Question 1. (20 pts.) Stock Prices, Y, are assumed to be affected by the annual rate of dividend of stock, X. A simple linear regression analysis was performed on 20 observations and the results obtained is

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level
INTERCEPT	-7.964633	3.11101359	-2.560	0.0166
X1	12.548580	1.27081204	9.874	0.0001

Circle the correct answer.

1. What statistical conclusion should you make about the effect of the dividend on average stock price?
   A. Since 11.30869 > table value, reject the null hypothesis.
   B. Since 12.54858 > table value, reject the null hypothesis.
   C. Since 9.874 < table value, reject the null hypothesis.
   D. Since 9.874 > table value, reject the null hypothesis.
   E. Since 0.7895 < table value, fail to reject the null hypothesis.

2. What is the 95% confidence interval for a value of Y given an X value of 2.36? You are given the standard error of this estimate is 3.351. I am 95% confident that
   A. the stock price for a stock with a dividend rate of 2.36% falls between 14.61 and 28.69.
   B. the mean stock price for all stocks with a dividend rate of 2.36 % falls between 14.61 and 28.69.
   C. the variance in stock price for all stocks falls between 14.61 and 28.69.
   D. the dividend rate for all stocks falls between 14.61 and 28.69.
   E. for each one point increase in dividend rate, the stock price will increase from 14.61 and 28.69

3. Which one of the following assumptions is incorrectly stated?
   A. The stock price is normally distributed for any dividend rate.
   B. The stock price has the same variability for any dividend rate.
   C. The stock price for any dividend rate is a linear function of dividend rate. (The mean stock price falls on the line).
   D. The difference between the stock price and the expected stock price given the dividend rate is independent from company to company.

4. The interpretation of 0.7895, the value of R-square is
   A. 78.95% of the sample stock prices (around the mean stock price) can be attributed to a linear relationship with the dividend rate in the population.
   B. the mean stock price will be estimated to increase 97.50 for each point increase in the rate.
   C. the mean stock price will be increase 78.95 for each point increase in the rate.
   D. the stock price will increase 78.95 for each point increase in the rate.
   E. 78.95% of the sample variability in stock price (around the mean stock price) can be attributed to a linear relationship with the dividend rate.

5. What is the estimate of the change in expected stock prices when the dividend rate increases by one point?
   A. 97.50
   B. -7.964633
   C. This is a parameter not a statistic.
   D. 12.54858
   E. 5.36546

6. The estimate of the standard deviation of $\hat{\beta}$ is:
   A. 3.36284
   B. 3.14983
   C. 0.39274
   D. 12.54858
   E. 1.27081

2 (Explanation)

95% Confidence interval of mean stock price at dividend 2.36 =

mean stock price $\pm$ standard error $\times$ $t_{n-2,0.975}$

$= (-7.964 + 12.54 \times 2.36) \pm 3.351 \times 2.10$

$= [14.61, 28.69]$

Question 2. (20 pts.) A study was conducted among a group of people who underwent coronary angiography. A group of 1493 people with coronary-artery disease were identified and were compared with a group of 707 people without the disease (controls). Risk factor information was collected on each group. Among cases, the mean serum cholesterol was 234.8 mg/dL with standard deviation = 47.3 mg/dL. Among controls, the mean serum cholesterol was 215.5 mg/dL with standard deviation = 47.3 mg/dL. What test is appropriate to determine whether the true mean serum cholesterol is different between the two groups?

Let x_i be the serum cholesterol for the ith case and y_j be the serum cholesterol level for the jth control.

$x_i \sim N(\mu_1, \sigma_1^2)$, $y_j \sim N(\mu_2, \sigma_2^2)$.

We wish to test $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$.

Since $s_1 = s_2$, we will assume equal variances i.e., $\sigma_1^2 = \sigma_2^2$. Then we use two-sample t-test for independent samples with equal variances

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{234.8 - 215.5}{\sqrt{47.3^2\left(\frac{1}{1493} + \frac{1}{707}\right)}} = \frac{19.3}{2.159}$$

$$= 8.94 \sim t_{1493+707-2} = t_{2198}$$

Since $t > t_{120,\,0.9995} = 3.373 > t_{2198,\,0.9995}$

it follows that p-value $< 2 \times (1 - 0.9995)$
0.001

Question 3. (20 pts.) Much controversy has risen concerning the possible association between myocardial infarction (MI) and coffee drinking. Suppose the information in Table 1 on coffee drinking and prior MI status is obtained from 200 60-64-year old males in the general population.

Table 1: MI and coffee drinking

Coffee drinking (cups/day)	MI in last 5 years	Number of people
0	Yes	3
0	No	57
1	Yes	7
1	No	43
2	Yes	8
2	No	42
3 or more	Yes	12
3 or more	No	28

Test for the association between history of MI and coffee drinking status, which is categorized as follows: 0 cups, 1 or more cups.

Ho: association

H_1: not association

Construct the ~~following~~ table

MI \ Coff	0	1+	~~2~~	~~3+~~	
Yes	3	27			30
No	57	113			170
	60	140			

Use chisquare test for 2×2 table

Since all expected value > 5 (smallest expected value $= \frac{60 \times 30}{200} = 9$)

$$\chi^2 = \frac{n\left(|ad-bc| - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{200\left(|3\times113 - 57\times27| - \frac{200}{2}\right)^2}{30\times170\times60\times140}$$

$$= 5.65$$

$\chi^2_{1,0.975} = 5.02$. Reject at 5%