Stat GR4205/5205, Fall 2017

Midterm, October 20, 2017

Instructor: Jingchen Liu

Instructions. This exam has four problems, each with several sub-problems. Write your answer in the blank space provided below each sub-problem. The space provided should be more than enough (e.g., if you find yourself needing a lot more space than provided you should be concerned with the approach you're using), but in case you do need more space you can use the back of the corresponding page as long as you clearly indicate that you are doing so. The sub-problems are arranged in a logical order and thus you should try to complete them in the order given. However, you should not "get stuck" on a sub-problem if you are unable to complete it. You should move on and try to complete the rest by assuming any stated results in the preceding parts are true even if you cannot prove them. Show your work in order to receive full credit (i.e., don't just provide a numerical answer – show, at least briefly, how you arrive at your answer). The total score is 100.

IMPORTANT

(1) Print your name below and then sign

NAME: ____________________

SIGNATURE:______________________

(2) Put your initial at the upper right corner of each page

Distribution of credits

Problem 1 (25)	
Problem 2 (40)	
Problem 3 (25)	
Problem 4 (10)	
Total (100)	

[You must include intermediate steps for each problem, otherwise the solution is considered as incomplete!]

Problem 1 (25 Points)

Consider the simple linear regression $y_i = \beta_1 x_i + \varepsilon_i$ for $i = 1, ..., n$, that is, the intercept is set to zero. Derive the least squares estimator of β_1 without an intercept and show your intermediate steps. You are not supposed to cite any results for least squares estimators developed in the class.

Problem 2 (40 Points)

Consider bivariate i.i.d. random variables. $(X_1, Y_1), ...(X_n, Y_n)$ where $n = 100$. The summary statistics (mean, standard deviation, and correlation) is given as follows

	Mean	SD	Correlation between X and Y
X	0.50	0.50	0.21
Y	-0.04	0.91	

Consider a simple linear regression model

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2)$. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the least squares estimates.

(a) (25 points). Based on the summary statistics, complete the following two regression tables. You must include intermediate steps, otherwise the solution is considered as incomplete!

Least square estimate

	Estimate	Standard Deviation
$\hat{\beta}_0$	( )	( )
$\hat{\beta}_1$	( )	( )

Analysis of variance table

	Sum of squares	d.f.	Mean squares
Regression	( )	( )	( )
Sum of squares of residuals	( )	( )	( )
Total	( )	( )	

(b) (15 points). If X is a Bernoulli random variable taking values in $\{0, 1\}$. We order the observations in such a way that the first 50 observations have $X_i = 0$ $(1 \leq i \leq 50)$ and the later 50 observations have $X_i = 1$ $(51 \leq i \leq 100)$. Consider that X is a group indicator and Y is a response variable. Perform a two sample t-test for which

$$\begin{aligned} H_0 &: \mu_0 = \mu_1 \\ H_A &: \mu_0 \neq \mu_1 \end{aligned}$$

where $\mu_j = E(Y|X = j)$ for $j = 0, 1$. Compute the t-statistic.

Problem 3 (25 Points)

Consider simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \tag{1}$$

Let $z_i = a + bx_i$ and

$$y_i = \gamma_0 + \gamma_1 z_i + \delta_i. \tag{2}$$

Show that the predicted values of the least-square estimators in (1) and (2) are identical for all x_i and z_i.

Problem 4 (10 points)

Show that the distribution of the p-value of a t-test under the null hypothesis is uniformly distributed over the interval $[0, 1]$.

For Grading Purpose Only

Problem 1 (25)	
Problem 2 (40)	
Problem 3 (25)	
Problem 4 (10)	
Total (100)	