Midterm Exam Review AP Stat

Directions- answer all multiple choice questions and then complete the following four problems from your book. You should have access to all answers but the most important part is your ability to understand, apply and explain statistical concepts. Show as much work as possible and use your calculator as much as possible. This should serve as an adequate review for your exam. **All concepts on the midterm exam are not limited to the review.**

1. page 47 example 1.36 b

Show that 85 and 86 are outliers using **1.5 times IQR** approach and a **z-score** approach (more than 2 standard deviations away from the mean)

2. page 92 example 2.25

Use **z-scores** (show work when using the formula) and verify using **TableA** on your calculator.

3. page 188 example 4.1 a, f, g

nonlinear data.

4. page 161 example 3.45 a, b ,c only

Explain why the **residual** at age **51 is -.5**.

Additional Study Topics

- LSRL
  - With a calculator
  - With your formula sheet
- LSRL
  - "Is it a good fit?"
- Nonlinear Models
- Design an Experiment
- Design a Simulation
- Disjoint / Mutually Exclusive
- Law of Large Numbers
- Distributions for μ_{X+Y}, μ_{X-Y}, σ^2_{X+Y}, σ^2_{X-Y}
- Empirical Rule
- Expected Value E(X)
- Probability with Venn Diagrams
- Binomial Questions

midterm exam review

Multiple Choice
Identify the letter of the choice that best completes the statement or answers the question.

____ 1. When a histogram has a longer tail to the right, it is said to be
a. symmetrical
b. skewed to the left
c. skewed to the right
d. none of these alternatives is correct

Exhibit 2-2
A survey of 800 college seniors resulted in the following crosstabulation regarding their undergraduate major and whether or not they plan to go to graduate school.

	Undergraduate Major			
Graduate School	**Business**	**Engineering**	**Others**	**Total**
Yes	70	84	126	280
No	182	208	130	520
Total	252	292	256	800

____ 2. Refer to Exhibit 2-2. What percentage of the students does not plan to go to graduate school?
a. 280
b. 520
c. 65
d. 32

____ 3. Refer to Exhibit 2-2. What percentage of the students' undergraduate major is engineering?
a. 292
b. 520
c. 65
d. 36.5

____ 4. Refer to Exhibit 2-2. Of those students who are majoring in business, what percentage plans to go to graduate school?
a. 27.78
b. 8.75
c. 70
d. 72.22

____ 5. Refer to Exhibit 2-2. Among the students who plan to go to graduate school, what percentage indicated "Other" majors?
a. 15.75
b. 45
c. 54
d. 35

____ 6. When data are skewed right, the mean will usually be
a. greater than the median
b. smaller than the median
c. equal to the median
d. positive

____ 7. The measure of location which is the most likely to be influenced by extreme values in the data set is the
   a. range
   b. median
   c. mode
   d. mean

____ 8. In a five number summary, which of the following is **not** used for data summarization?
   a. the smallest value
   b. the largest value
   c. the mean
   d. the 25th percentile

____ 9. For the standard normal probability distribution, the area to the left of the mean is
   a. -0.5
   b. 0.5
   c. any value between 0 to 1
   d. 1

____ 10. Which of the following is **not** a characteristic of the normal probability distribution?
   a. symmetry
   b. The total area under the curve is always equal to 1.
   c. 99.72% of the time the random variable assumes a value within plus or minus 1 standard deviation of its mean
   d. The mean is equal to the median, which is also equal to the mode.

Exhibit 6-2
The weight of football players is normally distributed with a mean of 200 pounds and a standard deviation of 25 pounds.

____ 11. Refer to Exhibit 6-2. The probability of a player weighing more than 241.25 pounds is
   a. 0.4505
   b. 0.0495
   c. 0.9505
   d. 0.9010

____ 12. Refer to Exhibit 6-2. The probability of a player weighing less than 250 pounds is
   a. 0.4772
   b. 0.9772
   c. 0.0528
   d. 0.5000

____ 13. Refer to Exhibit 6-2. What percent of players weigh between 180 and 220 pounds?
   a. 28.81%
   b. 0.5762%
   c. 0.281%
   d. 57.62%

____ 14. Refer to Exhibit 6-2. What is the minimum weight of the middle 95% of the players?
   a. 196
   b. 151
   c. 249
   d. 190

____ 15. X is a normally distributed random variable with a mean of 8 and a standard deviation of 4. The probability that X is between 1.48 and 15.56 is
a. 0.0222
b. 0.4190
c. 0.5222
d. 0.9190

____ 16. Given that Z is a standard normal random variable, what is the value of Z if the are to the left of Z is 0.0559?
a. – 0.4441
b. – 1.59
c. 0.0000
d. – 1.50

Problem

17. The hourly wages of a **sample** of eight individuals is given below.

Individual	Hourly Wage (dollars)
A	27
B	25
C	20
D	10
E	12
F	14
G	17
H	19

For the above **sample**, determine the following measures:
a. The mean.
b. The standard deviation.
c. The 25th percentile.

18. A professor at a local university noted that the grades of her students were normally distributed with a mean of 73 and a standard deviation of 11. **DO NOT ROUND YOUR NUMBERS.**

a. The professor has informed us that 7.93 percent of her students received grades of A. What is the minimum score needed to receive a grade of A?
b. Students who made 57.93 or lower on the exam failed the course. What percent of students failed the course?
c. If 69.5 percent of the students received grades of C or better, what is the minimum score of those who received C's?

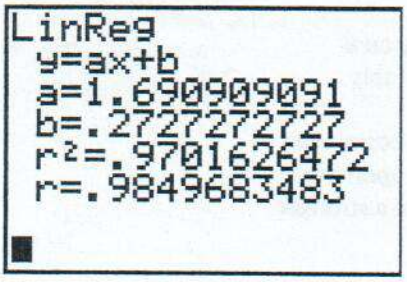
Statistics Midyear Concepts

1. Effect of extreme values, skewness, and/or outliers on the mean and median in relation to resistant and nonresistant measures. (mistake "does not change")
2. A "mathematical process" to define an outlier.
3. Compare and contrast a histogram to a bar graph in relation to qualitative (categorical) and quantitative data.
4. Explain the r^2 value (coefficient of determination) in the context of a problem. (not to be confuse with the correlation coefficient (r value). See the back
5. Define $\hat{y}$ and x if possible in a LSRL
6. Show your work when calculating a residual. $y - \hat{y}$
7. Explain what is meant by an interpolation and extrapolation.
8. Identify the two criteria for a linear relationship. ("good" r value **and** "good" residual plot)
9. Calculate an experimental and theoretical probability in relation to "flipping a coin".
10. Explain the concept of the "Law of Large Numbers" in the context of picking a heart. (mistake "cannot happen"

Correlation Coefficient

How well does your regression equation truly represent your set of data?

One of the ways to determine the answer to this question is to exam the *correlation coefficient* and the *coefficient of determination*.



The correlation coefficient, *r*, and the coefficient of determination, r^2, will appear on the screen that shows the regression equation information (be sure the Diagnostics are turned on — 2nd Catalog (above 0), arrow down to DiagnosticOn, press ENTER twice.)

In addition to appearing with the regression information, the values *r* and r^2 can be found under VARS, #5 Statistics → EQ #7 *r* and #8 r^2

Correlation Coefficient, r:

- The quantity r, called the *linear correlation coefficient*, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the *Pearson product moment correlation coefficient* in honor of its developer Karl Pearson.
- The mathematical formula for computing r is:

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

where n is the number of pairs of data.

(Aren't you glad you have a graphing calculator that computes this formula?)

- The value of r is such that $-1 \le r \le +1$. The + and – signs are used for positive linear correlations and negative linear correlations, respectively.
- *Positive correlation:* If x and y have a strong positive linear correlation, r is close to +1. An r value of exactly +1 indicates a perfect positive fit. Positive values indicate a relationship between x and y variables such that as values for x increases, values for y also increase.
- *Negative correlation:* If x and y have a strong negative linear correlation, r is close to -1. An r value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship between x and y such that as values for x increase, values for y decrease.
- *No correlation:* If there is no linear correlation or a weak linear correlation, r is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables
- Note that r is a dimensionless quantity; that is, it does not depend on the units employed.
- A *perfect* correlation of $\pm$ 1 occurs only when the data points all lie exactly on a straight line. If $r = +1$, the slope of this line is positive. If $r = -1$, the slope of this line is negative.
- A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*. These values can vary based upon the "type" of data being examined. A study utilizing scientific data may require a stronger correlation than a study using social science data.

Coefficient of Determination, r^2 or R^2:

- The *coefficient of determination*, r^2, is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph.
- The *coefficient of determination* is the ratio of the explained variation to the total variation.
- The *coefficient of determination* is such that $0 \le r^2 \le 1$, and denotes the strength of the linear association between x and y.
- The *coefficient of determination* represents the percent of the data that is the closest to the line of best fit. For example, if $r = 0.922$, then $r^2 = 0.850$, which means that 85% of the total variation in y can be explained by the linear relationship between x and y (as described by the regression equation). The other 15% of the total variation in y remains unexplained.
- The *coefficient of determination* is a measure of how well the regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation. The further the line is away from the points, the less it is able to explain.