $(p) = p(1-p)$

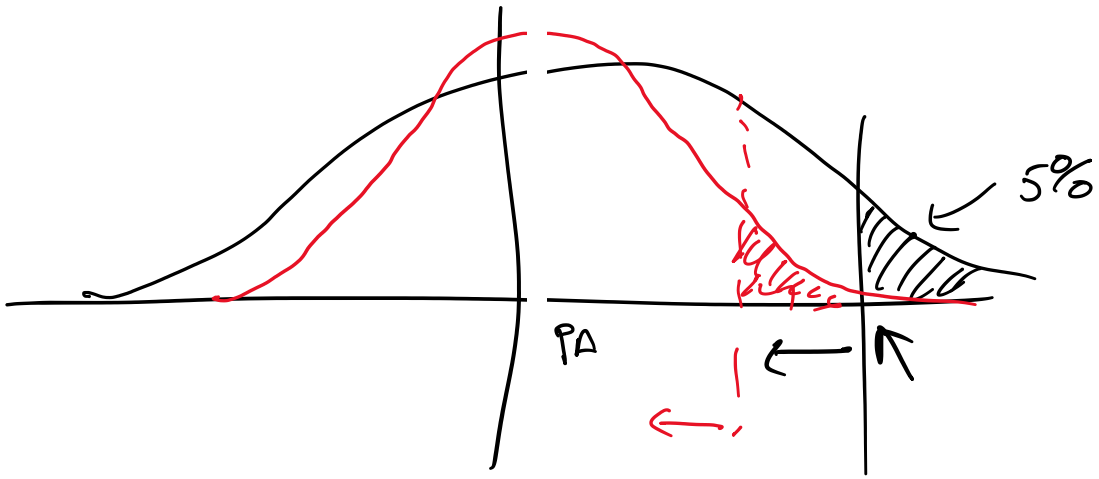
$\frac{1}{n}\sum_i (x_i - \bar{x})^2$

$$\mathrm{var}(\bar{X}) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n} \sum \mathrm{var}(X_i)$$

$$= \frac{1}{n^2} \cdot n\,p(1-p)$$

$$= \frac{1}{n}\,p(1-p)$$



Bernoulli: (p)

$X \in \{0, 1\}$

$p = p(X = 1)$

Linear Regression

Y: outcome

X: predictor

$Y = \beta_0 + \beta_1 X$

$$E[Y \mid X = x] = \beta_0 + \beta_1 x$$

Linear regression problem

Given data $\{\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle\}$

Estimate β_0, β_1 in linear model

$$y = \beta_0 + \beta_1 x$$

least squares problem

$$\arg\min_{\beta_0, \beta_1} \frac{1}{n} \sum_i (y_i - (\beta_0 + \beta_1 x_i))^2$$

Multiple predictors:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

salary ~ major

major $\in$ { CS, Bus, MATH }

$$\text{salary} \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$x = \begin{cases} 1 & \text{if major} = \text{Bus} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if major} = \text{MATH} \\ 0 & \text{o.w.} \end{cases}$$