Homework 3, STAT 5205

Zongyi Liu

Thu, Mar 13, 2025

1 Question 1

Let $\boldsymbol{X} \in \mathbb{R}^{n\times p}, \boldsymbol{y} \in \mathbb{R}^n$ be the covariates/regressors and response. Let $\hat{\boldsymbol{y}}$ be the fitted values of least square estimation, that is $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$. Please prove that

$$\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2+\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Answer

1.1 Prove in Simple Algebra

This is actually called the ANOVA Analysis, that is, **Total Sum of Squares**=**Regression Sum of Squares**+**Sum of Squared Errors**. Firstly we need to square both sides and summing over i:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2=\sum_{i=1}^{n}[(y_i-\hat{y}_i)+(\hat{y}_i-\bar{y})]^2$$

Expanding the square, we can get $\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$

Then we need to prove the cross term is 0. For the last term,

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}),$$

is zero because in least squares regression, the residuals $e_i = y_i - \hat{y}_i$ are orthogonal to the fitted values $\hat{y}_i$. That is, $\sum_{i=1}^{n} e_i\hat{y}_i = 0$

Since the cross-term can be canceled, we obtain the desired identity:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2.$$

1.2 Prove in Matrices

The residual vector $\hat{e}$ is $y-\hat{y}=y-X\hat{\beta}=y-X\left(X^TX\right)^{-1}X^Ty$, so the residual sum of squares $\hat{e}^T\hat{e}$ is $y^Ty-y^TX\left(X^TX\right)^{-1}X^Ty$

We can rewrite the Total Sum of Squares as

$$TSS=(y-\bar{y})^T(y-\bar{y})=y^Ty-2y^T\bar{y}+\bar{y}^T\bar{y}$$

The Sum of Squared Errors is $ESS=(\hat{y}-\bar{y})^T(\hat{y}-\bar{y})=\hat{y}^T\hat{y}-2\hat{y}^T\bar{y}+\bar{y}^T\bar{y}$

Also,

$$\hat{y}^T\hat{y}=y^TX\left(X^TX\right)^{-1}X^TX\left(X^TX\right)^{-1}X^Ty=y^TX\left(X^TX\right)^{-1}X^Ty=\hat{y}^Ty$$

Thus we can have:

$$\begin{aligned}TSS&=\|y-\bar{y}\|^2=\|y-\hat{y}+\hat{y}-\bar{y}\|^2\\&=\|y-\hat{y}\|^2+\|\hat{y}-\bar{y}\|^2+2\langle y-\hat{y},\hat{y}-\bar{y}\rangle\\&=RSS+ESS+2y^T\hat{y}-2\hat{y}^T\hat{y}-2y^T\bar{y}+2\hat{y}^T\bar{y}\\&=RSS+ESS-2y^T\bar{y}+2\hat{y}^T\bar{y}\end{aligned}$$

which again gives the result that $TSS = ESS + RSS$:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

since $(y - \hat{y})^T \bar{y} = 0$.

2 Question 2

We consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \text{iid for } i = 1, \dots, n.$$

Given the sample data points generated from the above model,

$$\begin{aligned}
(x_1, y_1) &= (1, 3.9)\\
(x_2, y_2) &= (2, 6.05)\\
(x_3, y_3) &= (3, 8.84)\\
(x_4, y_4) &= (4, 12.36)\\
(x_5, y_5) &= (5, 14.16)\\
(x_6, y_6) &= (6, 17.40)\\
(x_7, y_7) &= (7, 21.12)\\
(x_8, y_8) &= (8, 25.21)\\
(x_9, y_9) &= (9, 27.97)\\
(x_{10}, y_{10}) &= (10, 29.67)
\end{aligned}$$

complete the following two tasks:

2.1 Part a

construct a 95% confidence for β_1. You need to compute $\hat{\beta}_1$, SSX, SSE, and briefly explain why

$$\mathbb{P}\{\beta_1 \in \text{your confidence interval}\} = 95\%.$$

Answer

Compute β_1 and β_0 First we need to compute the Least Squares Estimate of $\beta_1$to get $\hat{\beta}_1$; the simple linear regression model is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, and the estimate of β_1 is given by $\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$.

We have $\bar{x} = \frac{1+2+3+...+10}{10} = 5.5$, and $\bar{y} = \frac{3.9+6.05+8.84+12.36+14.16+17.40+21.12+25.21+27.97+29.67}{10} = 16.668$. Thus $\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - 5.5)(y_i - 16.668)$

For $\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - 5.5)(y_i - 16.668)$, we compute separately:

$$\begin{aligned}
(1 - 5.5)(3.9 - 16.668) &= 57.456\\
(2 - 5.5)(6.05 - 16.668) &= 37.163\\
(3 - 5.5)(8.84 - 16.668) &= 19.57\\
(4 - 5.5)(12.36 - 16.668) &= 6.462\\
(5 - 5.5)(14.16 - 16.668) &= 1.254\\
(6 - 5.5)(17.4 - 16.668) &= 0.366\\
(7 - 5.5)(21.12 - 16.668) &= 6.678\\
(8 - 5.5)(25.21 - 16.668) &= 21.355\\
(9 - 5.5)(27.97 - 16.668) &= 39.357\\
(10 - 5.5)(29.67 - 16.668) &= 58.509
\end{aligned}$$

57.456+37.163+19.57+6.462+1.254+0.366+6.678+21.355+39.557+58.509=248.37

$$SSX = 4.5^2 \times 2 + 3.5^2 \times 2 + 2.5^2 \times 2 + 1.5^2 \times 2 + 0.5^2 \times 2 = 82.5$$

Thus $\hat{\beta}_1 = \frac{248.37}{82.5} = 3.0105$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 16.668 - (3.0105 \times 5.5) = 0.11025$

Compute SSE

Then we need to compute the Sum of Squared Errors (SSE), as the residuals are given by:

$$\epsilon_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

We compute the fitted values $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

$\hat{y} = [0.1100 + 3.0105 \times x_i] = [3.1205, 6.1310, 9.1415, 12.1520, 15.1625, 18.1730, 21.1835, 24.1940, 27.2045, 30.2150]$

Compute residuals:

$$\epsilon_i = y_i - \hat{y}_i = [3.9 - 3.1205, 6.05 - 6.1310, 8.84 - 9.1415, ..., 29.67 - 30.2150]$$

$$= [0.7795, -0.0810, -0.3015, 0.2080, -1.0025, -0.7730, -0.0635, 1.0160, 0.7655, -0.5450]$$

Compute SSE:

$$SSE = \sum \epsilon_i^2 = 4.2702$$

Compute Standard Error of $\hat{\beta}_1$

$$\sigma^2_{\hat{\beta}_1} = \frac{SSE}{n-2} \times \frac{1}{SSX}$$

where $n = 10$, so degrees of freedom $= n - 2 = 8$.

$$\sigma^2_{\hat{\beta}_1} = \frac{4.2702}{8} \times \frac{1}{82.5} = 0.002578$$

$$\text{SE}(\hat{\beta}_1) = \sqrt{0.002578} = 0.0508$$

Get the 95% Confidence Interval

The confidence interval for β_1 is given by:

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} \cdot \text{SE}(\hat{\beta}_1)$$

For $\alpha = 0.05$, the t-critical value for 8 degrees of freedom is:

$$t_{0.025,8} = 2.306$$

Compute the margin of error:

$$ME = 2.306 \times 0.0508 = 0.1855$$

Confidence interval:

$$(3.0105 - 0.1855, 3.0105 + 0.1855) = (2.8251, 3.1960)$$

This means that we are 95% confident that the true slope β_1 lies within the interval $(2.8251, 3.1960)$. That is:

$$\mathbb{P}\{\beta_1 \in (2.8251, 3.1960)\} = 95\%$$

2.2 Part b

Hypothesis testing for

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$$

You need to construct your test statistic (t statistic), compute its value based on the sample points, and then decide to whether accept or reject the null hypothesis.

Answer

We conduct a two-tailed t-test to determine if β_1 is significantly different from zero. Firstly we need to get the t-statistic for testing $H_0 : \beta_1 = 0$, which is given as $t = \frac{\hat{\beta}_1}{\text{SE}(\hat{\beta}_1)}$, and we substitute the values $t = \frac{3.0105}{0.0804} = 37.43$

Then we need to compute the p-value for a two-tailed test:

$$p = 2 \times (1 - \text{CDF of t-distribution at } |t|)$$

Using a t-distribution with $n - 2 = 8$ degrees of freedom:

$$p = 2.85 \times 10^{-10}$$

As we learned previously, the significance level is $\alpha = 0.05$, if $p < \alpha$, we reject H_0; otherwise, we fail to reject H_0. Since $p = 2.85 \times 10^{-10}$ is much smaller than 0.05, we reject the null hypothesis.

2.3 Do it in R

I also double checked it in R

```r
x <- c(1,2,3,4,5,6,7,8,9,10)
y <- c(3.9,6.05,8.84,12.36,14.16,17.40,21.12,25.21,27.97,29.67)

model <- lm(y ~ x)

beta_0_hat <- coef(model)[1]  # Intercept
beta_1_hat <- coef(model)[2]  # Slope

# Compute SSX (Sum of Squares of x)
x_mean <- mean(x)
SSX <- sum((x - x_mean)^2)

# Compute residuals and SSE
residuals <- model$residuals
SSE <- sum(residuals^2)

# Compute standard error of beta_1_hat
n <- length(x)
sigma_squared_hat <- SSE / (n - 2)
se_beta_1_hat <- sqrt(sigma_squared_hat / SSX)

# Compute 95% confidence interval for beta_1
alpha <- 0.05
t_crit <- qt(1 - alpha/2, df=n-2)  # t-critical value

CI_lower <- beta_1_hat - t_crit * se_beta_1_hat
CI_upper <- beta_1_hat + t_crit * se_beta_1_hat

cat("Estimate of beta1:", beta_1_hat, "\n")
cat("SSX:", SSX, "\n")
cat("SSE:", SSE, "\n")
cat("95% Confidence Interval for beta1: (", CI_lower, ",", CI_upper, ")\n")

plot(x, y, main="Simple Linear Regression", xlab="x", ylab="y", pch=19, col="blue")
abline(model, col="red", lwd=2)  # Add regression line
legend("topleft", legend=c("Observed Data", "Regression Line"),
col=c("blue", "red"), pch=c(19, NA), lwd=c(NA, 2))

summary(model)
```

And the printout is:

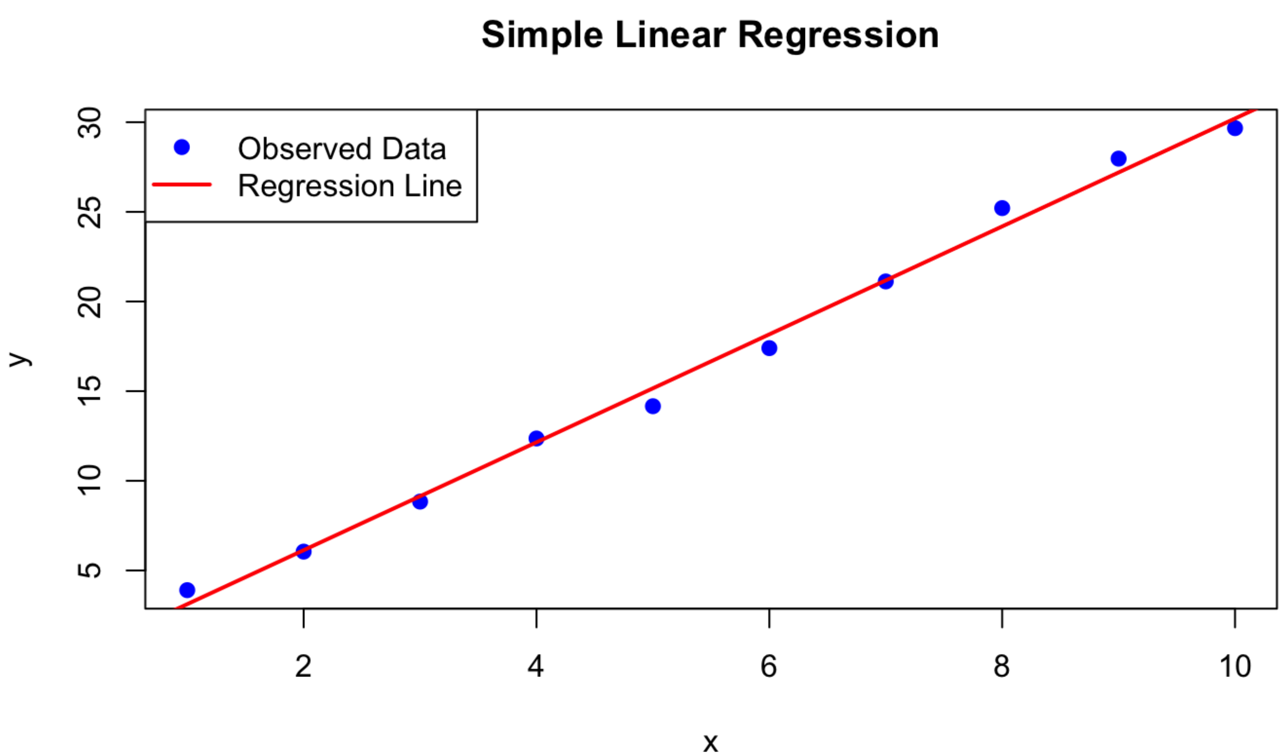
```
Estimate of beta1: 3.010545
SSX: 82.5
SSE: 4.270185
95% Confidence Interval for beta1: ( 2.825059 , 3.196032 )

Call:
lm(formula = y ~ x)

Residuals:
Min       1Q   Median       3Q      Max
-1.00273 -0.48450 -0.07245  0.62577  1.01564

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.11000    0.49909    0.22    0.831
x            3.01055    0.08044   37.43 2.85e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7306 on 8 degrees of freedom
Multiple R-squared:  0.9943,        Adjusted R-squared:  0.9936
F-statistic:  1401 on 1 and 8 DF,  p-value: 2.849e-10
```



3 Question 3

Let $M \in \mathbb{R}^{d\times d}$ be a real symmetric matrix (that is, all its elements are real numbers and $M_{ij} = M_{ji}$ for $i, j = 1, \ldots, d$).

3.1 Part a

Read and understand "Spectral theorem" and "Dyadic decomposition" on this document.

Summary of the Reading

Spectral Theorem
Let $A \in \mathbb{R}^{n\times n}$ be real and symmetric. Then:

1. The eigenvalues of A are real.
2. A is diagonalizable.
3. There is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

In short, A may be orthonormally diagonalized: $A = V\Lambda V^\top$ where $V \in \mathbb{R}^{n\times n}$ is an orthonormal matrix of eigenvectors of A, and $\Lambda \in \mathbb{R}^{n\times n}$ is a real diagonal matrix of eigenvalues.
Dyadic Decomposition
Dyadic decomposition is a concept from linear algebra that involves representing a matrix as a sum of dyadic products. Dyadic products, also known as outer products, are matrices obtained from two vectors. Specifically, if $\mathbf{u}$ and $\vec{v}$ are vectors, their dyadic product is a matrix $\mathbf{u}\vec{v}^T$, where $\vec{v}^T$ denotes the transpose of $\vec{v}$. 3.1 Dyadic Decomposition for Symmetric Matrices
Any symmetric matrix $A \in \mathbb{R}^{n\times n}$ can be represented as a sum of outer products of vectors, which can be considered a form of dyadic decomposition. This representation is closely related to the spectral decomposition of A. For a symmetric matrix, the dyadic decomposition can be expressed in terms of its eigenvectors and eigenvalues as follows:

$$A = \sum_{i=1}^{r} \lambda_i u_i u_i^T$$

where u_i are the eigenvectors of A, λ_i are the corresponding eigenvalues, and r is the rank of A. Each term $u_i u_i^T$ represents a dyadic product, contributing to the overall structure of A.

3.2 Part b

Suppose all the eigenvalues of $\boldsymbol{M}$ are non-negative, show that all the eigenvalues of $\boldsymbol{M} + \sigma^2 \boldsymbol{I}_{d\times d}$ are larger or equal to σ^2, where $\sigma > 0$ and $\boldsymbol{I}_{d\times d}$ is the $d \times d$ identity matrix.

Answer

We can apply the definition of spectrum theorem as learned before, since $\boldsymbol{M}$ is a symmetric matrix (which is a necessary condition for applying the theorem), it can be decomposed as:

$$\boldsymbol{M} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$$

where $\boldsymbol{Q}$ is an orthogonal matrix (i.e., $\boldsymbol{Q}^T = \boldsymbol{Q}^{-1}$), and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$ is a diagonal matrix consisting of the eigenvalues λ_i of $\boldsymbol{M}$, all of which are non-negative by assumption.

For eigenvalues of $\boldsymbol{M} + \sigma^2 \boldsymbol{I}_{d\times d}$, we can rewrite the combination as:

$$\boldsymbol{M} + \sigma^2 \boldsymbol{I}_{d\times d} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T + \sigma^2 \boldsymbol{I}_{d\times d}.$$

Using the property that the identity matrix commutes with any matrix,

$$\boldsymbol{M} + \sigma^2 \boldsymbol{I}_{d\times d} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T + \boldsymbol{Q}(\sigma^2 \boldsymbol{I}_{d\times d})\boldsymbol{Q}^T.$$

Since $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}_{d\times d}$, this simplifies to:

$$\boldsymbol{M} + \sigma^2 \boldsymbol{I}_{d\times d} = \boldsymbol{Q}(\boldsymbol{\Lambda} + \sigma^2 \boldsymbol{I}_{d\times d})\boldsymbol{Q}^T.$$

Since $\boldsymbol{\Lambda} + \sigma^2 \boldsymbol{I}_{d\times d}$ is a diagonal matrix with entries $\lambda_i + \sigma^2$, its eigenvalues are simply $\lambda_i + \sigma^2, \quad \text{for } i = 1, 2, \ldots, d.$

Since we assumed that $\lambda_i \geq 0$ for all i, it follows that $\lambda_i + \sigma^2 \geq \sigma^2$. Thus, every eigenvalue of $\boldsymbol{M} + \sigma^2 \boldsymbol{I}_{d \times d}$ is at least σ^2. $\square$