Q1

1.

(a.) $\text{trace}(X(X'X)^{-1}X') = \text{trace}((X'X)^{-1}X'X) = \text{trace}(I_k) = k = \text{trace}(P)$

$\text{trace}(M) = \text{trace}(I_n - P) = n - k$ #

(b.) $c \neq 0$

$c'Pc = c'P^2c = |Pc|^2 \geq 0$

$c'Mc = c'M^2c = |Mc|^2 \geq 0$

P, M are positive semi-definite.

Q2

2.

$$E[\hat{\sigma}_Y^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i^2 - 2\bar{Y}Y_i + \bar{Y}^2)\right]$$

$$= \frac{1}{n}\cdot\left(\sigma_y^2 + \mu_y^2 - 2\mu_y^2 - \frac{\sigma_y^2}{n} + \mu_y^2\right)\cdot n$$

$$= \frac{n-1}{n}\sigma_y^2$$

$$E[\hat{\sigma}_y^2] - \sigma_y^2 = \frac{-\sigma_y^2}{n}$$ #

Q3.

Direct computation

```
# 使用 (X'X)^(-1)*(X'Y) 計算 beta 的函數
calculate_beta_direct <- function(X, y) {
  X <- as.matrix(cbind(1, X))  # 添加截距
  solve(t(X) %*% X) %*% t(X) %*% y
}
```

```
Coefficients:
               Estimate Std.
(Intercept)    0.010799    0.
X_interest     0.144389    0.
x_dfy          0.997910    0.
x_infl        -0.554243    0.
x_svar        -0.449504    0.
x_tms          0.233549    0.
x_dfr          0.067591    0.
```

```r
# 準備數據
X <- dt[, c("x_dfy", "x_infl", "x_svar", "x_tms", "x_tbl", "x_dfr")]
y <- dt$y

# 計算 beta
beta_direct <- calculate_beta_direct(X, y)
beta_fwl <- calculate_beta_fwl(X, y)
```

FWL theorem

```r
# Define dependent variable (y) and independent variables (X)
X <- dt[, c("x_dfy", "x_infl", "x_svar", "x_tms", "x_tbl", "x_dfr")]
y <- dt$y

# Select 'x_tbl' as the variable of interest and the rest as control variables
X_interest <- X$x_tbl
X_controls <- X[, !colnames(X) %in% "x_tbl"]

# Step 1: Regress y on control variables (X_controls) and get residuals
lm_y_controls <- lm(y ~ ., data = as.data.frame(X_controls))
res_y_controls <- residuals(lm_y_controls)

# Step 2: Regress x_tbl on control variables (X_controls) and get residuals
lm_X_interest_controls <- lm(X_interest ~ ., data = as.data.frame(X_controls))
res_X_interest_controls <- residuals(lm_X_interest_controls)

# Step 3: Regress residuals of y on residuals of x_tbl
lm_residuals <- lm(res_y_controls ~ res_X_interest_controls)

# Display the results for FWL Theorem regression
summary(lm_residuals)

# Full model for comparison (regress y on both x_tbl and control variables)
lm_full <- lm(y ~ X_interest + ., data = as.data.frame(X_controls))
summary(lm_full)
```

```
均方誤差: 0.002002167
(Intercept) Intercept   0.01079947
x_dfy           x_dfy   0.99791023
x_infl         x_infl  -0.55424292
x_svar         x_svar  -0.44950368
x_tms           x_tms   0.23354924
x_tbl           x_tbl   0.14438939
x_dfr           x_dfr   0.06759062
```

Q4.

	Model	R_squared
1	M1	0.00000000
2	M2	0.01537088
3	M3	0.01545069
4	M4	0.02152345
5	M5	0.02250799
6	M6	0.03008154
7	M7	0.03059011

```r
x_dfy <- dt$x_dfy  # We'll use this as our condition variable
x_infl <- dt$x_infl
x_svar <- dt$x_svar
x_tms <- dt$x_tms
x_tbl <- dt$x_tbl
x_dfr <- dt$x_dfr
# Define the formulas for each regression
formulas <- list(
  y ~ 1,
  y ~ 1 + x_dfy,
  y ~ 1 + x_dfy + x_infl,
  y ~ 1 + x_dfy + x_infl + x_svar,
  y ~ 1 + x_dfy + x_infl + x_svar + x_tms,
  y ~ 1 + x_dfy + x_infl + x_svar + x_tms + x_tbl,
  y ~ 1 + x_dfy + x_infl + x_svar + x_tms + x_tbl + x_dfr
)
# Calculate R-squared for each formula
r_squared_values <- sapply(formulas, calculate_r_squared, data = dt)
# Create a data frame with the results
results <- data.frame(
  Model = c("M1", "M2", "M3", "M4", "M5", "M6", "M7"),
  R_squared = r_squared_values
)
# Print the results
print(results)
# Calculate the average R-squared
average_r_squared <- mean(r_squared_values)
cat("\nAverage R-squared:", average_r_squared)
# Plot the R-squared values
ggplot(results, aes(x = Model, y = R_squared)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  theme_minimal() +
  labs(title = "R-squared Values for Different Models",
       x = "Model",
       y = "R-squared")
```