THIS IS NOT REPRESNTATIVE OF CURRENT CLASS MATERIAL

STOR 455 Midterm 1 September 28, 2010

INSTRUCTIONS:
BOTH THE EXAM AND THE BUBBLE SHEET WILL BE COLLECTED. YOU MUST PRINT YOUR NAME AND SIGN THE HONOR PLEDGE ON THE BUBBLE SHEET. YOU MUST BUBBLE-IN YOUR NAME & YOUR STUDENT IDENTIFICATION NUMBER.

EACH QUESTION HAS ONLY ONE CORRECT CHOICE (decimals may need rounding).

USE "NUMBER 2" PENCIL ONLY - DO NOT USE INK - FILL BUBBLE COMPLETELY.

NO NOTES OR REMARKS ARE ACCEPTED - DO NOT TEAR OR FOLD THE BUBBLE SHEET.

A GRADE OF ZERO WILL BE ASSIGNED FOR THE ENTIRE EXAM IF THE BUBBLE SHEET IS NOT FILLED OUT ACCORDING TO THE ABOVE INSTRUCTIONS.

QUESTIONS are worth **1 point** each.

1. Using the standard normal distribution tables, what is the area under the standard
normal curve corresponding to $-1.2 < Z < 0.5$?
A) 0.5764 B) 0.3085 C) 0.2815 D) 0.8849

Use the following to answer questions 2-5: A friendly STOR455 instructor wants to find out if there is a difference between the final exam scores and the midterm exam scores. He takes a random sample of five students and finds the following scores (assume normal distribution).

Student	1	2	3	4	5
Midterm	25	22	24	27	25
Final	20	24	18	29	29

2. The target population of numbers is
A) All midterm scores in the class B) All final scores in the class
C) All students in the class **D) Bivariate midterm and final scores of all the students in the class** E) None of the above

3. What is a 90% confidence interval for μ, the mean difference between final and midterm scores?
A) (-3.95, 2.75) B) (-6.26, 5.06) **C) (-4.95, 3.75)** D) (-4.60, 3.40)
E) Not within ± 0.05 of any of the above.

4. Based on the confidence interval previously calculated, we wish to test H_0: μ = 0 versus H_a: μ ≠ 0 at the 5% significance level. Determine which of the following statements is true:
A) We cannot make a decision since the confidence level we used to calculate the confidence interval is 90%, and we would need a 95% confidence interval.
B) We accept H_0, since the value 0 falls in the 90% confidence interval and would therefore also fall in the wider 95% confidence interval.
C) We reject H_0, since the value 0 falls in the 90% confidence interval.
D) We cannot make a decision since the confidence interval is too wide.
E) None of the above

5. It is suspected that the material on the final is harder than on the midterm. Therefore you would like to prove that the final exam scores are lower than the midterm scores. The appropriate hypothesis for μ, the mean difference between second and first exam scores is:
A) H_0 μ=0 vs. H_a μ ≠0 **B) H_0 μ=0 vs. H_a μ<0** C) H_0 μ=0 vs. H_a μ>0
D) None of the above

6. In statistical tests of hypotheses, we say the data are significant at level α if
A) α = 0.01 **B) the p-value is less than or equal to α** C) α is small
D) the p-value is larger than α E) α has nothing to do with statistical significance.

Use the following to answer questions 7 - 9:
Do heavier cars use more gasoline? To answer this question, a researcher randomly selected 15 cars registered in North Carolina. He collected data about the weight (in hundreds of pounds) and the mileage (mpg) for each car. From a scatterplot made with the data, a linear model seems appropriate.
7. The study population of items in this study are
A) all cars in NC B) mileage C) weight D) both B, C
E) none of the above

8. The response variable in this study is
A) weight **B) mileage** C) both D) neither E) not enough info

9. The equation of the least-squares regression line is

$$\hat{y} = 40.4 - 0.521 \cdot x$$

Which of the following descriptions of the value of the slope is the correct description?

A) The mileage is expected to increase by 0.521 when the weight of a car increases by 1 pound.

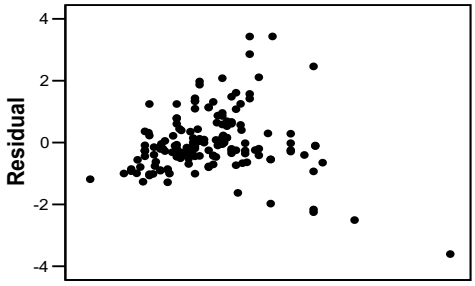
B) The mileage is expected to decrease by 0.521 when the weight of a car increases by 100 pounds.

C) We cannot interpret the slope because we cannot have a negative weight of a car.

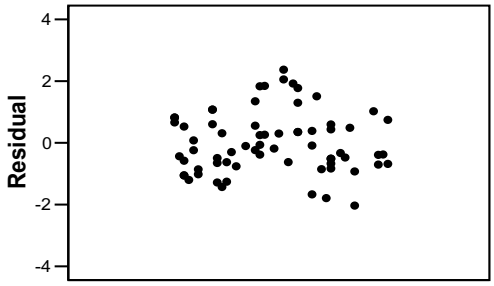
D) None of the above

10. Four different residual plots are shown below. Which plots indicate that the linear regression model is not appropriate?

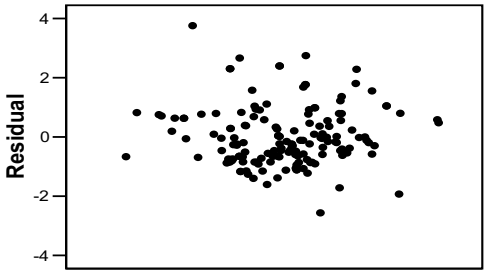
A)



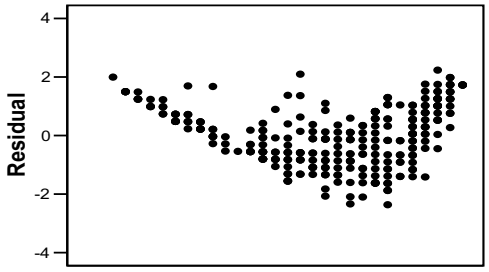
B)



C)



D)



E) Both A and D

11. John's parents recorded his height at various ages between 36 and 66 months. Below is a record of the results:

Age (months)	36	48	54	60	66
Height (inches)	34	38	41	43	45

Which of the following is the equation of the least-squares regression line of John's height on age?

A) Height = 12.(Age) B) Height = Age/12 C) Height = 60 – 0.22.(Age)

D) Height = 20.46 + 0.37.(Age) E) None of the above

12. What do we hope to capture within a confidence interval?
A) The parameter estimate. B) The unknown confidence level.
C) The unknown parameter value. D) The sample size.
E) None of the above.

Use the following to answer questions 13 – 14

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 & 3 \\ 10 & -3 & 3 & 5 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 3 & 2 & 1 \\ 5 & 3 & 1 \\ -1 & -1 & -1 \\ -2 & -5 & -1 \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} 3 & 2 & 1 \\ 5 & 3 & 1 \\ -1 & -1 & -1 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} 3 & 2 & 1 \\ 5 & 3 & 1 \end{pmatrix}$$

13. Which of the following matrices can be multiplied together?
A) **C.D** B) **A.D** **C) B.A** D) **A.B**
E) None of the above

14. Compute **D.C**

$$A) \begin{pmatrix} 14 & 22 \\ 22 & 35 \\ -6 & -9 \end{pmatrix} \quad B) \begin{pmatrix} 10 & -6 & -1 \\ 2 & -17 & -1 \end{pmatrix} \quad C) \begin{pmatrix} 18 & 11 & 4 \\ 29 & 18 & 7 \end{pmatrix} \quad D) \begin{pmatrix} 18 & 11 & 4 \\ 29 & 18 & 7 \\ -7 & -4 & -1 \\ -30 & -18 & -6 \end{pmatrix}$$

E) None of the above C is correct

15. Find the inverse of $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$

$$A) \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad B) \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad C) \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad D) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

E) None of the above B is correct

In questions 16 – 20 indicate what (if anything) will the SAS code produce. (Assume that a data set blah containing variables x and y exists in SAS memory.)

16.
```
proc reg data=blah noprint;
model y=x;
plot y*x;
run;
```

A) nothing, the code is incorrect
B) normal quantile (rankit) plot
C) plot of standardized residuals vs predicted values
D) plot of y vs x with the regression line overlaid
E) none of the above

17.
```
proc reg data=blah noprint;
model y=x;
plot student.*nqq.;
run;
```
A) nothing, the code is incorrect
B) normal quantile (rankit) plot for standardized residuals
C) plot of standardized residuals vs predicted values
D) plot of y vs x with the regression line overlaid
E) none of the above

18.
```
proc reg data=blah noprint;
model y=x;
output out=blah1 student=sr p=yhat;
run;
```
A) nothing, the code is incorrect
B) normal quantile (rankit) plot
C) plot of standardized residuals vs predicted values
D) plot of y vs x with the regression line overlaid
E) none of the above (the code creates a data set blah1 with variables sr and yhat added to the old data set)

19.
```
proc reg data=blah
model y=x;
plot r.*p;
run;
```
A) nothing, the code is incorrect
B) normal quantile (rankit) plot
C) plot of standardized residuals vs predicted values
D) plot of y vs x with the regression line overlaid
E) none of the above

Use the following SAS output to answer questions 20 - 22:

The UNIVARIATE Procedure
Variable: MPG

Basic Confidence Limits Assuming Normality

Parameter	Estimate	95% Confidence Limits	
Mean	25.30200	24.84470	25.75930
Std Deviation	0.63927	0.43971	1.16705
Variance	0.40866	0.19335	1.36201

20. The 95% confidence interval for σ is
A) 25.30 B) [0.19, 1.36] C) [24.84, 25.76] **D) [0.44, 1.17]**
E) None of the above is close

21. The value of the sample mean is
A) 25.30 B) [0.19, 1.36] C) [24.84, 25.76] D) [0.44, 1.17]
E) None of the above is close

22. This output was produced using
A) proc means B) proc reg C) proc print D) proc gplot
E) none of the above (proc univariate)

23 Which of the following functions is linear in unknown parameters (symbols β)?

$A)\beta_0 + \beta_1 x^4 \quad B)\beta_0 + x^{\beta_1} \quad C)\beta_1\sqrt{\beta_1 x_1 + \beta_2 x_2} \quad D)\beta_0 + \sin(\beta_1 x)$ E) none
A is correct

Use the following SAS output to answer questions 24 - 28:

Model: MODEL1
Dependent Variable: Y

Number of Observations Read 84
Number of Observations Used 84

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	93462942	93462942	16.83	<.0001
Error	82	455273165	5552112		
Corrected Total	83	548736108			

Root MSE	2356.29195	R-Square	0.1703
Dependent Mean	7111.20238	Adj R-Sq	0.1602
Coeff Var	33.13493		

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	20518	3277.64269	6.26	<.0001	13997	27038
X	1	-170.57519	41.57433	-4.10	<.0001	-253.27977	-87.87061

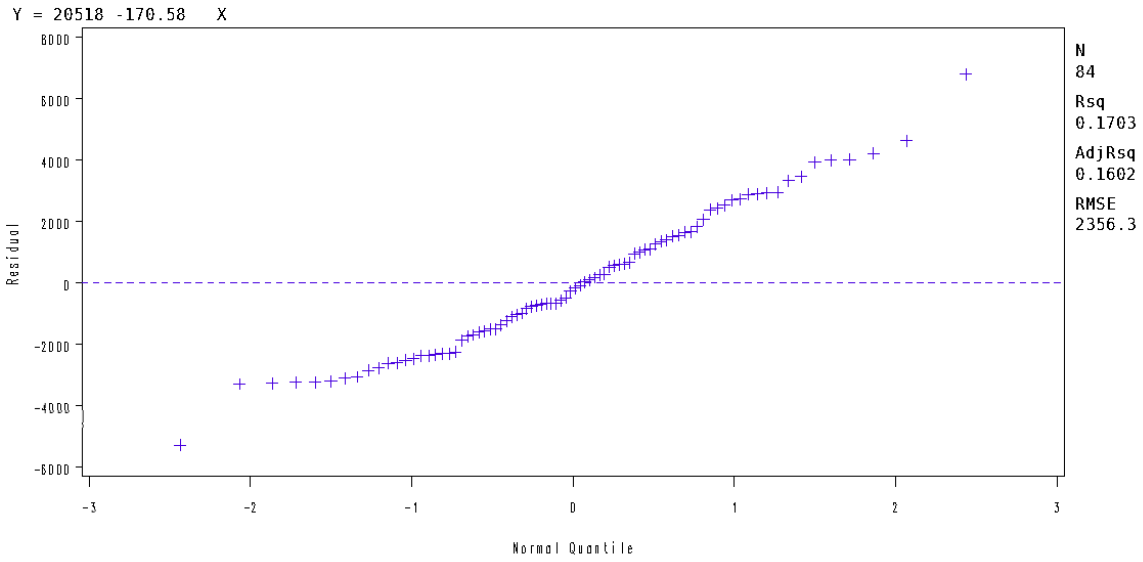
24. The estimate of the standard deviation is
A) 2356.29 B) 20518 C) 170.57519 D) 3277.64
E) none of the above is close

25. The estimate of the intercept
A) 2356.29 **B) 20518** C) -170.57519 D) 3277.64
E) none of the above is close

26. The estimate of the slope
A) 2356.29 B) 20518 **C) -170.57519** D) 3277.64
E) none of the above is close

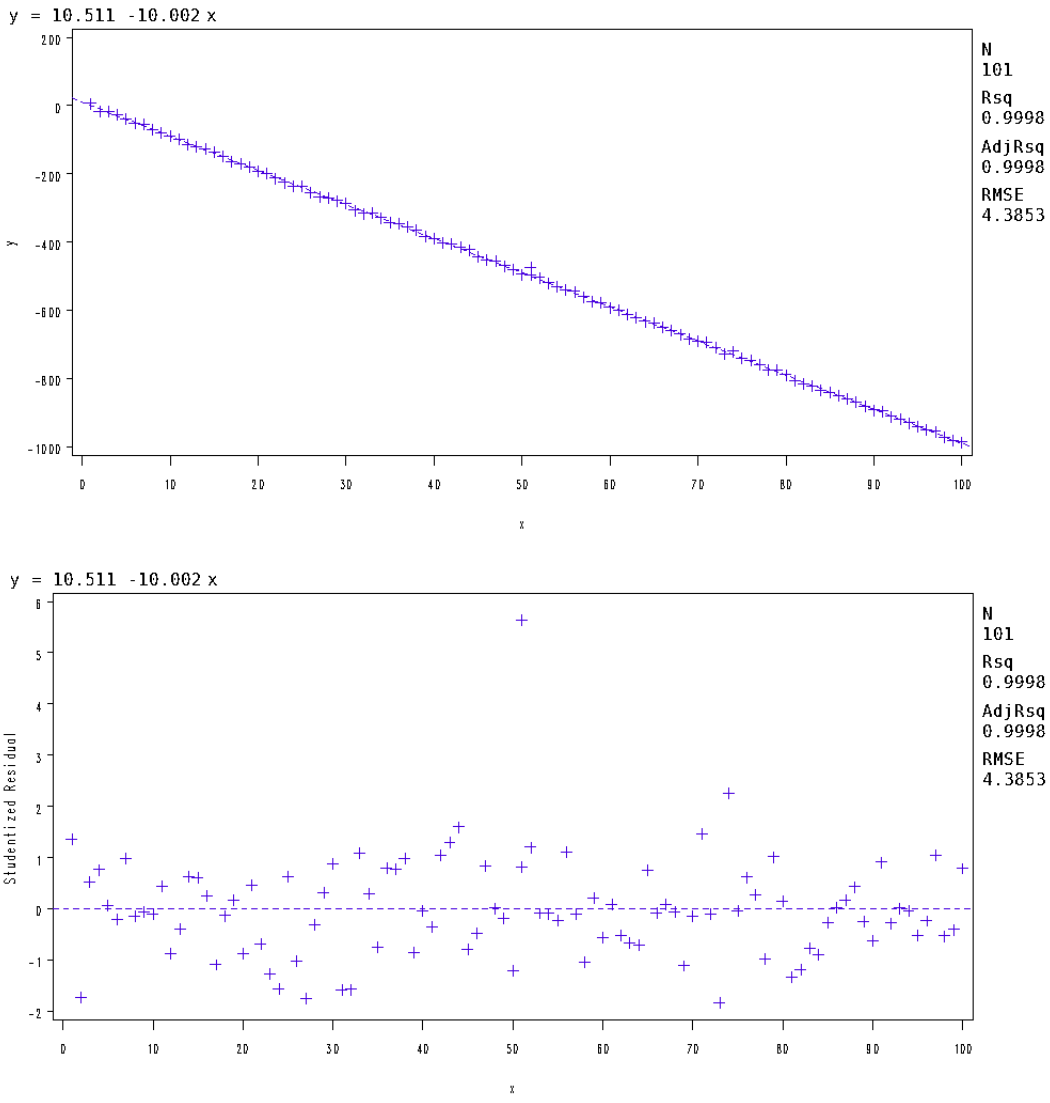
27. Is the slope significantly different from 0?
A) yes B) no C) not enough info D) none of the above

28. Based on the normal quantile plot below is the assumption of normal distribution of the errors appropriate?



A) yes because the plot follows roughly a line
B) yes, since Rsq is big enough
C) no, because Rsq<.5
D) no because there are outliers
E) none of the above

29. What is wrong with the following regression plots?



A) The variance is increasing B) The regression line does not fit
C) There is an outlier D) nothing, the plot indicates a good fit
E) None of the above

30. When computing a CI for standard deviation which table will I use?
A) Normal Table B) T table **C) Chi square table**
D)None of the above