2021 Midterm Exam (1 hour and 50 minutes)

Critical Values		
$F_{2,N,1-\alpha} = 3.00$	$F_{3,N,1-\alpha} = 2.60$	$F_{6,N,1-\alpha} = 2.10$
$t_{12,1-\alpha/2} = 2.18$	$t_{27,1-\alpha/2} = 2.05$	$t_{21,1-\alpha/2} = 2.08$
$\alpha = 0.05,\ N > 200$		

Problem 1. (**20 points**) Consider a simple linear regression model:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + U_i,\ i = 1, \ldots, n; \\ \mathrm{E}\left(U_i \mid X_1, \ldots, X_n\right) &= 0, \quad (1) \\ \mathrm{E}\left(U_i^2 \mid X_1, \ldots, X_n\right) &= \sigma^2, \quad (2) \\ \mathrm{E}\left(U_i U_j \mid X_1, \ldots, X_n\right) &= 0 \text{ for } i \neq j. \quad (3) \end{aligned}$$

Let $\hat{\beta}_1$ denote the OLS estimator of β_1 in this model. For each of the following statements, indicate true or false and **explain your answers**.

1. $\hat{\beta}_1$ is biased if condition (1) is replaced by $\mathrm{E}\left(U_i X_i\right) = 0$.
2. $\hat{\beta}_1$ is biased if condition (2) fails.
3. $\hat{\beta}_1$ is biased if condition (3) fails.
4. The variance formula

$$\mathrm{Var}\left(\hat{\beta}_1 \mid X_1, \ldots, X_n\right) = \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

   is invalid if conditions (2) and (3) fail.

5. Suppose that an econometrician generates $Y_i^* = Y_i - \overline{Y}$ and $X_i^* = X_i - \overline{X}$. He runs an OLS regression **without an intercept** of Y^* on X^* and finds that the OLS estimate is equal to $\hat{\beta}_1$.

Problem 2. (10 points) Let Y and X be two random variables.

1. Show that $\mathrm{E}[Y \mid X]$ and $Y - \mathrm{E}[Y \mid X]$ are uncorrelated (i.e., $\mathrm{Cov}\left(\mathrm{E}[Y \mid X], Y - \mathrm{E}[Y \mid X]\right) = 0$). **Hint**: Use law of iterated expectations.
2. Show that $\mathrm{Var}(Y) \geq \mathrm{Var}(Y - \mathrm{E}[Y \mid X])$. **Hint**: Use (i).

Problem 3. (12 points) Find A-C in the Stata output below, which was obtained by running a regression of the dependent variable y agains a single regressor x (**without an intercept**). **Hint**: be careful about the degree of freedom. The number of observations was 13. Is the regressor x significant? The significance level is 5%.

```
------------------------------------------------------------------------------
           y       Coef.  Std. Err.       t    P>t     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x    1.014672  A          1.083    0.301             B          C
------------------------------------------------------------------------------
```

Problem 4. (18 Points) You are conducting an econometric investigation into the hourly wage rates of male and female employees. Your particular interest is in comparing the determinants of wage rates for female and male workers. The sample data consist of a random sample of observations on 526 paid employees, 252 of whom are females and 274 of whom are males. The sample data provide observations on the following variables:

W = hourly wage rate, measured in dollars per hour

S = number of years of formal education completed, in years

A = age, in years;

T = firm tenure of employee, in years;

F = 1 if employee is female and = 0 if employee is male

The following model is proposed

$$\begin{aligned}\log(W) &= \beta_1 + \beta_2 S + \beta_3 A + \beta_4 A^2 + \beta_5 T + \beta_6 (S \cdot T) + \beta_7 F + \beta_8 (F \cdot S) \\ &\quad + \beta_9 (F{\cdot}A) + \beta_{10}(F{\cdot}A2) + \beta_{11}(F{\cdot}T) + \beta_{12}(F{\cdot}S{\cdot}T) + U.\end{aligned}$$

The observed sample provides the following estimates (standard errors of each estimate in parenthesis),

$$\begin{aligned}\log(W) &= \underset{(\ 0.2385\ )}{-0.5667} + \underset{(0.01104)}{0.05937} S + \underset{(0.01216)}{0.07980} A - \underset{(0.00015)}{0.00093} A^2 - \underset{(0.01128)}{0.01057} T \\ &\quad + \underset{(0.00087)}{0.00227} (S{\cdot}T) + \underset{(0.3373)}{0.03593} F + \underset{(0.01684)}{0.01684} (F{\cdot}S) \\ &\quad - \underset{(0.01715)}{0.03847} (F{\cdot}A) + \underset{(0.000219)}{0.000422} (F{\cdot}A^2) + \underset{(0.02652)}{0.01850} (F{\cdot}T) - \underset{(0.002187)}{0.002107} (F{\cdot}S{\cdot}T).\end{aligned}$$

The corresponding residual sum of squares (SSR) and the total sum of squares (SST) of the n = 526 observations are:

$$SSR = \sum_{i=1}^{n} \hat{U}_i^2 = 80.57$$

$$SST = \sum_{i=1}^{n} (\log(W_i) - \overline{\log W})^2 = 148.33.$$

1. Write the expression for the marginal effect of A over $\log(W)$ for male employees in terms of the unknown parameters, implied by regression equation. Repeat the exercise for female employees. Provide the null hypothesis that the marginal effect of A on $\log(W)$ for male employees is equal to the marginal effect of A on $\log(W)$ for female employees. The restricted OLS estimation provides $SSR = 81.7242$. Perform a F test ($q = 2$)and state your conclusion from the test.

2. Write the restriction over the parameters that the marginal effects of S over $\log(W)$ and of T over $\log(W)$ are equal for male and female employees. The restricted OLS estimation provides $SSR = 80.8747$. Perform a F test ($q = 3$) and state your conclusion from the test.

3. Write the restriction over the parameters that the regression equations are identical for male and female employees, that is, the mean log-wage of female employees with any given values of S, A and T equals the mean log-wage of male employees with the same values of S, A and T. The restricted OLS estimation provides $SSR = 93.1805$. Perform a F test ($q = 6$) and state your conclusion from the test.

Problem 5. (20 Points) Consider a regression of Y_i against a constant and X_i. Let $\hat{\beta}_0$, $\hat{\beta}_1$, and s^2 denote the estimated intercept, estimated slope parameter, and estimator of the variance of errors from that regression. Let T denote the t-statistic for testing H_0 that the slope parameter is zero in that regression. Let *pval* be the corresponding p-value. Now, let c_1 and c_2 be two constants ($c_2 \neq 0$). Define a new dependent variable and a new regressor as

$$Y_i^* = c_1 Y_i,$$
$$X_i^* = c_2 X_i.$$

Let $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and s_*^2 denote the estimated intercept, estimated slope parameter, and estimator of the variance of errors from the regression of Y_i^* against a constant and X_i^*. Let T^* denote the t-statistic for testing H_0 that the slope parameter in the regression of Y_i^* against a constant and X_i^* is zero. Let $pval^*$ be the corresponding p-value.

1. Find an expression for $\hat{\beta}_1^*$ in terms of $\hat{\beta}_1, c_1$, and c_2.
2. Find an expression for $\hat{\beta}_0^*$ in terms of $\hat{\beta}_0$ and c_1.
3. Find an expression for s_*^2 in terms of s^2 and c_1.
4. What is the relationship between T and T^*?
5. What is the relationship between $pval$ and $pval^*$?

Problem 6. (20 Points) Consider the following estimated regression model:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{U}_i,$$

where $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ are the OLS estimators and $\hat{U}_i$'s are the OLS residuals, i.e. $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$ are such that for $\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}$,

$$\sum_{i=1}^n \hat{U}_i = \sum_{i=1}^n \hat{U}_i X_{1i} = \sum_{i=1}^n \hat{U}_i X_{2i} = 0.$$

Let $\tilde{X}_{2i}$ be the OLS residuals from the regression of X_{2i} against a constant and X_{1i}:

$$\tilde{X}_{2i} = X_{2i} - \hat{\gamma}_0 - \hat{\gamma}_1 X_{1i},$$

where $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are such that

$$\sum_{i=1}^n \tilde{X}_{2i} = \sum_{i=1}^n \tilde{X}_{2i} X_{1i} = 0.$$

Let $\tilde{Y}_i$ be the OLS residuals from the regression of Y_i against a constant and X_{1i}:

$$\tilde{Y}_i = Y_i - \hat{\delta}_0 - \hat{\delta}_1 X_{1i},$$

where $\hat{\delta}_0$ and $\hat{\delta}_1$ are such that

$$\sum_{i=1}^n \tilde{Y}_i = \sum_{i=1}^n \tilde{Y}_i X_{1i} = 0.$$

Show that

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n \tilde{X}_{2i} \tilde{Y}_i}{\sum_{i=1}^n \tilde{X}_{2i}^2}.$$

Hint: Show $\sum_{i=1}^n \tilde{X}_{2i} Y_i = \hat{\beta}_2 \sum_{i=1}^n \tilde{X}_{2i} X_{2i} + \sum_{i=1}^n \tilde{X}_{2i} \hat{U}_i$ and then show $\sum_{i=1}^n \tilde{X}_{2i} X_{2i} = \sum_{i=1}^n \tilde{X}_{2i}^2$, $\sum_{i=1}^n \tilde{X}_{2i} \hat{U}_i = 0$ and

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n \tilde{X}_{2i} Y_i}{\sum_{i=1}^n \tilde{X}_{2i}^2}.$$