18. Which of the following are valid statistical hypotheses? Choose all that are valid.

    (a) $H_0: \pi > 0.05$ versus $H_a: \pi < 0.05$

    (b) $H_0: \mu_1 - \mu_2 = 10$ versus $H_a: \mu_1 - \mu_2 > 10$

    (c) $H_0: S_x = 1$ versus $H_a: S_x \neq 1$

    (d) $H_0: \pi_1 - \pi_2 = 0$ versus $H_a: \pi_1 - \pi_2 < 0$

Short answer and computational questions

19. An investigator desires to estimate π to within 0.05 with confidence 0.95. The investigator knows that the value of π is no larger than 0.25. What sample size should the investigator use? (7 pts)

20. Consider a population of $N = 500$ numerical values. An investigator is interested in studying the sampling distribution of the sample median based on samples of size $n = 11$. Describe how the investigator could obtain the sampling distribution assuming that she has sufficient resources. (5 pts)

21. A consumer product testing group is interested in the lifetimes of name brand televisions and discount brand televisions. Independent random samples of each type of television were selected resulting in the summary table below.

Type	Sample Size	Sample Mean lifetime	Sample SD
Name Brand	12	7.2 years	1.2
Discount	8	5.9 years	1.5

    The consumer group wants to know if mean lifetimes differ between name brand and discount televisions.

    (a) State the appropriate null and alternative hypotheses. (4 pts)

    (b) Is the pooled two sample t test appropriate? Why or why not? (1 pts)

    (c) Compute the test statistic and the p-value. (4 pts)

    (d) Make a decision and draw conclusions. Use $\alpha = 0.05$. (4 pts)

11.5 Final Exam: Fall 2000

True/False: 3 points each

1. The value of r (correlation coefficient) depends on which variable is labeled x and which is labeled y.

2. The value of r is always between 0 and 1.

3. If $r = 0$, then you can conclude that there is no relationship between x and y.

4. The coefficient of determination, r^2, measures the proportion of variability in y that can be explained by a linear relationship between x and y.

5. In simple linear regression, s_e is a point estimate of the standard deviation of y when x has the value x^*.

6. In the simple linear regression model, b is an unbiased estimator of β.

7. The standard deviation of the sampling distribution of $a + bx^*$ increases as the difference between x^* and $\bar{x}$ increases.

8. In a simple linear regression, the mean of y when x has the value x^* is $\alpha + \beta x^* + \varepsilon$.

9. In a simple linear regression model, the residual for the i^{th} observation is $y_i - (\alpha + \beta x_i)$.

10. Residuals in simple linear regression do not have the same standard deviation because the standard deviation of a residual depends on x.

11. If a simple random sample of size n is obtained and each member of the sample is classified into one of k categories, then the X^2 statistic for testing goodness of fit has $n - 1$ degrees of freedom.

12. A small X^2 value reveals that the observed counts are close to the counts that would be expected if H_0 were true.

13. If two categorical variables are independent, then

$$P(\text{level 1 of the first variable and level 1 of the second variable}) =$$

$$P(\text{level 1 of the first variable}) + P(\text{level 1 of the second variable}).$$

14. The objective of the one-way analysis of variance is to test the hypothesis of equality of all population variances.

15. When using the Tukey-Kramer multiple comparison procedure, if the confidence interval for $\mu_1 - \mu_2$ contains 0, then μ_1 and μ_2 <u>are</u> declared significantly different.

16. The F test for testing equality of population means can safely be used if the <u>difference</u> between the largest and the smallest of the sample standard deviations is no more than 2.

17. If $H_0 \colon \mu_1 = \mu_2 = \cdots = \mu_k$ is rejected, then it can be concluded that $\mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$.

Multiple choice: 4 points each

18. If X is a continuous variable, then $P(X = 10)$

    (a) is the area under the density curve where $x < 10$
    (b) is the area under the density curve where $x > 10$
    (c) is 0
    (d) depends on the specific distribution of X

19. A good fitting regression line should have

    (a) small r^2 and large s_e.
    (b) small r^2 and small s_e.
    (c) large r^2 and large s_e.
    (d) large r^2 and small s_e.

20. If the slope of the regression line is negative and the coefficient of determination is 0.64, then the Pearson's correlation is

    (a) 0.64.
    (b) -0.64.
    (c) 0.80
    (d) -0.80

21. A point is an influential observation in simple linear regression if

    (a) it has a large residual.
    (b) it has a small residual.
    (c) it plays a big role in determining the slope of the least squares line.
    (d) the x value is equal to $\bar{x}$.

22. Which of the following is <u>not</u> an assumption made about the random deviation ε in a simple linear regression model?

    (a) The distribution of ε is normal.
    (b) The mean of ε is $\alpha + \beta x$
    (c) The standard deviation of ε is constant for all x values.
    (d) $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent of one another.

23. In a simple linear regression, the predicted value of y, when x has value x^*, is

    (a) $\alpha + \beta x^* + \varepsilon$.
    (b) $\alpha + \beta x^*$.
    (c) $a + bx^* + e$.
    (d) $a + bx^*$.

24. A 95% confidence interval for the mean of y when x has the value x^*

    (a) is wider than a 95% prediction interval for the value y when x has the value x^*.
    (b) is narrower than a 95% prediction interval for the value y when x has the value x^*.
    (c) is exactly the same as a 95% prediction interval for the value y when x has the value x^*.
    (d) may be wider or narrower than a 95% prediction interval for the value y when x has the value x^*.

25. Two variables, x and y, having a bivariate normal distribution are independent if

    (a) $y = \alpha + \beta x$
    (b) $\rho = 0$
    (c) $r = 0$
    (d) $\alpha = 0$

26. The use of the X^2 test statistic is appropriate if

    (a) Sample size is at least 30.
    (b) All observed counts are at least 5
    (c) All expected counts are at least 5.
    (d) One or more expected count is at least 5

27. If $\mathrm{H}_0\colon \mu_1 = \mu_2 = \cdots = \mu_k$ is true, then

    (a) Only *MSE* estimates σ^2.
    (b) Only *MS Treat* estimates σ^2.
    (c) *MSE* estimates σ^2 and *MS Treat* estimates σ^2.
    (d) Only *MS Treat* / *MSE* estimates σ^2

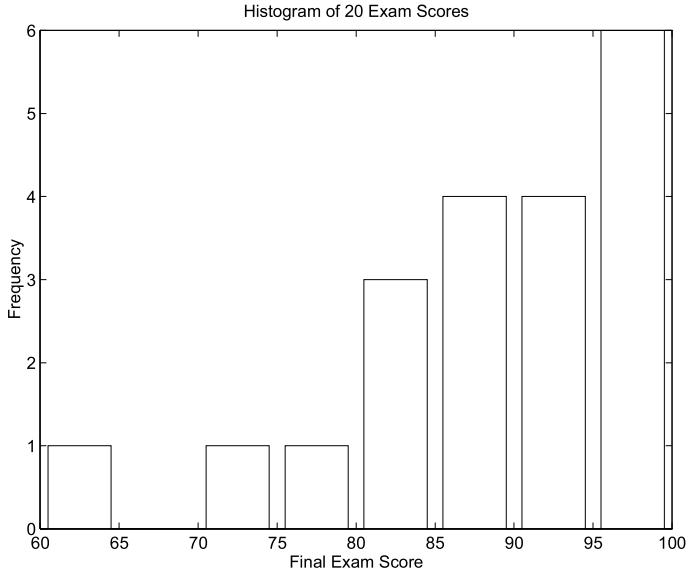
28. The assumption that observations are independent is required for

    (a) one-way ANOVA.
    (b) simple linear regression.
    (c) X^2 test of independence.
    (d) all of the above.

29. The following data set consists of the price of a GE toaster at different stores. What is the 5-number summary for this data set? 16 17 18 18.5 19 21 22.5 28

    (a) 16.0 17.5 18.75 21.75 28.0
    (b) 17.0 17.5 20.00 21.75 22.5
    (c) 16.0 17.5 20.00 21.75 28.0
    (d) 16.0 17.0 20.00 22.50 28.0
    (e) 16.0 17.0 18.75 22.50 28.0

30. The distribution of final exam scores for 20 statistics students is displayed below.



For this distribution,

(a) The median is approximately equal to 70.

(b) The median is approximately equal to 80.

(c) The median is approximately equal to 90.

(d) The median is approximately equal to 100.

(e) The location of the median cannot be determined from the histogram.

31. An investigator obtained a simple random sample of adults from each of three Western states. Each adult was classified into one of three categories: (1) in favor of opening the Alaska National Wildlife Refuge (ANWR) to oil exploration, (2) opposed to opening the ANWR to oil exploration, or (3) no opinion about opening the ANWR to oil exploration. The data were analyzed using the X^2 test statistic.

(a) This is a test of independence.

(b) This is a test of lack of fit.

(c) This is a test of homogeneity.

(d) This is an F test of equality of means.

32. To determine whether variable x causes changes in variable y, it is best to

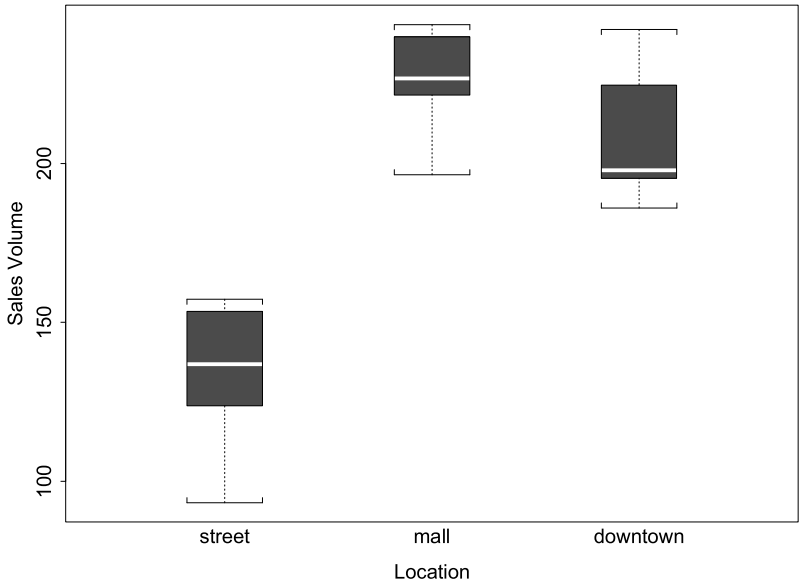
(a) randomly sample from the population of (x, y) values.

(b) randomly assign subjects to different levels of the x variable.

(c) randomly assign subjects to different levels of the y variable.

(d) randomly sample from each of the populations corresponding to different x values.

33. Stereo World is interested in studying the distribution of sales volume at each of three locations. The box-plot below summarizes the results. Which location has the least variability as measured by the interquartile range?



(a) The Street location
(b) The Mall location
(c) The Downtown location
(d) The variability is the same for all three
(e) The variability cannot be determined from the plot.

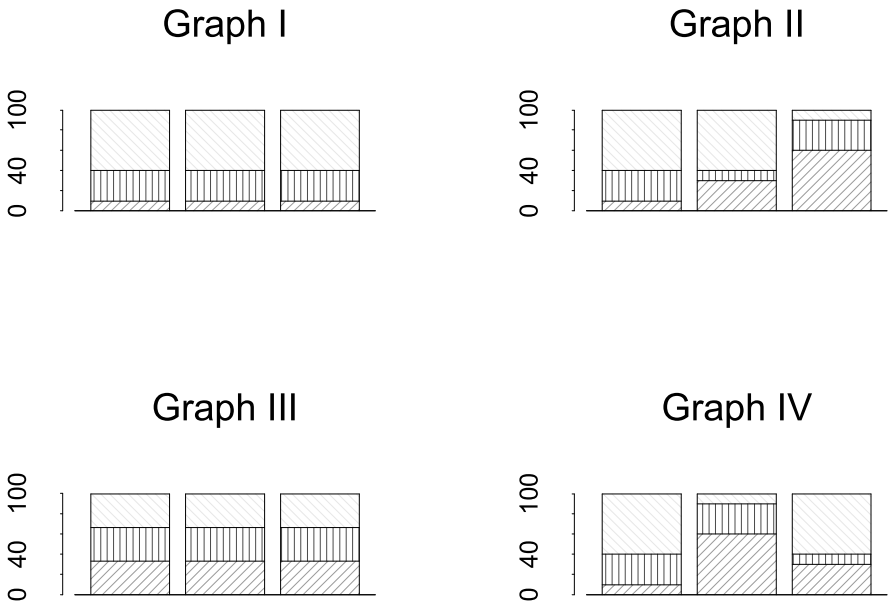
34. For which location is the median sales volume the largest?

(a) The Street location
(b) The Mall location
(c) The Downtown location
(d) The median is the same for all three
(e) The median cannot be determined from the plot.

35. The plot that is most useful for displaying the relationship between a numerical variable and a categorical variable is a

(a) normal plot
(b) histogram
(c) set of side-by-side box-plots
(d) segmented bar graph

36. Examine the segmented bar graphs below. For which graphs are the two variables independent?.



(a) I only
(b) III only
(c) II and IV only
(d) I and III only
(e) I, II, III, and IV

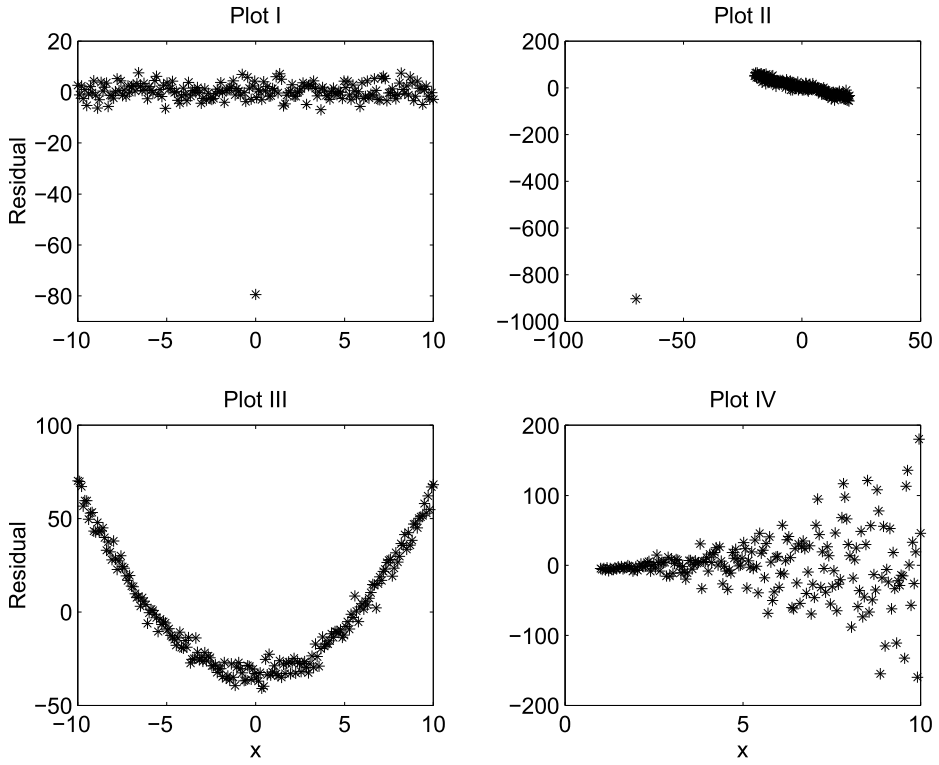
37. The plot that is most useful for displaying the relationship between two numerical variables is a

(a) scatterplot
(b) histogram
(c) set of side-by-side box-plots
(d) segmented bar graph

38. The plot that is most useful for displaying the relationship between two categorical variables is a

(a) normal plot
(b) histogram
(c) set of side-by-side box-plots
(d) segmented bar graph

39. Examine the four residual plots below. For which plot is the mean residual equal to zero?



(a) Plot I only

(b) Plot IV only

(c) Plot I and Plot IV

(d) Plots I and III only.

(e) The mean residual is equal to zero in <u>all</u> of the plots

40. Which of the residual plots in question 39 displays an outlier with large influence on the slope of the least squares line?

(a) Plot I only

(b) Plot II only

(c) Plots I and II

(d) All plots display an outlier with large influence.

41. Which of the residual plots in question 39 displays non-constant variance?

(a) Plot III only

(b) Plot IV only

(c) Plots III and IV

(d) All plots show non-constant variance

Short answer and computational questions

42. Compute the sample mean and sample variance of the following data set: 7, 11. (5 pts)

43. A construction engineer has designed an experiment to investigate three different types of pressure treatment that can be applied to raw lumber. Nine pieces of lumber were randomly assigned to each pressure treatment. After treatment, each piece of lumber was tested for water repellency. Low scores on the water repellency measure indicate better repellency than do high scores. Selected statistics are reported below.

Treatment	Sample Size	Sample Mean
I	9	51
II	9	60
III	9	48

(a) Complete the ANOVA table (2 pts for each entry)

(b) Give a point estimate for σ (2 pts)

(c) Give the critical value of the F statistic and test the hypothesis $H_0: \mu_1 = \mu_2 = \mu_3$ at $\alpha = 0.05$. (3 pts)

(d) Use Tukey-Kramer 95% confidence intervals to determine which, if any, means are significantly different. Show all work. (7 pts)

(e) What proportion of the variability among the repellency measures is explained by their relationship to the type of pressure treatment? (3 pts)

Source	Sum of Squares	*df*	Mean Square	*F*
Pressure Treatments	_____	_____	_____	_____
Error	2,112	_____	_____	
Total	_____	_____		

44. A simple linear regression model was fit to data concerned with eruptions of Old Faithful. The response was the duration of the eruption (in minutes) and the explanatory variable was the waiting time since the last eruption (in minutes). The sample mean of waiting time is $\bar{x} = 70.9$ minutes and the sample variance of waiting time is $s_x^2 = 184.96$. Rweb output appears below:

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.29917 -0.37689  0.03508  0.34909  1.19329

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.874016   0.160143  -11.70   <2e-16
waiting      0.075628   0.002219   34.09   <2e-16

Residual standard error: 0.4965 on 270 degrees of freedom
Multiple R-Squared: 0.8115,     Adjusted R-squared: 0.8108
F-statistic:  1162 on 1 and 270 degrees of freedom,     p-value:     0
```

(a) Give point estimates for α, β, and σ. (4 pts)

(b) Estimate the mean eruption duration when waiting time is one hour. (3 pts)

(c) Predict the value of y when waiting time is one hour. (3 pts)

(d) Compute a 95% confidence interval for the mean eruption duration when waiting time is one hour. Show all work. (7 pts)

11.6 Final Exam: Spring 2001

1. Mr. Iago Bald, a specialist in meditation and hypnosis claims that 75% of the men he has treated for hair loss have fully recovered their receding hair lines. We think this claim is exaggerated, so to disprove the claim using an appropriate hypothesis test (and an experiment) we would state our null and alternative hypotheses as

   A. $H_0: \pi = .75$ versus $H_a: \pi > .75$

   B. $H_0: p = .75$ versus $H_a: p < .75$

   C. $H_0: \pi = .75$ versus $H_a: \pi < .75$

   D. None of the above.

2. A valid hypothesis test is conducted at the 5% significance level and the p-value is calculated to be 0.044. The correct decision is

   A. Sufficient evidence exists to accept H_0.

   B. Sufficient evidence exists to reject H_0.

   C. Insufficient evidence exists to reject H_0.

   D. Fail to reject H_0.

3. The p-value of a test statistic is defined as

   A. The probability that the null hypothesis is true.

   B. The sample proportion.

   C. The probability that the test statistic would take on a value as extreme or more extreme than the value actually observed assuming that H_0 is true.

   D. The probability proportion, assuming we know it.

   E. The probability that the test statistic would take on a value as extreme or more extreme than the value actually observed assuming that H_a is true.

4. The surgeon general is worried that more teenagers are starting to smoke. Suppose that samples (SRS) were taken in 1980 and in 1998 from people in the 17–20 year old age group and a 95% confidence interval for the difference in proportion of smokers, $\pi_1 - \pi_2$ is $(-0.093, 0.087)$. Pick the best interpretation of this confidence interval.

   A. There is a probability of 0.95 that $\pi_1 - \pi_2$ falls in the interval $(-0.093, 0.087)$.

   B. If we repeat the process again and again, we will get the same confidence interval 95% of the time.

   C. If we repeat the process again and again, in the long run, 95% of our confidence intervals will contain $p_1 - p_2$.

   D. If we repeat the process again and again, in the long run, approximately 95% of our confidence intervals will contain the true mean difference, $\pi_1 - \pi_2$.

5. Which of the following is **not true** about the difference between sample proportions from independent samples? (If all are true choose D.)

   A. The mean of the sampling distribution is $\pi_1 - \pi_2$.

   B. $p_1 - p_2$ is a statistic.

   C. The variance of the sampling distribution is $\sigma^2_{p_1} + \sigma^2_{p_2}$.

   D. All of the above are true.

6. Alcohol and nicotine consumption during pregnancy may harm an unborn child. One study classified 452 mothers according to their alcohol intake prior to pregnancy recognition and their nicotine intake during pregnancy. The two-way table is

	Nicotine (mg/day)		
Alcohol (oz/day)	None	1-15	16 or more
None	105	7	11
.01-.10	58	5	13
.11-.99	84	37	42
1.00 or more	57	16	17

An appropriate null hypothesis for the above table can be stated as

A. Pregnancy and alcohol consumption are associated.

B. Pregnancy and alcohol consumption are independent.

C. Alcohol and nicotine consumption of (future) mothers are associated.

D. Alcohol and nicotine consumption of (future) mothers are independent.

7. General Motors claims that its model 6.5L diesel produces less NO_2 emissions than does the 5.9L Cummins diesel engine that Dodge puts in their trucks. For a random sample of $n_1 = 17$ GM vehicles, a mean level of 2.5 gm/liter was observed with a standard deviation of 4.2 gm/liter. Similarly, a random sample of $n_2 = 18$ Dodge vehicles was taken and an observed mean and standard deviation of 3.5 gm/liter and 3.2 gm/liter respectively were computed. To test that the General Motors claim is correct, what type of test should be used? Assume that NO_2 emission measures are approximately normal.

A. An independent sample t procedure, with H_a being two sided.

B. An independent sample t procedure, with H_a being one sided.

C. A dependent sample t procedure, with H_a being two sided.

D. A dependent sample t procedure, with H_a being one sided.

8. A 3×4 two-way table containing a total of n observations, was used to determine whether the *row variable* **A** and the *column variable* **B** are independent. Suppose that the computed test statistic is $X^2 = 10.25$. We can conclude at the $\alpha = .05$ level that

A. There is a significant association between A and B.

B. There is not a significant association between A and B.

C. The χ^2 test is not appropriate for this problem.

D. None of the above.

9. A scientist is interested in testing the hypotheses $H_0 : \mu = 105$ versus $H_a : \mu \neq 105$, using $\alpha = .01$ level. He computed a 99% C.I. for μ to be $(104.13, 105.91)$. Which of the following is a correct conclusion for the above problem?

A. The true mean value is different from 105.

B. The true mean value is not different from 105.

C. Reject H_0.

D. There is not much confidence in our method.

10. Suppose that a simple linear regression model was fit to data and that $H_0: \beta = 0$ was rejected in favor of $H_a: \beta > 0$. If all of the assumptions of the simple linear regression model are satisfied, then the investigator can conclude that

    A. there is a causal relationship between x and y

    B. there is no association between x and y

    C. there is a nonlinear association between x and y

    D. as x increases, $\mu_{y|x}$ also increases

11. The equation for the simple linear regression model is $y_i = \alpha + \beta x_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma_\varepsilon)$. The parameters that need to be estimated are

    A. $\alpha, \beta,$ and σ_ε

    B. α, β and ε_i

    C. a and b

    D. $a, b,$ and s_e

12. Which of the following is the correct interpretation for β in the simple linear regression model?

    A. β is the estimated mean response when the value of the explanatory variable is 0.

    B. β is the mean response when the value of the explanatory variable is 0.

    C. β is the estimated change in mean response per unit increase in the explanatory variable.

    D. β is the change in mean response per unit increase in the explanatory variable.

13. Suppose that a population of (x, y) pairs follows the simple linear regression model. Then the mean of y when $x = x^*$ is

    A. β

    B. σ_ε

    C. $a + bx^*$

    D. $\alpha + \beta x^*$

    E. Both C and D

14. Which of the following statements is not true?

    A. Correlation and simple linear regression can be used to investigate the relationship between 2 categorical variables.

    B. Correlation and simple linear regression are not resistant to outliers.

    C. The presence of confounding variables can make results misleading.

    D. A strong correlation does not imply causation.

    E. Linear relationships are measured by r and b.

15. When testing $H_0: \beta = 0$ against $H_a: \beta \neq 0$ in simple linear regression, the SAS computer output gave the p-value as 0.015. It is appropriate for the researcher to

    A. conclude that $H_0: \beta = 0$ is true with probability 0.015.

    B. reject $H_0: \beta = 0$ at $\alpha = 0.05$.

    C. conclude that $H_a: \beta \neq 0$ is true with probability 0.015.

    D. reject $H_a: \beta \neq 0$ at $\alpha = 0.015$.

    E. Both A and B.

16. Which of the following is not true in a one-way ANOVA table?

    A. $df_{\text{tot}} = df_{\text{treat}} + df_{\text{error}}$
    B. $SS_{\text{tot}} = SS_{\text{treat}} + SS_{\text{error}}$
    C. $MS_{\text{tot}} = MS_{\text{treat}} + MS_{\text{error}}$
    D. $F = MS_{\text{treat}} \div MS_{\text{error}}$

17. Which of the following statements about outliers and influential points is FALSE?

    A. An outlier has a large vertical deviation from the fitted regression line.
    B. An influential point is an observation that has a marked effect on the fitted regression line when it is removed from the data set.
    C. It is possible for an observation to be both an outlier and an influential point.
    D. If an observation is influential it must be an outlier.

18. Select the correct hypotheses for conducting an ANOVA test where there are $k = 3$ treatments.

    A. $\text{H}_0\colon \mu_1 = \mu_2 = \mu_3$ versus $\text{H}_a\colon \mu_1 \neq \mu_2 \neq \mu_3$
    B. $\text{H}_0\colon \mu_1 = \mu_2 = \mu_3$ versus $\text{H}_a\colon \mu_i \neq \mu_j$ for at least one pair of treatment means.
    C. $\text{H}_0\colon \mu_1 = \mu_2 = \mu_3$ versus $\text{H}_a\colon$ the treatment means are not all equal.
    D. $\text{H}_0\colon \mu_1 = \mu_2 = \mu_3 = 0$ versus $\text{H}_a\colon$ the treatment means are not all equal to 0.
    E. Both **B** and **C** are correct and equivalent statements.

19. Which of the following statements is not true as it pertains to one-way ANOVA?

    A. The ε_{ij}s are assumed to be a SRS from the $\text{N}(0, \sigma_\varepsilon)$ distribution.
    B. The population standard deviations are assumed to be all equal.
    C. The ANOVA model is $x_{ij} = \mu_i + \varepsilon_{ij}$.
    D. If H_a is true, then MS_{treat} and MS_{error} both are estimates of σ_ε^2.
    E. $\overline{X}_i$ is an unbiased estimator of μ_i.

20. Below is a table of summary statistics for a designed experiment:

Treatment	n_i	$\bar{x}_i$	s_i^2
One	4	6	16
Two	5	10	36
Three	6	18	49

    Is it reasonable to make the assumption of equal population variances when analyzing these data via ANOVA? Choose the best answer.

    A. Yes, because $\frac{7}{4} = 1.75$ and $1.75 < 2$.
    B. No, because $\frac{7}{4} = 1.75$ and $1.75 < 2$.
    C. No, because $\frac{18}{6} = 3$ and $3 > 2$.
    D. No, because $\frac{49}{16} = 3.0625$ and $3.0625 > 2$.
    E. No, by inspection of the table, the population standard deviations are not the same.

21. A set of Tukey 95% confidence intervals is shown.

```
Tukey's Studentized Range (HSD) Test for variable: WEIGHT

                   Simultaneous            Simultaneous
                      Lower    Difference     Upper
      BRAND        Confidence    Between   Confidence
Comparison           Limit       Means       Limit
1     - 2          -0.6397      0.5779      1.7956
1     - 3          -1.6695     -0.4519      0.7656
2     - 3          -2.2473     -1.0298     -0.1878
```

The set of confidence intervals indicate

A. All the means are different from one another.

B. None of the means are different from one another.

C. μ_1 is not different from μ_2 or μ_3.

D. μ_2 is different from μ_3.

E. Both **C** and **D** are true.

22. If Tukey's method is used to construct a set of 95% confidence intervals, then we can be

A. 95% confident that none of the confidence intervals contains their respective true difference between the means.

B. 95% confident that all of the sample means are equal to the true means.

C. 95% confident that none of the sample means are equal to the true means.

D. 95% confident that at least one of the confidence intervals contains its true difference between the means.

E. 95% confident that all of the confidence intervals contains their respective true differences between the means.

Word Problems and Discussion. Show ALL WORK.

23. Suppose that the regression equation relating systolic blood pressure (y) and age (x) is $y_i = \alpha + \beta x_i + \varepsilon_{ij}$, where $\alpha = 98$, $\beta = .95$, and $\sigma_\varepsilon = 17$. Assume that $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_\varepsilon)$. Compute the probability that the systolic blood pressure of a randomly selected 51 year old is less that 120. (10 pts)

24. Suppose that the mean weight of all books in the library is 5.6 pounds and the standard deviation is 2 pounds. Compute the probability that the mean weight of a random sample of 25 books is greater than 6 pounds. (10 pts)

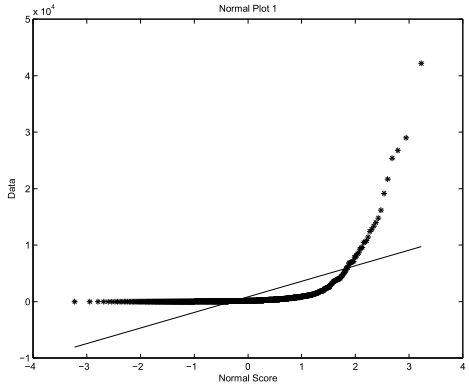
25. Houndstongue (a noxious weed) is found in abundance on private and public lands that have been grazed by cattle. Houndstongue is rarely found on lands that have been grazed by mountain goats. One investigator concluded that houndstongue infestations could be reduced by importing mountain goats to the infested areas.

(a) Define the term "confounding variable." (6 pts)

(b) Describe a confounding variable that was ignored by the investigator in this study. (4 pts)

26. *Occupational Outlook Quarterly* (Spring 1993) reported that less than 2% of all drywall installers employed in the construction industry are women. How many drywall installers should we sample to estimate π to within 0.01 of the true value with 95% confidence. (10 pts)

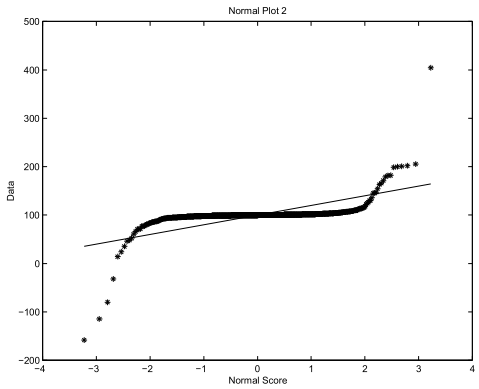
27. Examine this normal plot.



(a) Describe the shape of the distribution (4 pts)

(b) Is there a value of λ (power transformation) that can be used to help normalize the data? If so, is λ less that or greater than 1? (4 pts)

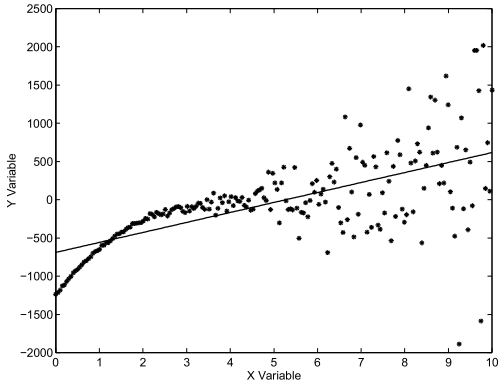
28. Examine this normal plot.



(a) Describe the shape of the distribution (4 pts)

(b) Is there a value of λ (power transformation) that can be used to help normalize the data? If so, is λ less that or greater than 1? (4 pts)

29. Examine the scatterplot below. The least squares regression line is drawn on the plot.



List the linear regression assumptions that appear to be violated or other problems that are revealed by the plot. If no problems are detected, then write none. (8 pts)

30. A researcher hypothesized that babies who are born in summer and fall in areas with cold winters take longer to learn to crawl because they are slowed down by the clothing. A sample of 425 northern babies was obtained. Each baby was classified according to season of birth and early versus late crawling. The data are summarized below.

Season of Birth	Early Crawlers	Late Crawlers	Total
Winter	64	51	115
Spring	43	36	79
Summer	44	53	97
Fall	50	84	134
Total	201	224	

Partial SAS output is given below.

Table of Season by Crawl

Season	Crawl		
Frequency / Expected / Row Pct / Col Pct	Early	Late	Total
Winter	64 54.388 55.65 31.84	51 60.612 44.35 22.77	115
Spring	43 37.362 54.43 21.39	36 41.638 45.57 16.07	79
Summer	44 45.875 45.36 21.89	53 54.64 23.66	97
Fall	50 63.374 37.31 24.88	84 70.626 62.69 37.50	134
Total	201	224	425

Statistics for Table of Season by Crawl

Statistic	DF	Value
Chi-Square		10.3373
Cramer's V		0.1560

(a) State the null hypothesis for this study. (4 pts)

(b) Compute the expected count for Late Crawlers born in the Summer. (5 pts)

(c) Sketch a segmented bar graph. Use Season as the variable on the X axis. (6 pts)

(d) How many degrees of freedom are associated with the test statistic? (2pts)

(e) At $\alpha = .05$ do we reject or fail to reject H_0? Explain how you made your decision. (5 pts)

(f) Based on the segmented bar graph and the test statistic, does it appear that the researchers hypothesis is supported? Explain. (5 pts).