**EE325: Midterm Exam Semester 1, 2018**

## 1. (15 points) Probability and Basic Statistics

| | A = -5 | A = 0 | A = 5 |
|---|---|---|---|
| B = 1 | 0.10 | 0.05 | 0.25 |
| B = 2 | 0.05 | 0.15 | 0.05 |
| B = 3 | 0.20 | 0.10 | 0.05 |

(a) (5 points) What are the expectations of A and B?
(b) (5 points) Are A and B independent, why or why not? (show your calculation)
(c) (5 points) What is the conditional expectation of B given A = 5

**2. (20 points) Simple Linear Regression**

All the variables have the same meaning and same properties as stated in class.

Given the following linear regression function.

$$Y_i = \beta_0 + u_i,$$

where $\beta_0$ is a constant parameter and $u_i$ is the error term that is independently and identically distributed with expected value of 0 and variance $\sigma^2$.

(a) (10 points) Find the OLS estimator of $\beta_0$.

(b) (10 points) Calculate $Var(\hat{\beta}_0)$.

**3. (15 points) Multiple regression**

Suppose that you are asked to conduct a study to determine whether the time spent on commuting to the university has an impact on B.E. student's GPA. Given that you are under a 1-month time constraint and has no financial support, explain step-by-step how you would conduct your research on order to obtain an unbiased estimator of the impact of commuting time to the university and B.E. student's GPA. (Also state what variables you will collect.) *<You can continue writing on the back of this page>*

**4. (15 points) Simple Regression**

You would like to conduct a study to find the impact of number of hours each person sleeps at night and their yearly earnings. Suppose you obtained a random draw of 706 samples. Let $Y_i$ be yearly earnings (measured in USD) and $X_i$ be the average number of hours slept per night (measured in minutes). Preliminary analysis of the sample data produces the following sample information:

$$\sum Y_i = 653{,}250 \sum X_i = 155{,}374 \sum \hat{u}_i^2 = 8{,}842{,}612{,}828 \sum \hat{u}_i = 0.02062$$

$$\sum x_i y_i = -69{,}201{,}310 \sum y_i^2 = 9{,}212{,}351{,}250 \sum x_i^2 = 12{,}951{,}888.48 \quad N = 50$$

where $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$, $\hat{u}_i = Y_i - \widehat{Y}_i$, $N =$ sample size

Use the above information to answer the following questions. **Show all the formulas and your calculation.**

(a) (7 points) Use the above information to compute OLS estimates of the intercept coefficient $\beta_0$ and that of the slope coefficient $\beta_1$.

(b) (3 points) Interpret the slope coefficient estimate you calculated in part (2.1) – i.e., explain in words what the numeric value you calculated for $\beta_1$ means.

(c) (5 points) Compute the value of $r^2$. Explain the rationale behind the calculation of $r^2$.

**5. Multiple Regression (30 points)**

Using the data in NBASAL.DTA, we estimate the below model and obtain the following results:

$$wage = \beta_0 + \beta_1 game + \beta_2 age + u.,$$

```
. regress wage game age

      Source |       SS       df       MS              Number of obs =     269
-------------+------------------------------           F(  2,   266) =   30.44
       Model |  49885642.7     2  24942821.4           Prob > F      =  0.0000
    Residual |   217993283   266  819523.619           R-squared     =  0.1862
-------------+------------------------------           Adj R-squared =  0.1801
       Total |   267878925   268  999548.229           Root MSE      =  905.28

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       games |   14.03942   2.957143     4.75   0.000     8.217038    19.86181
         age |   91.09241   16.43781     5.54   0.000     58.72763    123.4572
       _cons |  -1994.302   470.6061    -4.24   0.000    -2920.889   -1067.715
------------------------------------------------------------------------------
```

where

wage = annual salary of US National Basketball Association (NBA) players (in thousand US dollars)

games = average games that the player played per year (games/year)

age = age of the player (years)

(a) (5 points) Why do we usually need to include more than 1 explanatory variable in a regression model?

(b) (5 points) Do "*games*" and "*age*" have the expected effects? Explain.

(c) (5 points) If "*experience*" is omitted from this equation, what should be the direction of bias of $\beta_1$ and $\beta_2$? Explain.

(d) (5 points) If you suspect that "*age*" has a positive impact on "*wage*", but the magnitude of impact is diminishing every year, how would you modify this regression specification? Write down your suggested regression function and give explanation.

(e) (5 points) Do you think this model includes all the factors that explain wage? If not, propose 2 other factors. Explain what signs you think their $\beta$ will be.

(f) (5 points) Suppose the two factors that you proposed in part 3.5 are irrelevant (meaning that they are not in the population regression model), but you include them in your regression analysis anyway. What would happen to your estimator of $\beta_1$ and $\beta_2$?

<The End of Exam>