



امتحان الفصل
للعام الجامعي 2025/2024

المادة: احصاء 3	المرحلة: السنة المنهجية: الثانية
المدة: دقيقة 90	الاستاذ: عباس رمال
الدورة: الاولى	

Exercise 1

The swiss dataset comes with base R. It contains socio-economic data for 47 French-speaking provinces of Switzerland in 1888. Load swiss dataset
data("swiss")
head(swiss)

Variable	Description
Fertility	Common standardized fertility measure (dependent variable)
Agriculture	% of males involved in agriculture as occupation
Examination	% of draftees with highest exam level
Education	% of draftees with secondary education
Catholic	% Catholic (religion)
Infant.Mortality	Infant mortality (per 1000 live births)

Part 1: Multiple Linear Regression Analysis

Fit the following multiple regression model:

$$Fertility = \beta_0 + \beta_1 Agriculture + \beta_2 Education + \beta_3 Catholic + \varepsilon$$

Then:

1. Parameter Estimates

- Display regression coefficients $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$

2. Confidence Intervals

- Compute 95% confidence intervals for all coefficients.

3. Residual Analysis

- Compute residuals and verify that their mean is approximately zero.
- Test whether the residuals are normally distributed and test the homogeneity of variances.

4. Variance Estimators

- Compute variance estimates for each parameter.

5. Coefficient of Determination

- Compute R^2 and Adjusted R^2 .
 - Interpret the model's explanatory power.
6. **Correlation Analysis**
- Compute pairwise correlations between Fertility, Agriculture, Education, and Catholic.
 - Comment on strength and direction.
7. **Hypothesis Testing for Coefficients**
- Test $H_0: \beta_i = 0$ for each coefficient at $\alpha = 0.05$ using t-tests.
 - Perform an F-test for overall model significance
8. **ANOVA Table**
- Construct the ANOVA table and interpret significance.
9. **Matrix Form**
- Write the design matrix X for the regression model.
10. **Leverages (Hat Values)**
- Calculate leverages.
 - Verify that sum of leverages = $p + 1$.
 - Identify how many leverage values exceed $2*(p+1)/n$.
11. **Studentized Residuals**
- Calculate the internal studentized residuals. How many points are suspected?
 - Calculate the external studentized residuals. How many points are suspected?
12. **Cook's Distance**
- Compute Cook's distance for each observation and identify influential points.

Part 2: Model Selection

Use **all predictors**: Agriculture, Examination, Education, Catholic, Infant.Mortality.

1. **Exhaustive Search**
 - Evaluate all possible subsets.
 - Select the best model using Adjusted R^2 , Mallows' C_p , R^2 , and BIC.
2. **Backward Elimination (Top-Down)**
 - Start with the full model.
 - Iteratively remove predictors using AIC and F-tests.
3. **Forward Selection (Bottom-Up)**
 - Start with intercept-only model.
 - Iteratively add predictors using AIC and F-tests.
4. **Stepwise Selection (Both Directions)**
 - Combine forward and backward procedures.
 - Select final model using AIC and F-tests.
5. **Stagewise Regression**
 - Perform incremental fitting of predictors.
 - Report the order in which predictors are added and comment on the final model.