MSDS 596 Regression & Time Series

Fall 2018

Homework 2

Due: Tues 10/02/18 @ 6:40pm

rutgers.instructure.com/courses/17597

Problem 1. Let $X = (X_1, X_2, \ldots, X_n)$ be a n dimensional random vector with covariance matrix Σ_X. Let $\boldsymbol{A}$ be a $m \times n$ matrix, and define $Y = \boldsymbol{A}X$. Show that the covariance matrix of Y is

$$\text{Cov}(Y) = \boldsymbol{A}\Sigma_X \boldsymbol{A}'.$$

Problem 2. Let $X = (X_1, X_2)$ be bivariate normal with mean zero and covariance matrix $\Sigma = (\sigma_{ij})$.

(a) Write $X_2 = \alpha X_1 + Z$, where $\text{Cov}(Z, X_1) = 0$. Provide an explicit formula for α in terms of σ_{ij}.

(b) Explain why Z is normally distributed and independent of X_1. What is the variance of Z?

(c) Since Z is independent of X_1, conditional on $X_1 = x$ we have $X_2 = \alpha x + Z$. Derive the conditional mean and variance of X_2 given $X_1 = x$, without referring to the formula given in class.

Problem 3. Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. That is to say, $\boldsymbol{X} = (X_1, \ldots, X_n)' \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n)$. We estimate σ^2 with the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Let $\boldsymbol{1}_n \in \mathbb{R}^n$ denote the vector of 1's, and $\boldsymbol{P} = \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n'$. Recall what you've learned about $\boldsymbol{P}$ from the last homework.

(a) Show that

$$(n-1)\, s^2 = \boldsymbol{X}'\boldsymbol{P}\boldsymbol{X}.$$

(b) Find $\mathbb{E}(s^2)$.

(c) Find $Var(s^2)$.

(d) Find the distribution of $\frac{(n-1)s^2}{\sigma^2}$.

(e) Argue that s^2 and $\bar{X}$ are independent.

Problem 4. *(Two-sided one-sample Z-test).* Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where σ^2 is known. We want to test $H_0 : \mu = 1$ versus $H_1 : \mu \neq 1$. Consider the following quantity

$$Z := \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu).$$

(a) Show that Z is a pivotal quantity, and find its probability distribution.

(b) Construct a test statistic $T(\boldsymbol{X})$ based on the above pivotal quantity, and write down the form of the rejection region for this test in terms of a critical value c.

(c) Write down the power function for this test.

(d) Setting the test size at α, find an expression for the critical value c.

(e) Suppose $n = 9$, $\sigma^2 = 1$, and the observed $\bar{x} = 1.75$. What's the p-value for this test? Would you reject the null hypothesis at level 5%? What about 1%?

Problem 5. Using the `teengamb` dataset from the last homework, fit a model with expenditure on gambling (`gamble`) as the response, and all remaining variables as predictors.

(a) Which variables are significant at the 5% level? Give the respective confidence intervals for their associated coefficients.

(b) What interpretation should be given to the coefficient for `sex`?

(c) Fit a model with just `income` as a predictor. Compare it to the full model using an F-test.

(d) State the null hypothesis of the above F-test as an assertion involving the regression coefficients from the full model.

Problem 6. The following is the R output of a simple linear regression analysis. Figure out the three entries indicated by "???". Show your work in detail.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.0983     0.1589   6.912 1.63e-07 ***
x             0.7190     0.1535     ???      ??? ***
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.8662 on 28 degrees of freedom
Multiple R-squared: ??? ,
F-statistic: 21.94 on 1 and 28 DF,  p-value: 6.579e-05
```