FINAL EXAM. ECON 229, FALL 2015. NAME: ________________________________

Show your work. Box your final answers. Each question is worth one point.

Part I: Calculation questions

1. Simple regression coefficients. Label and use the blank columns to find a and b, the OLS estimates of α and β in the regression model $y_i = \alpha + \beta x_i + \varepsilon_i$.

x	y				
2	17				
2	11				
4	9				
4	7				
6	3				
6	1				

2. Correlation. Label and use the blank columns to find r, the correlation between x and y below.

x	y					
8	24					
8	21					
9	20					
10	19					
11	18					
14	18					

3. Residuals. We are estimating the regression model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2_i} + \varepsilon_i$ using the data below. We find that the OLS estimates of the βs are $b_0 = 10$, $b_1 = 1$, and $b_2 = 1$. Taking this as given, label and use the blank columns to find s^2, the estimated variance of the error term ε_i. (Hint: with two x variables, we write $K = 3$ for the number of regressors including the intercept term.)

x_1	x_2	y			
1	4	13			
4	1	17			
3	3	16			
4	4	18			
5	5	20			
7	4	19			
4	7	23			

4. Confidence intervals and hypothesis testing. Suppose that we are estimating a linear regression model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$. We find that the OLS estimate of β_1 is $b_1 = 35$, and that the standard error of b_1 is $SE(b_1) = 20$. Find the t-statistic, the p-value associated with the null hypothesis that $\beta_1 = 0$ (i.e. x_1 has no effect on y), and a 95% confidence interval for β_1. You can use a normal approximation for the distribution of b_1; see e.g. the attached table for reference.

5. Polling. We have conducted a random poll of 36 voters, in which 21 support Zack, and 15 support Jessie. Give an estimate of Zack's support with a margin of error at the 95% confidence level. You can report your answers in terms of fractions or percentages. You can approximate $\Phi^{-1}\left(\frac{.95+1}{2}\right) \approx 1.96$ as 2 for simplicity, if you like.

Part II: Derivation and explanation questions

6. Regression coefficients, part 1. We believe that a dependent variable y is affected in a linear way by one measurable variable x, and by a random error term which we denote as ε. We want to estimate the true data-generating process $y_i = \alpha + \beta x_i + \varepsilon_i$ with $y_i = a + bx_i + e_i$. In mathematical terms, what exactly do the OLS estimates a and b minimize? Write this out in terms of a, b, the x_is, and the $y_i s$. What is the motivation for minimizing this? Illustrate with a diagram.

7. Regression coefficients, part 2. The minimization problem described above implies that $\sum_i \left(-2X_i(Y_i - bX_i)\right) = 0$, where $X_i \equiv x_i - \bar{x}$ and $Y_i \equiv y_i - \bar{y}$. Use this to solve for b in terms of the data.

8. Single vs. multiple regression analysis. Suppose that a dependent variable y is impacted by two independent variables, x_1 and x_2 in a linear manner; i.e. $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$. I run an OLS regression of y on x_1 and x_2, and find that $b_1 > 0$, and $b_2 < 0$. However, if I regress y only on x_1, I find that the estimated slope coefficient, b, has a negative sign. (You can assume that all of the p-values in both regressions are very small.) What does all of this imply about the sign of the correlation between x_1 and x_2? Explain as clearly as you can.

9. Variance of the slope term. We are estimating $Y_i = \beta X_i + \varepsilon_i$ with $Y_i = bx_i + e_i$, where $X_i = x_i - \bar{x}$ and $Y_i = y_i - \bar{y}$. Using the formula for b as a starting point, derive the formula for $\mathrm{var}(b)$, the variance of the OLS estimate of β, in terms of the data and in terms of σ^2, the true variance of the error term ε_i.