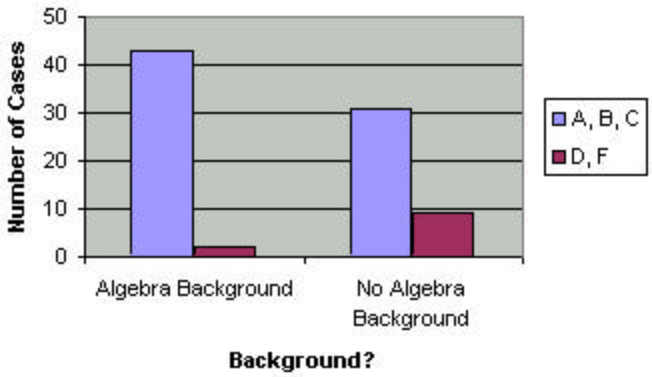
Stat 2300 International, Sample Midterm, Fall 2004

1.Suppose that the correlation coefficient between two variables is zero. This implies that

a. there is no relationship between the two variables.
b. there is no linear relationship between the two variables.
c. there is no curved relationship between the two variables, but there may be a linear relationship.
d. there are probably outliers.
e. no answer or skip this item

2.Suppose that you suspect that poor performance in a first college level mathematics course is related to whether or not a student has taken a complete high school algebra sequence of courses. You have received the following data:

	High School Preparation?		
Grade	Algebra Background	No Algebra Background	Total
A, B, C	43	31	74
D, F	2	9	11
Total	45	40	85



Which of the following statements is not a correct conclusion from the graph and/or table?

a. There were fewer students without an algebra background than students with an algebra background.
b. There were fewer students receiving a grade of "A, B, C," having no algebra background than there were students receiving a grade of "A, B, C," having an algebra background.
c. Students were more likely to receive a grade of "A, B, C," if they had an algebra background than if they did not have an algebra background.
d. The percent of students not having an algebra background that received a grade of "A, B, C," was more than 10% higher than the percent of students having an algebra background that received a grade of "A, B, C,".
e. no answer or skip this item

3. A new automobile model has a surface area of 2.5 square meters. Paint blemishes occur at the rate of 0.1 blemishes per square meter. How many blemishes will the automobile have on average?

a. 0.1
b. 0.25
c. 0.35
d. 25
e. no answer or skip this item

4. The distribution of scores on a placement test is approximated by a normal curve with a mean equal to 500 and a standard deviation equal to 80. Use a normal curve calculator of your choice to answer this question. What proportion of scores is in the interval from 380 to 620?

a. 0.9332
b. 0.0668
c. 0.8664
d. 0.5000
e. no answer or skip this item

5. Which of the following situations are most likely to be exponential: Situation I: The distribution of hand spans of students in a statistics class; Situation II: Land areas in square miles of the 77 U.S. cities with populations over 200,000 in 1992?

a. Hand spans of students; you measure the hand spans at a constant rate.
b. Hand spans of students; the hand spans are independent of each other.
c. Land areas in square miles; the land areas are independent of each other.
d. Neither hand spans of students, nor land areas in square miles
e. no answer or skip this item

6. A study was conducted to compare the final exam scores of students of three different instructors. The same final exam was given to all students enrolled in classes taught by the three instructors. The summary statistics for twenty-five randomly selected students are given in the following table. An analysis of variance was performed on the data. The F-statistic for testing whether the mean final exam scores are equal for all three instructors is 2.00.

Instructor	Number of students	Mean final exam score	Variance
A	10	79.70	46.46
B	7	86.29	40.24
C	8	79.37	85.13

What are the degrees of freedom associated with this analysis of variance?

a. 2 and 21
b. 2 and 22
c. 2 and 23
d. 2 and 24
e. no answer or skip this item

7. The recommended dietary allowance of folic acid for adult females is 400 mcg. Folic acid is found naturally in leafy dark green vegetables, legumes (dried beans and peas), citrus fruits and juices, and most berries. A vitamin supplement is supposed to contain $m = 400$ mcg of folic acid. A random sample of 100 such vitamin tablets was obtained and the amount of folic acid contained in each tablet was determined. The sample mean was 399.92 mcg, and the sample standard deviation was s = 0.5 mcg. Any deviation from the null is of interest. Use an Interactive Tool of your choice. What is the standard error of the sample mean?

a. 399.92
b. 0.5
c. 0.05
d. 39.92
e. no answer or skip this item

8. A university financial aid officer is interested in performing a hypothesis test to determine if the average amount of financial aid his university's students receive is greater than $5000. Which standard deviation would give the highest power for the hypothesis test?

a. A standard deviation of $100
b. A standard deviation of $200
c. A standard deviation of $300
d. A standard deviation of $400
e. no answer or skip this item

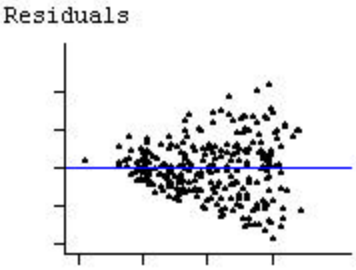
9. A student is interested in determining if a gasoline additive improves the performance of a car in terms of miles/gallon. She selects 16 cars at random. Eight of the cars are provided with the gasoline additive and eight are provided with no additive. She then determines the miles per gallon after each car has driven 500 miles. Is this a paired data design?

a. This is a paired data design, because two different groups of cars are paired and matched.
b. This is not a paired data design, because two different groups of cars, unpaired or unmatched in any way, receive the treatments - a gasoline additive or no additive.
c. This is not a paired data design, because two different groups of cars, paired and matched, receive the treatments - a gasoline additive or no additive.

d. This is a paired data design, because two different groups of cars are unpaired or unmatched.
e. no answer or skip this item

10. For 64 students, the least squares line for X = Socialhours (# of hours per week spent socializing) versus Y = GPA is GPA = 3.33 - 0.01×(Socialhours). On average, if one student spends one more hour socializing than another, how much lower would the student's GPA be with the greater number of hours spent socializing?
a. 0.01
b. 0.02
c. 0.03
d. 3.32
e. no answer or skip this item

11. Here is a residual plot: What does this plot indicate about the standard deviations of the residuals?



a. This residual plot indicates an equality of standard deviations of the residuals. The residuals fan out as X gets larger, indicating an increasing standard deviation of the residuals.
b. This residual plot indicates an equality of standard deviations of the residuals. The residuals fan out as X gets larger, indicating a decreasing standard deviation of the residuals.
c. This residual plot indicates an inequality of standard deviations of the residuals. The residuals fan out as X gets larger, indicating an increasing standard deviation of the residuals.
d. This residual plot indicates an inequality of standard deviations of the residuals. The residuals fan out as X gets smaller, indicating a decreasing standard deviation of the residuals.
e. no answer or skip this item

12. Data for the 50 states plus Washington, D. C. includes the following three variables: Income: median household income in thousands of dollars. TeacherSalary: the average salary of primary and secondary school teachers, in thousands of dollars, Poverty: the percentage of people in poverty. The table below contains least squares estimates for some linear models:

Model	Estimates
Income = a + b×TeacherSalary + E	a: 14.05, b: +0.5227
Income = a + b×Poverty + E	a: 42.78, b: -0.7380
Income = a + b×TeacherSalary + c×Poverty + E	a: 25.84, b: +0.4214, c: -0.6218

If the poverty rate is held fixed, what is the expected change in median household income associated with an increase of $1000 in the average salary of primary and secondary school teachers?

a. $522.7
b. $738.0
c. $421.4
d. $621.8
e. Cannot determine from the information given
f. no answer or skip this item

13. Why do randomized block designs produce more accurate estimates of treatment differences than a completely randomized design?

a. The comparability of treatment groups is controlled and results in improved precision.
b. Two or more measurements are collected from each participant in the experiment.
c. More than one experimental unit is observed under the same treatment conditions.
d. Experimental units are randomized to treatments.
e. no answer or skip this item

14. Students created a paper airplane and attached a weight to it to observe the effect on the distance the plane would travel. The four categories of weights were: 0 paper clips, 1 paper clip, 2 paper clips and 3 paper clips. Each plane was thrown 6 times and the distance traveled was recorded. The following ANOVA table was produced:

Analysis of Variance Results					
Source	df	SS	MS	F-ratio	P
Weight	3	183.74			
Error	20	79.13			
Total	23	262.87			

How will you compute Error MS?

a. 79.13 times 20
b. 79.13 divided by 20
c. 183.74 minus 79.13
d. 262.87 minus 79.13
e. no answer or skip this item

15. Students in a technology class performed a study examining the pollution level of automobiles. Three categories of exhaust filters were: standard, new1 and new2; each of the three filters was randomly placed on each of twelve vehicles. The pollution levels of the automobiles in each of the three categories were recorded. The following ANOVA table was produced (use an Interactive Tool of your choice):

Analysis of Variance Results					
Source	df	SS	MS	F-ratio	P
Factor (Exhaust)	2	3501	1751	2.19	0.1278
Error	3	26373	799		
Total	35	29874			

What percentage of variation in pollution level is explained by Factor (Exhaust)?

a. 13.3%
b. 11.7%
c. 86.7%
d. 33.3%
e. no answer or skip this item

16. In most applications of one-way ANOVA, a key statistical question of interest is:

a. Only this question: "Are the differences in the t population means $m_1, m_2, \ldots, m_t$ supported by the data?"
b. Only this question: "Could the observed differences in t sample means have occurred just by chance?"
c. Both of the options presented
d. Neither of the options presented
e. no answer or skip this item

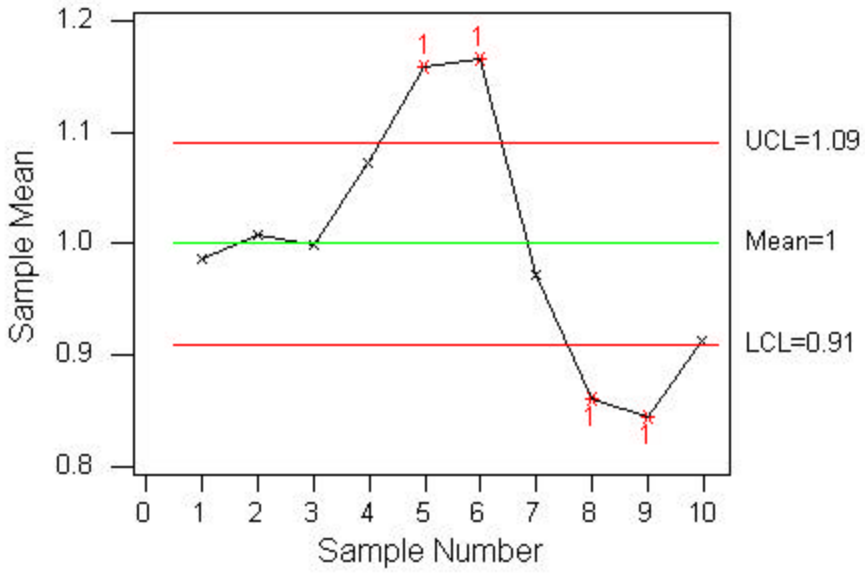
17. When more than two population means are compared, the probability of finding at least one significant pairwise mean difference when in fact none exists is _____.

a. the pairwise comparison rate
b. the experimentwise error rate
c. the comparisonwise error rate
d. $(1-\alpha)^3$
e. no answer or skip this item

18. Data collected over time from processes that behaves as if they had been obtained from more than one population:

a. can be used to predict the process behavior in the near future
b. indicate the presence of special causes of variation
c. indicates an in-control processes
d. all of the options presented
e. none of the options presented
f. no answer or skip this item

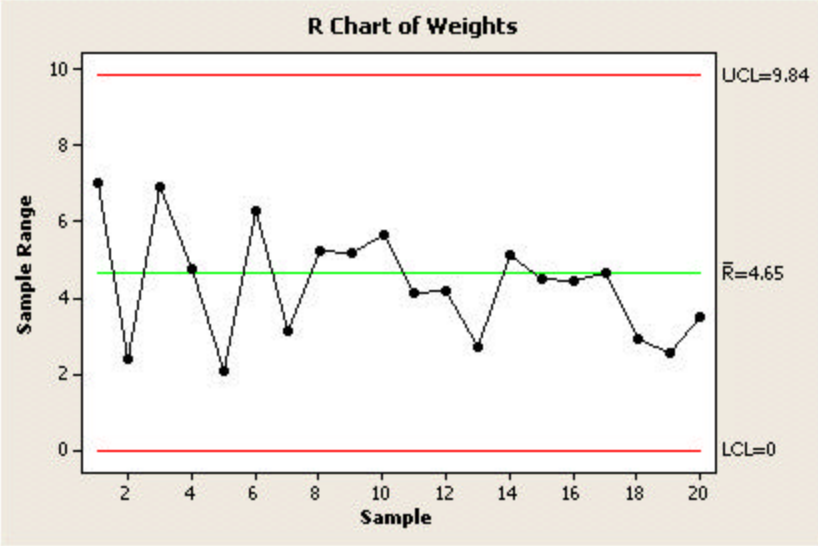
19. A control chart for the sample means obtained in a study of the thickness of silicon oxide on wafers used in integrated circuits follows. Ten samples of four wafers each were obtained, the thickness of the silicon oxide measured, the sample means calculated, and the accompanying chart prepared. Measurements are in microns.



Is the centering of the thickness of the silicon oxide in control?

a. We have reason to believe that the process is not in control; we can use the center line as a valid estimate of process centering.
b. We have reason to believe that the process is in control; no runs, trends, or sawtooth patterns are present.
c. The process mean is in control, we can not use the center line as a valid estimate of process centering.
d. The process mean is not in control, we can not use the center line as a valid estimate of process centering.
e. no answer or skip this item

20. A control chart for the sample ranges obtained in a study of the weights of metal wires follows. Twenty samples of five wires each were obtained, the weights measured, the sample means calculated, and the accompanying chart prepared. Measurements are in grams. What is the estimate of the standard deviation of the weights of the metal wires?



a. approximately 1.73 grams
b. approximately 2.60 grams
c. approximately 3.87 grams
d. approximately 8.65 grams
e. no answer or skip this item

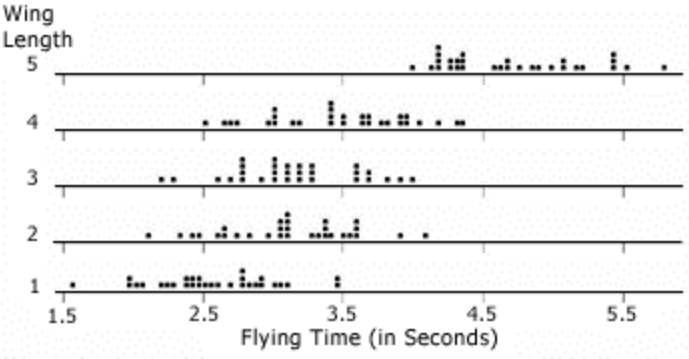
21. Since the seasonal component is a measure of the discrepancy between the time series from its mean value, the average of the seasonal components should be (i) _____ . In the classical decomposition model, the trend and cyclic components are merged into one piece.

Without resorting to additional assumptions on the nature of the time series, it is virtually impossible to distinguish between these two components.

Cycles of unknown duration can often be analyzed using (ii) _____ .

Fill in the missing two words.

22. Suppose the flying times were recorded for a total of 150 flights, using 5 different wing-lengths as shown in the above dotplot.



Find the degrees of freedom for (i) wing-length, (ii) Error, and (iii) Total.

23. The number of unsatisfied customers complaining to the store manager each day might plausibly be modeled using the Poisson distribution.

If the average number of unsatisfied customers per day is 5.5, what is the chance that no customer complains during a day? Use the Poisson calculator.

24. The distribution of vehicle speeds at a certain highway location is approximated by a normal curve with a mean of 61 miles per hour and a standard deviation of 6 miles per hour. Use the empirical rule to find the values that would complete the following statement.

About 99.7% of the vehicle speeds at this location are between ____ and ____.

25. Boxes of a certain breakfast cereal are supposed to contain 12 ounces of cereal. The machine that fills the boxes puts on average m ounces in the boxes. A random sample of 100 such boxes was obtained and the weights of the cereal inside each was found. The mean sample was 11.78 ounces, and the sample standard deviation was s = .5 ounces.

(i) Estimate the standard error of the sample mean and (ii) find the z-statistic.

26. Use WebStat. Load from "Data > Sample Data" the data set "Textbook_spending_4_majors.dat". This is one out of 5 questions that will work with this data set. Conduct a one way ANOVA on the first 3 columns only (i.e., civil-eng, chem-eng, and elect-eng), assuming the data represent textbook spendings from samples of student within these three different majors.

Report the mean textbook spendings for these three majors.

27. Use the one way ANOVA from question 26.

State the null and alternative hypotheses that relate to this one way ANOVA.

28. Use the one way ANOVA from question 26.

In the ANOVA table, how do we calculate MS Treatments? First indicate a formula using terms such as SS Error, df Treatment, etc. and then fill in the numerical values.

29. Use the one way ANOVA from question 26.

Report the *p*-value and conclude whether we reject (or do not reject) the null hypothesis and draw a conclusion.

30. Use the one way ANOVA from question 26.

Compare your result in question 29 with the result from CyberStats Unit E-3, Uses 2, that compares the three majors from above and comp-eng. What is different here - and what is a possible conclusion?

31. Use WebStat. Load from "Data > Sample Data" the data set "Air_pressure_boiling_point.dat". This is one out of 5 questions that will work with this data set. Conduct a multiple linear regression analysis with AirPressure and log(AirPressure) as two explanatory variables and BoilingPoint as the response variable.

Report and interpret the *p*-value from the ANOVA table for the multiple regression model.

32. Use the multiple regression model from question 31.

Report and interpret the p-values associated with AirPressure and log(AirPressure).

33. Use the multiple regression model from question 31.

Aren't your answers to questions 31 and 32 contradictory - one saying that both parameters must be included and the other saying that none of these parameters should be included in the model? Use an appropriate scatterplot to explain what is happening here.

34. Use WebStat. Load from "Data > Sample Data" the data set "Air_pressure_boiling_point.dat". This is one out of 5 questions that will work with this data set.

Conduct two simple linear regression analyses: (i) one with AirPressure as the explanatory variable and BoilingPoint as the response variable and (ii) one with log(AirPressure) as the explanatory variable and BoilingPoint as the response variable. Report your two models.

35. Use the two simple regression models from question 34.

Which of the two models in question 34 should we ultimately use to predict BoilingPoint, given that we know both AirPressure and log(AirPressure)?