AP Statistics Review for Midterm Exam

Name _____

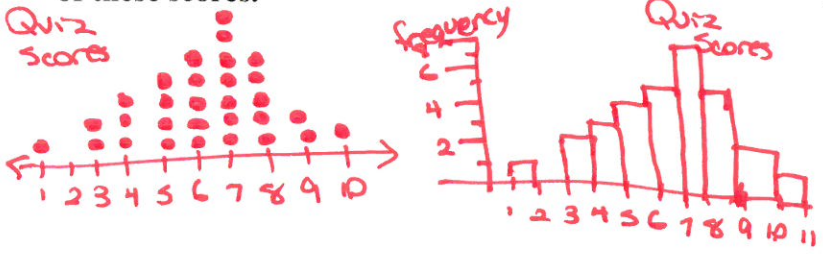
This Midterm Exam Review covers chapters 1 to 17.

Part I. Free Response.

1. The following were scores earned by students in a class of 30 on a quiz.

1	4	5	5	6	7	7	7	8	9	3	4	5	6	6
7	7	8	8	9	3	4	5	6	6	7	7	8	8	10

a. Make a dotplot or a histogram (your choice) of these scores.



b. Find the mean and the standard deviation of these scores. Show use of the formulas.

Since these scores are a sample find s!

$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$

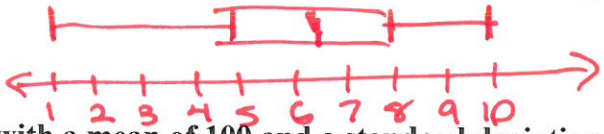
$\bar{x} = \frac{\sum x_i}{n} = 6.2$

$s = \frac{(1-6.2)^2 + (4-6.2)^2 + \ldots + (10-6.2)^2}{30-1}$

$s = 2.0240$

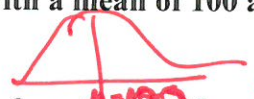
c. Give the 5-number summary of the scores and make a modified boxplot.

min = 1
Q1 = 5
med = 6.5
Q3 = 8
max = 10



8-5 = 3
3(1.5) = 4.5
5-4.5 = .5
8+4.5 = 12.5

2. IQ is distributed normally with a mean of 100 and a standard deviation of 15.5. Make & Label Sketches!



a. What is the percentile rank of someone with an IQ of 112?

$z = \frac{112-100}{15.5} = .7742$

Looking up on the Table

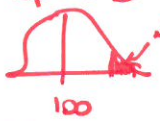
$P(x \le 112) = P(z < .77) = .7794$

or Calc Normcdf(-1E99, 112, 100, 15.5) = .7804

b. What IQ do you need to have to be in the top 5%?

InvNor(.95) = 1.6449 = z

Top 5% = 95%ile



$1.6449 = \frac{x-100}{15.5}$

$125.496 = x$

from table z = 1.64 so 125.42

c. What percent of people have an IQ between 75 and 115?

Table $z = \frac{75-100}{15.5} = -1.6129$

$z = \frac{115-100}{15.5} = .9677$

$P(75 < x < 115) = P(-1.61 < z < .97) =$

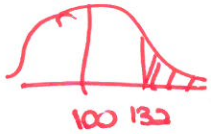
$.8340 - .0537 =$

.7805

Calc Normcdf(75, 115, 100, 15.5) = .7800

d. What percent have an IQ of at least 132?

Table $P(x \ge 132) = P(z > 2.06) = 1 - P(z < 2.06) =$



$z = \frac{132-100}{15.5} = 2.0645$

$1 - .9803 =$

.0197

or Calc Normcdf(132, 1E99, 100, 15.5) = .0195

AP Statistics Review for Midterm Exam

3. The following table shows the SAT scores of football players in the Big Ten vs. all students' SAT scores.

School	Football Players' SAT	All Students' SAT
Illinois	872	1140
Indiana	741	1007
Michigan	826	1190
Michigan State	788	998
Minnesota	838	1050
Northwestern	1034	1250
Ohio State	820	986
Penn State	897	1083
Purdue	881	1009
Wisconsin	825	1090
Note: Iowa scores were not available		

a. Which university's scores would be influential in a scatterplot? Justify your answer.

Northwestern's scores are influential. The x-value is much higher than all the other SAT scores for football players, so the SAT scores for all students is influential. (it's also quite high)

b. Using football players' SAT as your explanatory variable, find the LSRL for these data.

x = football player SAT scores

y = student SAT scores

$\hat{y} = 405.933 + .7913x$

c. What is the correlation for these data? Interpret this number in context.

$r = .6966$

There is a moderately strong linear association positive between football players' SATs and all student SATs

d. Iowa's football players have an average SAT score of 814. What score would you predict for the entire student body?

$\hat{y} = 405.933 + .7913(814)$

$\hat{y} = 1050.0512$

I would predict that Iowa's overall student body SAT scores would be about 1050.

e. Find the residual for Penn State. Show work.

$\hat{y} = 405.933 + .7913(897)$

$\hat{y} = 1115.7291$

$e = y - \hat{y}$

$e = 1083 - 1115.7291$

-32.7291

f. What is the coefficient of determination for these data? Interpret this number in the context of this problem.

$r^2 = .4852$

About 48.52% of the variation on student SAT scores can be explained by the LSRL of overall student SAT scores on football player SAT scores.

AP Statistics Review for Midterm Exam

4. According to *The New York Times* (April 2, 1993, page A1) the average monthly rate for basic television cable service increased as follows:

Year	1986	1987	1988	1989	1990	1991	1992
Rate ($)	11.00	13.20	13.90	15.20	16.80	18.00	20.00

a. Find the equation of the LSRL. Give the correlation and the coefficient of determination. Interpret these in the context of this problem.

x = years since 1980
y = cable rate

$\hat{y} = 2.7464 + 1.4107x$

$r = .9943$ – There is a strong positive linear association between years & rate

$r^2 = .9888$ About 98.88% of the variation of the rate can be explained by the LSRL of rate on year.

b. Interpret the slope in the context of this problem. Predict the monthly rate in 1993 (it was actually $21.00).

The cable rate increases by 1.4107 (about $1.41) per year.

$\hat{y} = 2.7464 + 1.4107(13) = \$21.0855$

c. In what year will the rate reach $50.00? Comment on your predictions.

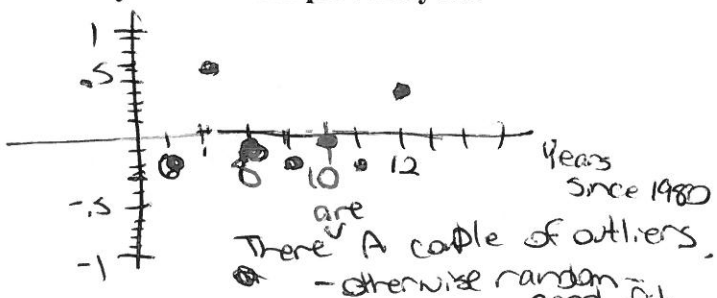
$50 = 2.7464 + 1.4107x$

$41.425 = x$

During the year between 2021 and 2022.

According to it will surpass it by 2022

d. Sketch a residual plot in the space below. What does your residual plot tell you?



There are a couple of outliers. A – otherwise random – good fit

5. Heart weight can be predicted form the length of the cavity of the left ventricle for certain mammals. Consider the following chart:

Mammal	Length of cavity of left ventricle (cm)	Heart Wgt (grams)
Mouse	.55	.13
Rat	1.0	.64
Rabbit	2.2	5.8
Dog	4.0	102
Sheep	6.5	210
Ox	12.0	2030
Horse	16.0	3900

Determine if a linear model is appropriate for this set of data. Use the language of statistics to justify your answer. Don't forget context!

– scatter plot appears curved

x = length of cavity of left ventrical
y = weight of heart

$\hat{y} = -572.6328 + 242.7692x$

$r = .9454$

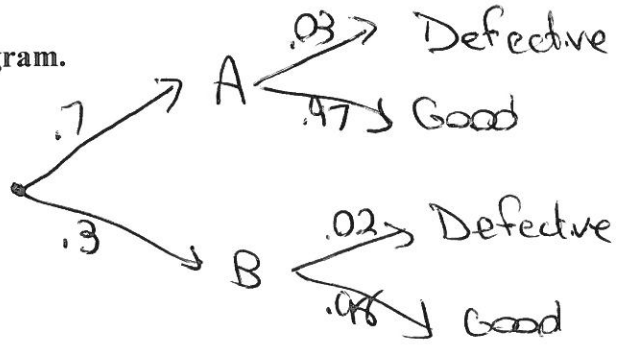
$r^2 = .8937$

No
A linear model is not good. There is a curved pattern in the scatterplot. While the correlation is strong, there is a clear pattern in the residuals plot

– clear pattern in residuals

6. A videocassette recorder (VCR) manufacturer receives 70% of his parts from factory A and the rest from factory B. Suppose that 3% of the output from A are defective and 2% of the output from B are defective.

a. Make a Tree Diagram.



A Defective	.021
A Good	.679
B Defective	.006
B Good	.294

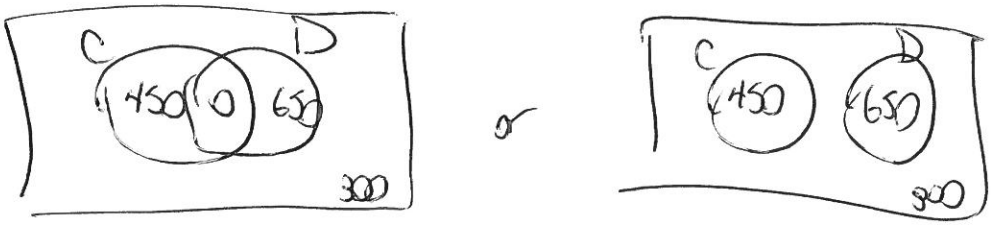
b. What is the probability that a received part is defective?

P(Defective) = .021 + .006 = .027

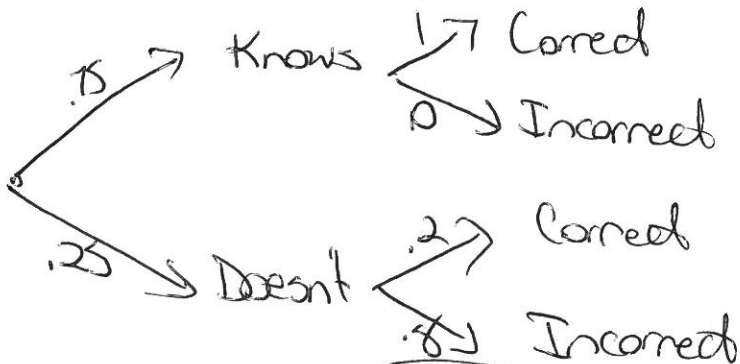
c. If a part is defective, what's the probability that it comes from factory A?

$P(A|D) = \frac{P(A \text{ Defective})}{P(\text{Defective})} = \frac{.021}{.027} \approx .7778$

7. Draw and label a Venn Diagram. Out of 1,400 households surveyed, 450 have cats, 650 have dogs, and 300 have no animal.



8. An exam consists of multiple-choice questions, each having five possible answers. Linda estimates that she has a probability 0.75 of knowing the answer to any question that may be asked. If she does not know the answer, she'll guess, with condition probability 1/5 of being correct. What is the probability that Linda gives the correct answer to a question? Draw a tree diagram to guide the calculation.



Know ✓	.75
Know X	0
Doesn't ✓	.05
Doesn't X	.20

P(Correct) = .75 + .05 = .8

AP Statistics Review for Midterm Exam

9. A certain school gives students grades of 6, 5, 4, 3, 2, 1. The following is a probability distribution for the random variable X that assigns a grade to each student.

Grade (X)	1	2	3	4	5	6
P (X)	.54	.23	.16	.05	.01	.01

a. In the space below, show the calculation of the mean, variance, and standard deviation of X.

$\mu_x = 1(.54) + 2(.23) + 3(.16) + 4(.05) + 5(.01) + 6(.01) =$ 1.79

$\sigma_x^2 = (1-1.79)^2(.54) + (2-1.79)^2(.23) + \ldots + (6-1.79)^2(.01)$

$\sigma_x^2 = 1.1059$

$\sigma_x = \sqrt{1.1059} = 1.0516$

b. A second school has probability distribution for random variable Y that assigns a grade to each student. Show the calculation for the mean, variance, and standard deviation of X-Y.

Grade (Y)	1	2	3	4	5	6
P (Y)	.30	.30	.25	.08	.04	.03

$\mu_y = 1(.30) + 2(.30) + 3(.25) + 4(.08) + 5(.04) + 6(.03) = 2.35$

$\sigma_y^2 = (1-2.35)^2(.3) + (2-2.35)^2(.3) + \ldots + (6-2.35)^2(.03) =$ 1.5875

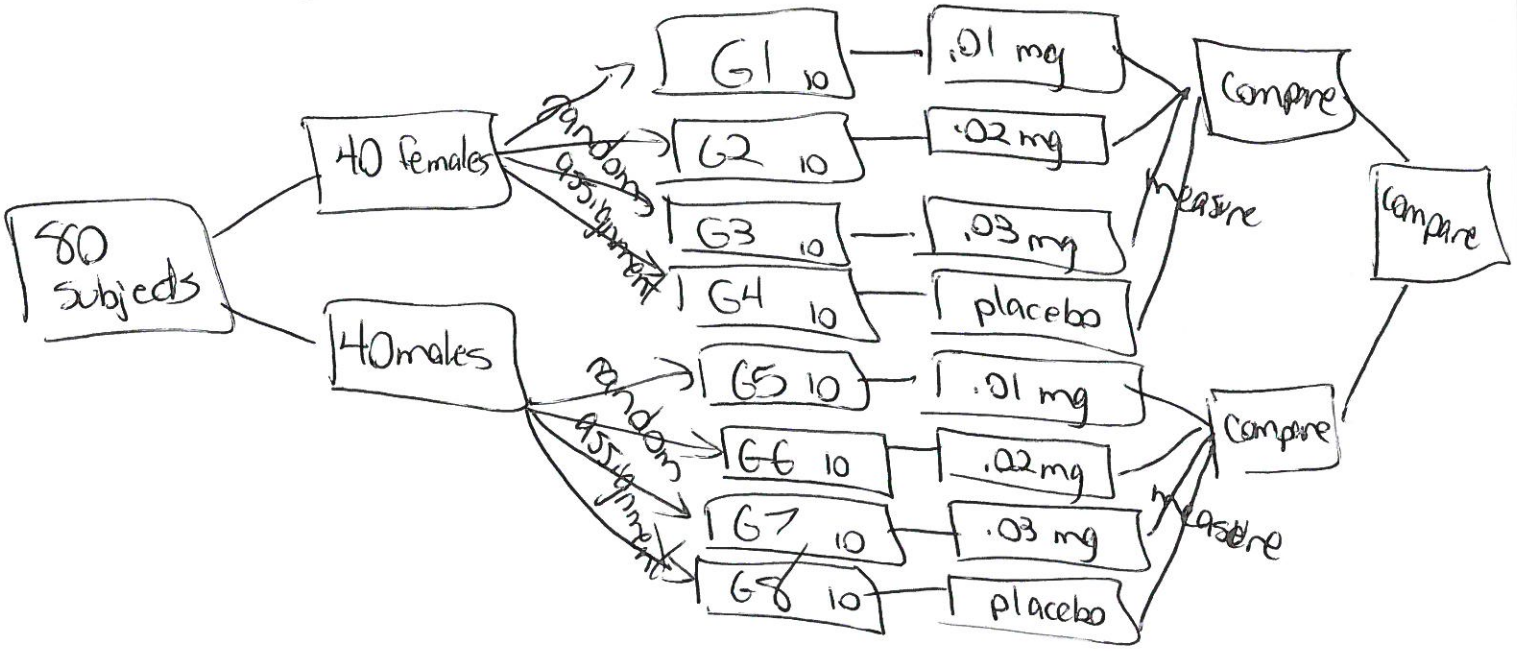
$\sigma_y = \sqrt{1.5875} = 1.26$

$\mu_{x-y} = 1.79 - 2.35 = -.56$

$\sigma_{x-y}^2 = 1.1059 + 1.5875 = 2.6934$

$\sigma_{x-y} = \sqrt{2.6934} = 1.6412$

10. Medical Patients were randomly selected for an experiment using a weight loss pill which contained various levels of magnesium, .01, .02, .and 03. 80 patients were selected, 40 men and 40 women and told to take the pill once a day for 2 weeks. Design an experiment that includes randomization, replication, control and blocking. Write a thorough description of the experiment AND make a sketch. Then perform the randomization to assign treatments to subjects, use random number generator on calculator, seed at 634.



Measure the starting weight of the 80 subjects, Separate the males and females into 2 blocks of 40. for a baseline. Randomly assign the 40 females to 4 different groups by first numbering them from 1 - 40. Then using RandInt(1,40) to select 10 females for group 1 (make sure to ignore repeats). Select 10 more using the same process with the remaining #s for group 2, then group 3. The last 10 females will be group 4. Use the same process for random assignments for males. Group 1 & Group 5 will receive the low (.01 mg) dosage, Groups 2 and 6 will receive the medium dosage (.02mg). Groups 3 and 7 will receive the high dosage. Groups 4 & 8 will ~~receive~~ Serve as our control groups and will receive a placebo. At the end of 2 weeks weigh the men & women again. ~~to~~ ~~compare the results~~ Compare the results. Repeat the experiment to verify results.

G1 25, 26, 36, 20, 15, 34, 31, 12, 27, 32
G2 24, 18, 40, 17, 37, 35, 4, 2, 8, 1
G3 36, 39, 5, 29, 13, 7, 9, 6, 30, 19
G4 3, 10, 11, ~~8~~ 14, 16, 21, 22, 23, 25, 33

AP Statistics Review for Midterm Exam

Part II. Multiple Choice.

B 1. In a 1927 to 1932 Western Electric Company study on the effect of lighting on worker productivity, productivity increased with each increase in lighting but then also increased with every decrease in lighting. If it is assumed that the workers knew a study was in progress, then this is an example of

a. The effect of a treatment unit.
b. The Placebo effect
c. The Control Group effect
d. Sampling Error
e. Voluntary response bias

B 2. In designing an experiment, blocking is used

a. To reduce bias
b. To reduce variation
c. As a substitute for a control group
d. As a first step in randomization
e. To control the level of the experiment

E 3. Consider the following studies being run by three different nursing home establishments.

I. One nursing home has pets brought in for an hour every day to see if patient morale is improved.

II. One nursing home allows hourly visits every day by kindergarten children to see if patient moral is improved.

III. One nursing home administers antidepressants to all patients to see if patient morale is improved.

Which of the following is true?

a. None of these studies uses randomization.
b. None of these studies uses control groups.
c. None of these studies uses blinding
d. Important information can be obtained from all these studies, but none will be able to establish causal relationships.
e. All of the above.

D 4. A consumer product agency tests miles per gallon for a sample of automobiles using each of four different octanes of gasoline. Which of the following is true?

a. There are four explanatory variables and one response variable.
b. There is one explanatory variable with four levels of response.
c. Miles per gallon is the only explanatory variable, but there are four response variables corresponding to the different octanes.
d. There are four levels of a single explanatory variable.
e. Each explanatory level has an associated level of response.

D 5. Which of the following is important in the design of experiments?

I. Control of confounding variables
II. Randomization in assigning subjects to different treatments.
III. Replication of the experiment using sufficient numbers of subjects.

a. I and II
b. I and III
c. II and III
d. I, II, and III
e. None of the above gives the complete set of true responses.

A 6. Which of the following are true about the design of matched-pair experiments?

I. Each subject might receive both treatments.
II. Each pair of subjects receives the identical treatment, and differences in their response are noted.
III. Blocking is one form of matched-pair design.

a. I only b. II only c. III only d. I and III e. II and III

B 7. Which of the following are true statements?

I. In general, strong association implies causation.
II. In well-designed, well-conducted experiments, strong association implies causation.
III. Causation and association are unrelated concepts.

a. I only b. II only c. III only d. I and II e. I, II, and III

D 8. In a study of 379 Marijuana uses, you recorded the user's gender, how long the user has been using, how much marijuana they use each day and the age at which the user first used Marijuana.

D The number of variables you recorded is:

a. 379
b. Five – gender, length of time used, amount of use, age of first use, the number of users
c. Four – length of time used, amount of use, age of first use, the number of users
d. Four – gender, length of time used, amount of use, age of first use
e. Three – length of time used, amount of use, age of first use

D 9. You want to use the empirical rule to test the following data for normalcy:

41 52 45 43 41 54 51

You'd expect 95% of these observations to be between what two values?

a. (36.5 , 56.9) b. (41.2 , 51.8) c. (30.2 , 63.2) d. (35.7 , 57.7) e. (31.4 , 62)

$\bar{x} = 46.7$ $s = 5.5$

E 10. You believe that a Normal Probability Plot of 30 data points provides evidence that the original data is normally distributed. If this is true, then the pattern of points on this Normal Probability Plot is:

a. Bell-Shaped b. u-shaped c. a logarithmic curve
d. randomly scattered in the plot. e. None of the above

You measured the weights of members of population W and found the weights to be normally distributed. The distribution has a mean $\mu = 160$ pounds and $\sigma = 25$.

B 11. For population W, what is the percentile for the weight 185 pounds?

a. 68th percentile b. 84th percentile c. 34th percentile
d. 16th percentile e. None of these

A 12. For population W, what weight corresponds to a z-score of z = -.66?

a. 143.5 lbs b. 176.5 lbs c. 170.3 lbs d. 140.4 lbs e. None of these

C 13. What weight is in the 88th percentile?

a. 179.4 lbs b. 130.6 lbs c. 189.4 lbs d. 210 lbs e. None of these

D 14. What proportion of the population weighs between 143.7 lbs and 191.4 lbs?

a. 52.3% b. 58.8% c. 68.3% d. 63.8% e. None of these

E 15. What proportion of the population weighs less than 135 lbs?

a. 68% b. 34% c. 13.5% d. 63.5% e. None of these