Lecture 5: Inference and Prediction

The Gaussian/Normal Linear Model

Ailin Zhang

2023-05-19

Recap: Gauss-Markov Model

Under the Gauss-Markov Model,

- $E(\hat{\beta}) = \beta$
- $\text{cov}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$
- $E(\hat{\sigma}^2) = \sigma^2$, where $\hat{\sigma}^2 = \frac{RSS}{n-(p+1)} = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{n-(p+1)}$

Statistical Inference

- We would like to further study the distribution of the OLS estimator
  - Enable statistical Inference
- To derive the distribution, we need stronger assumptions
  - The assumption will focus on the Gaussian/Normal linear model:
    - $\epsilon \sim \mathrm{N}(0, \sigma^2 I_n)$
    - $Y \sim \mathrm{N}(X\beta, \sigma^2 I_n) \Longleftrightarrow y_i \stackrel{\text{ind}}{\sim} N(x_i{}^T \beta, \sigma^2), \qquad i = 1, \ldots, n$
  - where X is fixed such that $X^T X$ is non-degenerate, and (β, σ^2) are fixed but unknown parameters.
- The modeling assumption is extremely strong, but it is canonical in statistics.
- It allows us to derive elegant formulas, and also justifies the output of the linear regression function in many statistical packages.

Recap: Normal and χ^2 Distribution

- If $Z_1, \ldots, Z_k$ are independent, standard normal random variables, then the sum of their squares,

$$Q = \sum_{i=1}^{k} Z_i^2$$

  is distributed according to the chi-squared distribution with k degrees of freedom
- We denote it as $Q \sim \chi_k^2$

Joint Distribution of $(\hat{\beta}, \hat{\sigma}^2)$

Theorem

Under the Gaussian linear model,

$$\begin{pmatrix} \hat{\beta} \\ \hat{\epsilon} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} \beta \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} (X^T X)^{-1} & 0 \\ 0 & I_n - H \end{pmatrix} \right\}$$

so $\hat{\beta} \perp\!\!\!\perp \hat{\epsilon}$; $\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-(p+1)}$, *and* $\hat{\beta} \perp\!\!\!\perp \hat{\sigma}^2$

Proof.

$$\begin{pmatrix} \hat{\beta} \\ \hat{\epsilon} \end{pmatrix} = \begin{pmatrix} (X^T X)^{-1} X^T Y \\ (I_n - H) Y \end{pmatrix} = \begin{pmatrix} (X^T X)^{-1} X^T \\ I_n - H \end{pmatrix} Y$$

This is a linear transformation of Y , so they are jointly normal.

Joint Distribution of $(\hat{\beta}, \hat{\sigma}^2)$

Proof. (Continued)

We have verified their means and variances, so we only need to show that their covariance is zero

$$\mathrm{cov}(\hat{\beta}, \hat{\epsilon}) = (X^T X)^{-1} X^T \mathrm{cov}(Y)(I_n - H)^T = \sigma^2 (X^T X)^{-1} X^T (I_n - H^T) = 0$$

- $\hat{\sigma}^2 = RSS/(n-p-1) = \hat{\epsilon}^T \hat{\epsilon}/(n-p-1)$ is a quadratic function of $\hat{\epsilon}$. We only need to show that it is a scaled chi-squared distribution, which follows from the Normality of $\hat{\epsilon}/\sigma$ with the projection matrix $I_n - H$ as its covariance matrix. (Not required)

Joint distribution of $(\hat{Y}, \hat{\epsilon})$

Theorem

Under the Gaussian linear model,

$$\begin{pmatrix} \hat{Y} \\ \hat{\epsilon} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} H & 0 \\ 0 & I_n - H \end{pmatrix} \right\}$$

so $\hat{Y} \perp\!\!\!\perp \hat{\epsilon}$

- Orthogonal: a linear algebra fact without assumptions.
- Independent: a statistical property under the Gaussian linear model.

Example: The Auto Data

Auto data of 9 variables about 392 car models in the 1980s. The variables include

- acceleration: Time to accelerate from 0 to 60 mph (in seconds)
- horsepower: Engine horsepower
- weight: Vehicle weight (lbs.)
- Description of all 9 variables: https://rdrr.io/cran/ISLR/man/Auto.html

```r
Auto = read.table("auto.txt", header=T)
```

How to run regression in R?

```
lm(acceleration ~ weight + horsepower, data=Auto)
```

```
##
## Call:
## lm(formula = acceleration ~ weight + horsepower, data = Aut
##
## Coefficients:
## (Intercept)       weight   horsepower
##   18.435791     0.002302    -0.093313
```

The lm() command above asks R to fit the model

acceleration $= \beta_0 + \beta_1$weight $+ \beta_2$horsepower $+ \epsilon$

and R gives us the regression equation

acceleration $= 18.4358 + 0.0023$weight -0.0933horsepower

More R Commands

```r
lm1 = lm(acceleration ~ weight + horsepower, data=Auto)
lm1$coef # show the estimated beta's
```

```
## (Intercept)      weight  horsepower
## 18.43579116  0.00230182 -0.09331277
```

```r
lm1$fit # show the fitted values
```

```
##         1         2         3         4         5         6
## 14.370709 11.539806 12.347930 12.341024 13.310981  9.952064
##         9        10        11        12        13        14
##  7.625972  9.568373 10.774006 11.813017 13.096021  4.543836
##        17        18        19        20        21        22
## 15.769702 16.459014 15.127144 18.367244 16.468044 15.631065
##        25        26        27        28        29        30
## 16.132862  8.996446  9.846002  8.926686 11.318640 15.127144
##        33        34        35        36        37        38
```

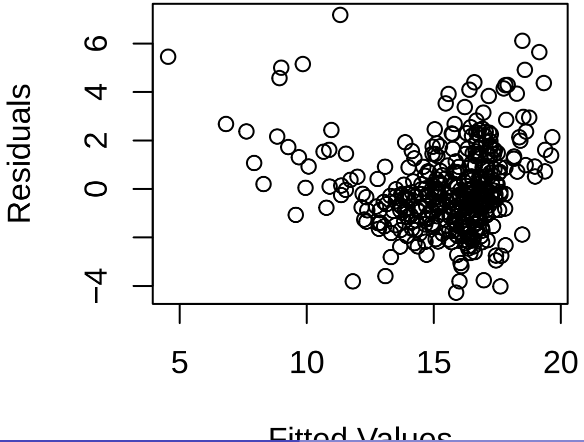
More R Commands

```r
lm1$res # show the residuals
```

```
##           1           2           3           4
## -2.370708958 -0.039806159 -1.347929897 -0.341024437 -2.8109
##           7           8           9          10
##  1.070892905  0.201005520  2.374027511 -1.068372771 -0.7740
##          13          14          15          16
## -3.596021395  5.456164483 -0.030995504 -0.592134521 -0.2697
##          19          20          21          22
## -0.627144421  2.132756343  1.031956377 -1.131064889  2.4620
##          25          26          27          28
## -1.132861648  5.003554062  5.153997564  4.573314295  7.1813
##          31          32          33          34
##  0.251037230 -0.699533425 -2.167508517 -1.053909786 -1.2672
##          37          38          39          40
## -1.172898793 -0.727545276 -0.881381723 -0.220698357 -0.8671
```

More R Commands

```
plot(lm1$fit,lm1$res,
xlab="Fitted Values",
ylab="Residuals")
```



More R Commands

```r
summary(lm1)
```

```
##
## Call:
## lm(formula = acceleration ~ weight + horsepower, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2802 -1.1236 -0.2544  0.9128  7.1814
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.4357912  0.3264888   56.47   <2e-16 ***
## weight       0.0023018  0.0002068   11.13   <2e-16 ***
## horsepower  -0.0933128  0.0045628  -20.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.745 on 389 degrees of freedom
## Multiple R-squared:  0.6018, Adjusted R-squared:  0.5998
## F-statistic:   294 on 2 and 389 DF,  p-value: < 2.2e-16
```

Interpreting the intercept β_0

- β_0 = intercept = the mean value of Y when all X_j's are 0.
- may have no practical meaning

  e.g., β_0 is meaningless in the Auto model as no car has 0 weight

Interpreting the coefficient β_j

- $\beta_j =$ the regression coefficient for X_j, is the mean change in the response Y when X_j is increased by one unit holding other X_i's constant.
- Also called the partial regression coefficients because they are adjusted for the other covariates.
- Interpretation of β_j depends on the presence of other predictors in the model. (Let's explore through an exercise!)