Midterm Exam

Instructions: This is a closed book exam. You may use a calculator. The use of other electronic devices, such as cellphones, laptops, or tablets, is prohibited. There are 34 points on this 70-minute exam. The possible points for each question are shown in parentheses following the question number. Answers should be written in English. Please write legibly, and do not forget to write your name and student id number on the answer sheet. Good luck!

Part I: Multiple choice questions. Select one or two choices that are correct unless stated otherwise. You do not have to explain your answer.

1. (2 points) Consider an i.i.d. sample of (Y_1, Y_2, … , Y_n) drawn from a population with mean μ_Y and variance σ_Y^2.
    a. Y_2 is a random variable.
    b. $\overline{Y}$ is the ordinary least squares (OLS) estimator of μ_Y.
    c. $\overline{Y}$ is normally distributed for any n.
    d. $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ is an unbiased and consistent estimator of σ_Y^2.
    e. None of the above.

2. (2 points) In econometrics, we typically do not rely on exact or finite sample distributions for statistical inference because
    a. variables typically are normally distributed.
    b. we have an infinite number of observations.
    c. observations in our data are correlated with one another.
    d. asymptotic distributions provide good approximations when the sample size is reasonably large.
    e. None of the above

3. (2 points) Suppose you estimate a simple linear regression model, $Y_i = \beta_0 + \beta_1 X_i + u_i$. To determine whether the estimated slope is small or large in magnitude,
    a. you should see if R^2 is close to one.
    b. you should see if the slope coefficient is statistically significantly different from zero.
    c. you should analyze the economic importance of the effect of X on Y, implied by the slope coefficient.
    d. you should change the scale of the regressor so that the slope coefficient appears to be large.
    e. None of the above.

4. (2 points) The standard error of the regression (*SER*)
   a. is the fraction of the sample variance of the dependent variable explained by the independent variable.
   b. is an estimator of the standard deviation of the regression error.
   c. ranges between -1 and 1.
   d. measures the goodness of fit of the regression model.
   e. None of the above.

5. (2 points) The slope estimator $\hat{\beta}_1$ in the simple linear regression model, $Y_i = \beta_0 + \beta_1 X_i + u_i$, has a larger standard error, other things equal, if
   a. the sample size is larger.
   b. the estimated intercept $\hat{\beta}_0$ is larger.
   c. there is more variation in the explanatory variable, X_i.
   d. there is a smaller variance of the error term, u_i.
   e. None of the above.

Part II: Longer questions. Show your work and explain your answers.

6. (13 points) Consider an exit poll of randomly selected voters after an election with two candidates (i.e., the incumbent and the challenger). When we denote each voter's choice by Y, $Y = 1$ if the person votes for the incumbent and $Y = 0$ if the person votes for the challenger. Y is a Bernoulli random variable with probability $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = 1 - p$. Think of the exit poll as reporting $\hat{p}$, the fraction of survey respondents who have voted for the incumbent. The following table presents the poll results by gender.

	Proportion selecting the incumbent ($\hat{p}$)	Number of observations (n)
Men	0.49	600
Women	0.43	500

   (a) (2 points) Estimate $\text{var}(\hat{p})$ for each gender: $\widehat{\text{var}(\hat{p}_m)}$ for men and $\widehat{\text{var}(\hat{p}_w)}$ for women. Report to five decimal places.

   (b) (4 points) Does the poll contain statistically significant evidence that the challenger is ahead of the incumbent among male voters?
      i. (1 point) Postulate a hypothesis test by stating the null and the alternative hypotheses.
      ii. (1 point) Compute the relevant *t*-statistic.
      iii. (1 point) Compute the *p*-value associated with the *t*-statistic.
      iv. (1 point) Can you reject the null hypothesis at the 10% significance level? Explain.

   (c) (7 points) Is there a statistically significant difference in the share of men and the share of

women voting for the incumbent (denoted by p_m and p_w, respectively)?

i. (1 point) Estimate the gender difference in the vote share, $p_m - p_w$.

ii. (2 points) Compute the standard error of the difference estimator, $\hat{p}_m - \hat{p}_w$.

iii. (2 points) Construct a 95% two-sided confidence interval for the gender difference in the vote share, $p_m - p_w$.

iv. (2 points) Can you reject the null hypothesis that there is no gender difference in the vote share, at the 5% significance level? Explain.

7. (11 points) An economic study analyzes observational data on chief executive officers (CEO) for U.S. corporations, drawn by simple random sampling. The study examines a relationship between CEO tenure and CEO pay by estimating the simple linear regression model, $salary_i = \beta_0 + \beta_1 ceoten_i + u_i$. Use the following R outputs to answer the questions. Recall that you should use heteroskedasticity-robust standard errors whenever you report standard errors, use them for further calculations, or conduct a hypothesis test.

Variables in the dataset:

`salary`	annual compensation (in 1,000 U.S. dollars)
`ceoten`	years as company CEO

R outputs:

```
> library(sandwich)
> library(lmtest)

> nrow(data)
[1] 177
> summary(data$ceoten)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   6.000   7.955  11.000  37.000

> ols <- lm(salary ~ ceoten, data)
> summary(ols)

Call:
lm(formula = salary ~ ceoten, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-977.8 -345.9 -169.4  263.8 4373.9

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  772.426     65.676  11.761   <2e-16 ***
ceoten        11.746      6.148   1.911   0.0577 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 583.2 on 175 degrees of freedom
Multiple R-squared:  0.02043,    Adjusted R-squared:  0.01484
F-statistic: 3.651 on 1 and 175 DF,  p-value: 0.05769
```

```
> Cov1 <- vcovHC(ols, type = "HC1")
> Cov1
             (Intercept)     ceoten
(Intercept)   2395.4089 -166.49852
ceoten        -166.4985   34.37488

> coeftest(ols, vcov. = Cov1)

t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  772.426     48.943 15.7822  < 2e-16 ***
ceoten        11.746      5.863  2.0034  0.04667 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(a) (5 points) Report the estimated coefficients and their standard errors. Interpret the estimated coefficients. Also, mention whether each coefficient is statistically significantly different from zero at the 5% level.

(b) (2 point) What is the average annual salary in the sample?

(c) (1 point) Report $\widehat{\text{var}}(\hat{\beta}_0)$ and $\widehat{\text{var}}(\hat{\beta}_1)$.

(d) (3 points) Suppose that we use monthly salary in U.S. dollars instead of annual salary in 1,000 US dollars as the dependent variable in the regression analysis. How will the regression results change? Compute $\hat{\beta}_0$, $\hat{\beta}_1$, $SE(\hat{\beta}_0)$, $SE(\hat{\beta}_1)$, R^2, and SER.