STAB57

An Introduction to Statistics

Agenda:

1. 知识点 1:bayesian inference
2. 知识点 2:inference
3. 知识点 3: prior 分类
4. 知识点 4: SLR【学校 week11 内容】

提醒：

1. Assignment2 在学生系统 class10，别忘记写～
2. 接下来的 quiz 都是硬算

- 3 月 25 日-3 月 29 日的 quiz：本节课知识点 1-3.题目集中在知识点 2 后
- 4 月 1 日-4 月 5 日的 quiz：本节课知识点 4

知识点 1:BAYESIAN INFERENCE

1. Idea:
   - 回顾
     1) marginal distribution:

        Continous RV - PDF: $f_X(x) = \int f_{X,Y}(x,y)dy$

        Discrete RV- PMF: $f_X(x) = \sum_y f_{X,Y}(x,y)$
     2) conditional distribution
        $$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$
     3) joint PMF:
        $$f_{X,Y}(x,y) = f_{Y|X}(y \mid x) f_X(x) = f_{X|Y}(x \mid y) f_Y(y)$$
   - Frequentist approach:

     Likelihood function 来estimate

2. Bayesian inference：
   - θ是random variable，并且相信有distribution

     这个distribution叫 **prior distribution** of θ，记为$\pi(\theta)$

     是θ的PDF
   - 用$\pi(\theta|s)$来研究data s

     $\pi(\theta|s)$为 **posterior distribution** of θ

     summarizes what you know after the data has been observed.
3. Marginal distribution of s
   $$m(s) = \int \pi(\theta)\, f(s|\theta)d\theta = \int \pi(\theta)\, L(s|\theta)d\theta$$
4. *Posterior density of θ:*
   $$\pi(\theta|s) = \frac{\pi(\theta) f(s|\theta)}{m(s)} = \frac{\pi(\theta) L(s|\theta)}{m(s)}$$

知识点 2:INFERENCE - ESTIMATION USING POSTERIOR DISTRBUTION

1. Posterior distribution可以计算对应的posterior mean， posterior variance，…

2. **Posterior median**可以通过posterior distribution的median计算

3. **Posterior mode**可以通过计算什么θ 可以使posterior density最大

4. Bayesian 的 CI 称为**credible interval**

5. **HPD intervals**:

   Credible interval中的一个，满足$C(s) = \{\varphi: \omega(\varphi|s) \geq c\}$

   $\omega(\varphi|s)$是marginal posterior density of φ, c要满足credible interval

EXAMPLE 1

Suppose that in a population of stendets in course with large enrollment the mark, out of 100, on final exam is approximately distributed as $N(\mu, 9)$.

The instructor places a prior $\mu \sim N(65,1)$ on unknown parameter. A sample of 10 marks is obtained as given below.

46, 68, 34, 86, 75, 56, 77, 73, 53, 64

a. Determine the 95% credible interval for μ

$(x_1, x_2, \ldots, x_n) \sim N(\mu, \sigma_0^2)$ where σ^2 is known and prior: $\mu \sim N(\mu_0, \tau_0^2)$

Posterior distribution is $N(\mu^*, \sigma^{*2})$ where,

$\mu^* = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma_0^2}\bar{x}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma_0^2}}$ and $\sigma^{*2} = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma_0^2}}$

复杂的这种一般他会直接给你

b. Use the 95% credible interval for μ to test the hypothesis $H_0: \mu = 65$

c. What is the mode of the posterior distribution of μ

EXAMPLE 2: E&R 7.1.2

determine the posterior mean and variance of θ for Bernoulli model

EXAMPLE 3: E&R 7.1.3

In Example 7.1.2, what is the posterior probability that μ is positive, given that $n = 10, \bar{x} = 1$ when $\sigma_0^2 = 1, \mu_0 = 0$, and $\tau_0^2 = 10$?

知识点 3:PRIOR 分类

1. **Conjugate prior**：

   result in a posterior distribution that belongs to the same family of distribution as the prior

2. **Improper priors**：

   $\pi(\theta)$不是 proper PDF

3. **Non-informative priors**:

   If we have no prior information, we want a prior with minimal influence on the inference.We then use priors that are non-informative or vague

   没有 information

知识点 4:SLR

1. Pearson Correlation Coefficient:

- 测量 strength + direction，线性关系的强弱
- Population correlation：

$$\text{Standardize } X \text{ and } Y: Z_x = \frac{x - \mu_x}{\sqrt{\mathrm{Var}\,(X)}}, Z_y = \frac{Y - \mu_y}{\sqrt{\mathrm{Var}\,(Y)}}$$

$$\mathrm{Corr}\,(X, Y) = \mathrm{Cov}\,(Z_x, Z_y) = \rho_{xy} = \frac{\mathrm{Cov}\,(X, Y)}{\sqrt{\mathrm{Var}\,(X)}\sqrt{\mathrm{Var}\,(\mathrm{Y})}}$$

- Sample correlation：

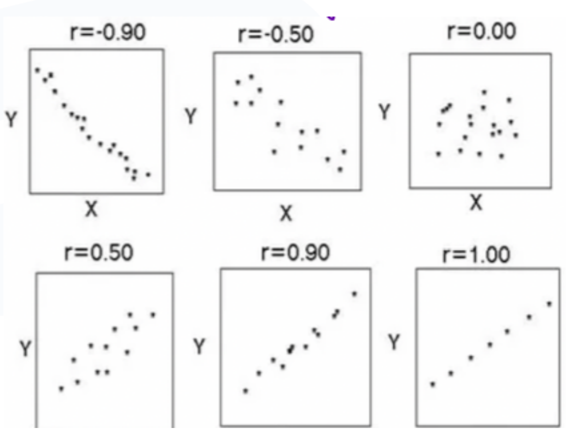
$$\text{Standardize } X \text{ and } Y: Z_x = \frac{X - X}{s_x}, Z_y = \frac{Y - Y}{s_y}$$

$$\mathrm{Corr}\,(X, Y) = \mathrm{Cov}\,(Z_x, Z_y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} = r = r_{xy}$$

- Sample correlation 是 population correlation 的 estimate

2. Correlation 的性质：

- $-1 \leq \mathrm{Corr} \leq 1$
- Strength 看绝对值大小
  - ｜Corr｜=1: perfect linear
  - ｜Corr｜ → 1 : strong association
  - ｜Corr｜=0: independent， uncorrelated
  - ｜Corr｜ → 0 : weak association
- Direction 看正负
  - Corr > 0 : positive association
  - Corr < 0 : negative association



3. 做 regression 的第一步其实是画 scatterplot 来确定 quantitative variables 之间的 relationship

Interpretation of scatterplot:

- Direction of the line:
  - the line is going upward ⇒ the correlation is positive.
  - the line is going downward ⇒ then the correlation is negative.
- Closeness of the points to the line suggests the strength of the correlation
  - points are closely clustered around the line ⇒ strong correlation
  - points are not so close to the line ⇒ moderate/weak correlation
- If the points look totally random ⇒ No relationship between X and Y

AUSUMPTIONS ABOUT SLR

➢ **Simple Linear Regression:**

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i$$

y_i: Response or dependent variable

x_i: Predictor or independent variable, treated as fixed

β_1 and β_2: parameters, regression coefficients

ϵ_i: Random Error

Goal: to be able to predict y for a given value of x

➢ **Assumptions about ϵ_i:**

- For purpose of deriving the statistical inferences only, utilized while constructing tests of hypothesis and confidence interval for parameters
- Identically: Have equal variance $Var(\epsilon_i) = \sigma^2$
- Independently: independent of each other
- Distributed: normally distributed: $\epsilon_i \sim N(0, \sigma^2)$

➢ **About y_i:**

- $E(y_i) = \beta_1 + \beta_2 x_i$
- $Var(y_i) = Var(\epsilon_i) = \sigma^2$
- Normal distribution: $y_i \sim N(\beta_1 + \beta_2 x_i, \sigma^2)$

THE METHOD OF LEAST SQAURES

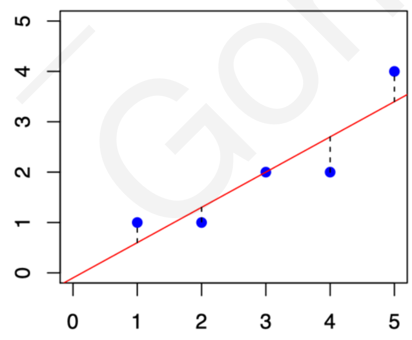
➢ **Interpretation of Regression Line:**

$$E(y_i) = \beta_1 + \beta_2 x_i$$

- β_1 is the y-intercept of the line: the point at which the line intersects the y-axis.
- $\beta_2$is the slope of the line: the change (amount of decrease or increase) in mean response y in its unit for every one unit increase in x
- Goal: fit the data points by finding the line that is closest to the data.

➢ **Method of least square:**

- The values of β_1 and β_2 are unknown
- The estimate of β_1 and β_2 are B_1 and B_2
- Goal: estimate β_1 and β_2 by minimizing the sum of squares of the difference between the observations and the line in the scatterplot.
  the **vertical difference** between the observations and the line in the scatterplot
  即： minimize SSE= $Q = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2$



- Estimates are least-square estimates or ordinary least square estimates(OLS)

- **OLS**
  - Fitted line or fitted linear regression model:

$$y = \hat{\beta}_1 + \hat{\beta}_2\, x$$

With intercept

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

With slope

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

  - Predicted value:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

  - Residual: the difference between the observed value and the fitted (or predicted value)

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_i)$$

INFERENCE IN SLR

- $\hat{\beta}_2$:
  - $E(\hat{\beta}_2) = \beta_2$
  - $Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
  - $\hat{\beta}_2$ is unbiased estimator of β_2

- $\hat{\beta}_1$:
  - $E(\hat{\beta}_1) = \beta_1$
  - $Var(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}}{S_{XX}}\right)^2$
  - $\hat{\beta}_1$is unbiased estimator of β_1

- $\widehat{\sigma^2}$：我直接写了和 anova 匹配的公式
  - OLS: $\widehat{\sigma^2} = \frac{SSE}{n-2} = MSE$
    Unbiased estimator of σ^2
  - MLE : $\frac{SSE}{n}$
    biased

- CI for β_2:

$$\hat{\beta}_2 \pm t_{(1+\gamma)/2,(df=n-2)} * SE(\hat{\beta}_2)$$

- T test for β_2:
  - Testing $H_0: \beta_2 = 0$ (no relationship between X and Y)
  - $T = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} \sim t_{(n-2)}$

MAXIMUM LIKELIHOOD FUNCTION(MLE)

- **Likelihood function for y_i:**

$$f(y_i|x,\beta_0,\beta_1,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\{-\frac{[y_i-(\beta_0+\beta_1 x)]^2}{2\sigma^2}\}$$

$$L(y_1,\dots,y_n) = \prod_{i=1}^{n} f(y_i|x,\beta_0,\beta_1,\sigma^2) = \left[\frac{1}{2\pi\sigma^2}\right]^{\frac{n}{2}} \exp\{-\frac{\sum_{i=1}^{n}[y_i-(\beta_0+\beta_1 x_i)]^2}{2\sigma^2}\}$$

- **Log- Likelihood function for y_i:**

$$LL(y_1,\dots,y_n) = \ln\left(\left[\frac{1}{2\pi\sigma^2}\right]^{\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^{n}[y_i-(\beta_0+\beta_1 x_i)]^2}{2\sigma^2}\right\}\right)$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i-(\beta_0+\beta_1 x_i)]^2$$

- **Normal equations:**

By find partial derivatives w.r.t. β_0, β_1 and set to 0

$$n\beta_0 + \beta_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\beta_0\sum_{i=1}^{n} x_i + \beta_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

ANOVA

- Sum of squares decomposition:

$$\text{Total sum of square (TSS)} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\text{Regression sum of square (RSS)} = \hat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\text{Error/Residual sum of square (ESS)} = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right)^2$$

$$TSS = RSS + ESS$$

- Anova table:

Source	df	Sum of Square (SS)	Mean SS = SS/df
X	1	$b_2^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$	$b_2^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$
Error	$n-2$	$\sum_{i=1}^{n}(y_i - b_1 - b_2 x_i)^2$	s^2
Total	$n-1$	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	-

- **F – test:**
  - State Hypothesis:
    $H_0: \beta_2 = 0 =$ x contributes no information for predicting y
    $H_a: \beta_2 \neq 0 = x$ is useful for predicting y
  - Find Test Statistics:

$$F^* = \frac{\frac{RSS}{1}}{\frac{ESS}{n-2}} = \frac{MSR}{MSE} \sim F(1, n-2)$$

  - Decision rule:

    1) P-value approach：p-value $= P(T > t_{\text{obs}})$

       Reject H_0 if p-value $< \alpha$

    2) Rejection region approach：Reject H_0 if $t_{\text{obs}} > F(\alpha; 1, n-2)$

- Coefficient of determination:

  the proportion of variation in Y that can be explained by the model.

$$R^2 = \frac{RSS}{TSS}$$

For simple linear regression (only one X variable),

$$r^2 = R^2 \Longrightarrow r = \sqrt{R^2}$$

EXAMPLE 4: E&R 10.3.7

A student takes weekly quizzes in a course and receives the following grades over 12 weeks. X = week and Y = grade.

Week	Grade	Week	Grade
1	65	7	74
2	55	8	76
3	62	9	48
4	73	10	80
5	68	11	85
6	76	12	90