

HW2 Due: February 28th

1 Probability integral transformation

Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Show that Y is uniformly distributed on $(0, 1)$, in other words $\Pr(Y \leq y) = y, 0 < y < 1$.

2 JF exercise 5.7

Consider the general multiple-regression equation

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon.$$

An alternative procedure for calculating the least-squares coefficient $\hat{\beta}_1$ is as follow:

1. Regress Y on X_2 through X_k , obtaining residuals $E_{Y|2\dots k}$.
2. Regress X_1 on X_2 through X_k , obtaining residuals $E_{1|2\dots k}$.
3. Regress the residuals $E_{Y|2\dots k}$ on the residuals $E_{1|2\dots k}$. The slope for this simple regression is the multiple-regression slope for X_1 , that is, $\hat{\beta}_1$.
 - (a) Apply this procedure to the multiple regression of the prestige on education and income. Confirm that the coefficient for education is properly recovered.
 - (b) The intercept for the simple regression in Step 3 is 0. Why is this the case?
 - (c) The procedure in this problem reduces the multiple regression to a series of simple regressions (in step 3). Can you see any practical application for this procedure?

3 Finish the following questions using R

1. Install the R package “carData”, read the documentation of the dataset “Highway1” under the package, list all variables in the “Highway1” dataset and explain what they are.
2. Use **rate** as the response variable, use all other variables **except** “htype” to fit a multiple linear regression and finish the following questions
 - (a) Calculate the total sum of squares, regression sum of squares and residual sum of squares
 - (b) Calculate the least square estimate by using equation $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$
 - (c) Get an estimate of the standard errors of the least square estimate of the coefficients.
 - (d) Test the null hypothesis of $H_0 : \beta_{trks} = 0$ vs $H_a : \beta_{trks} \neq 0$. Report the p-value.

- (e) Test the null hypothesis of $H_0 : \beta_{len} = \beta_{adt} = \dots = \beta_{twid} = 0$. Write out the alternative hypothesis. What test statistic do you get, report the associated p-value.
- (f) Calculate the variance inflation factors. Report your findings.