

STAT 135 Lecture 18

Henry Liev

5 August 2025

Remark 0.1

Fact: Mean of X 's, Mean of Y 's lie on the regression line

$$\mathbf{1}^T = [1 \quad 1 \quad 1 \quad \cdots \quad 1]$$

Last time: Mean residuals is 0 $\mathbf{Y} - \hat{\mathbf{Y}} = 0$

$$\mathbf{1}^T [\mathbf{Y} - \hat{\mathbf{Y}}] = 0$$

$$\mathbf{1}^T \mathbf{Y} = \mathbf{1}^T (\mathbf{X} \hat{\beta})$$

$$\mathbf{1}^T \mathbf{Y} = (\mathbf{1}^T \mathbf{X}) \hat{\beta}$$

$$\mathbf{1}^T \mathbf{Y} = [s_1 \quad s_2 \quad \cdots \quad s_p] \hat{\beta}$$

$$\frac{1}{n} \mathbf{1}^T \mathbf{Y} = \frac{1}{n} [\cdots] \hat{\beta} \quad s_j = j^{th} \text{ column sum}$$

Remark 0.2

Regression Diagnostics

Assumptions:

1. Normally distributed errors
2. Independence
3. Constant variance
4. Linearity

Look at residuals vs predicted values, if there is a cone, then the variance is not constant, or if there is a shape that does not look random the x-axis, then it is likely nonlinear, the residuals should also have mean 0

Remark 0.3

More about regression:

Linear in parameters, not the data (x 's)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 (X_{i2})^2 + \cdots + \beta_{p-1} (X_{i1} X_{i2})$$

Can transform Y 's and combine X 's

Method 0.4 (Can formulate ANOVA as a linear model)

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad \sum \alpha_j = 0$$

3 groups: μ, α_1, α_2

$$\mathbb{Y} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 0 \rightarrow \alpha_3 = -\alpha_1 - \alpha_2$$

Remark 0.5

- ANOVA is a special case of linear models (Regression)
- We can set up \mathbb{X} matrix to encode which observation falls into which group

Remark 0.6 (Opinion)

- I use linear models descriptively to understand simple relationships in data
- Step up from looking at correlations
- Don't take the inference very seriously due to strong assumptions

Definition 0.7 (Bayesian Statistics)

Have been working in the Frequentist

- Parameters are non-random constants
- Expressions of uncertainty are through sampling distributions of estimators

A realized 95% confidence interval is not a statement of uncertainty about a parameter value
Frequentism is awkward

We want to talk about uncertainty in parameters

Why “confidence”?

“The probability of observing a result at least as extreme as what we observed (under the null hypothesis)”

Frequentists approach focuses on $P(\text{Data}|\text{hypothesis})$

Really care about $P(\text{Hypothesis}|\text{Data})$

Method 0.8 (Bayes' Rule)

$$P(B|A) = P(A|B) \frac{P(B)}{P(A)}$$

$$P(Y = y|X = x) = P(X = x|Y = y) \cdot \frac{P(Y = y)}{P(X = x)}$$

Continuous version:

$$f_{Y|X}(y|x) = f_{X|Y}(x|y) \frac{f_Y(y)}{f_X(x)}$$

Continuous random variable X and Y have some joint distribution

$f_X(x)$ marginal density of X

$f_Y(y)$ marginal density of Y

$f_{Y|X}(y|x)$ conditional density of Y given $X = x$

Remark 0.9

Big idea:

Model parameter is a random variable

Sample data: $X_1, \dots, X_n \sim f(X|\Theta = \theta)$

$f_{\Theta|X}(\theta|x) = f_{X|\theta}(x|\theta) \frac{f_{\Theta}(\theta)}{f_X(x)}$ posterior = likelihood times prior

$f_X(x) = \int f_{X|\theta}(x|\theta) f_{\Theta}(\theta) d\theta$

$$f_{\Theta|X}(\theta|x) \propto f_{X|\theta}(x|\theta) f_{\Theta}(\theta)$$

Find a & b such that $P(a \leq \Theta \leq b | \text{data}) = 95\%$ Bayesian credible interval

Example 0.10 (Coin Flipping)

Flip a possible biased coin 10 times, observe 6 heads and 4 tails

Θ is the underlying probability of heads

Need a prior on Θ pick $U(0, 1)$

$$f_X(x|\theta) = \binom{10}{6} \theta^6 (1 - \theta)^4$$

$$f_{\Theta|X}(\theta|X = 6) \propto \theta^6 (1 - \theta)^4 \cdot 1 = \theta^6 (1 - \theta)^4 \sim \text{Beta}(7, 5)$$

Need to find K such that $\int_0^1 K \theta^6 (1 - \theta)^4 d\theta = 1 \rightarrow K = \frac{11!}{4!6!}$

Now we can construct a Bayesian Credible Interval

Remark 0.11 (Why is this not the default way to do things)

- Seems more natural
- Statistics is for science \Rightarrow want “objective” methods, priors seem subjective

“Real” priors may not have nice closed forms

More realistic setup \Rightarrow Computations get very difficult