AP Statistics – Semester 1 Final Exam Practice Problems

#1.

In a highly academic suburban school system, 45% of the girls and 40% of the boys take advanced placement classes. There are 2200 girls and 2100 boys enrolled in the high schools of the district. What is the expected number of students who take advanced placement courses in a random sample of 150 students?

a. 128
b. 64
c. 78
d. 90
e. 75

expected number $= \mu$

$\mu_{total} = \mu_{girls} + \mu_{boys}$
$= 990 + 840 = 1830$

Bernoulli settings:

girls: $n = 2200$, $p = .45$, $\mu = np = (2200)(.45) = 990$

boys: $n = 2100$, $p = .40$, $\mu = np = (2100)(.40) = 840$

$p_{total} = \frac{1830}{2200+2100} = .42558$

so if 150 students sample:

$(.42558)(150) = 63.8 \approx \boxed{64}$

#2.

One of the values in a normal distribution is 43 and its z score is 1.65. If the mean of the distribution is 40, what is the standard deviation of the distribution?

a. 3
b. −1.82
c. .55
d. 1.82
e. −.55

$\mu = 40$
$z = 1.65$
$x = 43$

$z = \frac{x-\mu}{\sigma}$

$1.65 = \frac{43-40}{\sigma}$

$\sigma = \frac{43-40}{1.65} = \boxed{1.818}$

#3.

Two plans are being considered for determining resistance to fading of a certain type of paint. Some 1500 homes of 9500 homes in a large city are known to have been painted with the paint in question. The plans are:

Plan A: (i) Random sample 100 homes from all the homes in the city.
(ii) Record the amount of fade over a 2-year period.
(iii) Generate a confidence interval for the average amount of fade for all 1500 homes with the paint in question.

Plan B: (i) Random sample 100 homes from the 1500 homes with the paint in question.
(ii) Record the amount of fade over a 2-year period.
(iii) Generate a confidence interval for the average amount of fade for all 1500 homes with the paint in question.

a. Choose Plan A over Plan B
b. Either plan is good—the confidence intervals will be the same.
c. Neither plan is good—neither addresses the concerns of the study.
d. Choose Plan B over Plan A
e. You can't make a choice—there isn't enough information given to evaluate the two plans.

Plan B is superior because it looks at fading only on houses with the certain type of paint (which is the intent of the study).

#4.

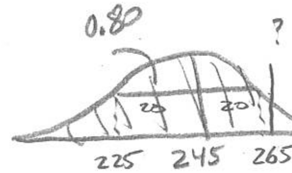
Which of the following is *not* a property of the sample standard deviation (s)?

a. sensitive to the variability of the distribution (yes, it measures this)
b. independent of the mean (yes, spread does not depend on center)
c. resistant to extreme data values no (outliers affect $\bar{x}$, and σ is calculated using $\bar{x}$)
d. independent of the median (yes, spread does not depend on center)
e. all of the above are properties of s

#5.

The weights of professional football players are approximately normally distributed with a mean of 245 lbs. with a standard deviation of 20 lbs. If Thor is at the 80th percentile in weight for football players, which of the following is closest to his weight in pounds?

a. 265
b. 255
c. 252
d. 270
e. 262



invNorm(0.80, 245, 20)
$= \boxed{261.8}$ lbs

#6.

In a famous study from the late 1920s, the Western Electric Company wanted to study the effect of lighting on productivity. They discovered that worker productivity increased with each change of lighting, whether the lighting was increased or decreased. The workers were aware that a study was in progress. What is the most likely cause of this phenomena? (This effect is known as the Hawthorne Effect.)

a. Response bias
b. Absence of a control group
c. Lack of randomization
d. Sampling variability
e. Undercoverage

Response bias:
when something about the study influences the responses.
Because the productivity improvement isn't a result of lighting level it must be due to perception that the lighting changed.
One explanation is workers may have believed that the light change meant they wanted worker output to improve.

#7.

Given $P(A) = .4$, $P(B) = .3$, $P(B|A) = .2$. What are $P(A \text{ and } B)$ and $P(A \text{ or } B)$?

a. $P(A \text{ and } B) = .12$, $P(A \text{ or } B) = .58$
b. $P(A \text{ and } B) = .08$, $P(A \text{ or } B) = .62$
c. $P(A \text{ and } B) = .12$, $P(A \text{ or } B) = .62$
d. $P(A \text{ and } B) = .08$, $P(A \text{ or } B) = .58$
e. $P(A \text{ and } B) = .08$, $P(A \text{ or } B) = .70$

$P(A \text{ and } B) = P(A) \cdot P(B|A)$
$= (.4)(.2) = \boxed{.08}$

$P(A \text{ or } B) = P(A) + P(B) - P(A \cap B)$
$= .4 + .3 - .08$
$= \boxed{.62}$

#8.

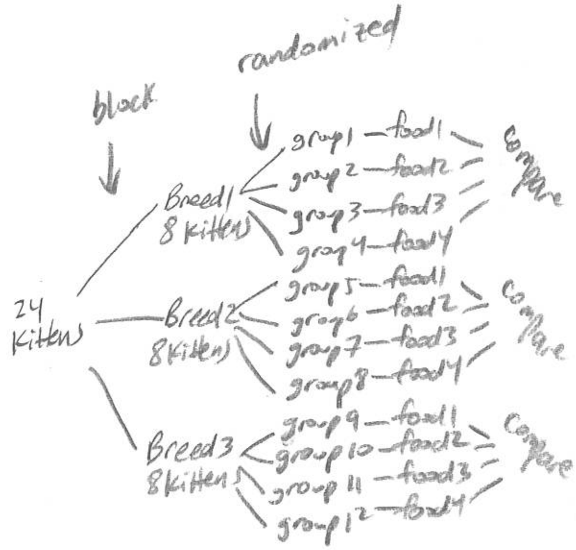
A researcher is interested in establishing a cause-and-effect relationship between exercise level and percentage of body fat. Which of the following should she use?

a. A survey with a stratified random sample
b. An observational study
c. A census
d. A controlled experiment
e. A longitudinal study

only an experiment can establish cause and effect.
For all others, even if an association is seen association does not imply causation.

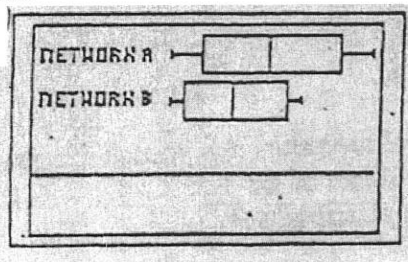
#9.

You are going to conduct an experiment to determine which of four different brands of cat food promotes growth best in kittens ages 4 months to 1 year. You are concerned that the effect might vary by the breed of the cat, so you divide the cats into three different categories by breed. This gives you eight kittens in each category. You randomly assign two of the kittens in each category to one of the four foods. The design of this study is best described as:



a. Randomized block, blocked by breed of cat and type of dog food
b. Randomized block, blocked by type of cat food
c. Matched pairs where each two cats are considered a pair
d. A controlled design in which the various breed of cats are the controls
e. Randomized block, blocked by breed of cat

#10.



The boxplots above compare the television ratings for two competing networks. What conclusion(s) can you draw from the boxplots?

nothing in the data about show

× I. Network A has more shows than Network B
✓ II. Network A has a greater range of ratings than Network B — max−min larger for A
✓ III. Network A is higher rated than Network B — Q1, MED, Q3 and max are all higher for A than B

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

#11.

A least-squares regression line for predicting performance on a college entrance exam based on high school grade point average (GPA) is determined to be *Score* = 273.5 + 91.2 (*GPA*). One student in the study had a high school GPA of 3.0 and a score of 510. What is the residual score for this student?

actual score = 510

predicted score = 273.5 + 91.2(3.0)
= 547.1

residual = actual − predicted
= 510 − 547.1
= $\boxed{-37.1}$

a. 26.2
b. 43.9
c. −37.1
d. −26.2
e. 37.1

#12.

The following table gives the probabilities of various outcomes for a gambling game.

Outcome	Lose \$1	Win \$1	Win \$2
Probability	.6	.25	.15

What is the player's expected return on a bet of \$1?

a. \$.05
b. −\$.60
c. −\$.05
d. −\$.10
e. You can't answer this question because this isn't a complete probability distribution.

$E(x) = \sum xp$

$= (-1)(.6) + (1)(.25) + (2)(.15)$

$= \boxed{-.05}$

or

L1	L2
−1	.6
1	.25
2	.15

1-Var Stats L1, L2

$E = \bar{x}$

(don't forget FreqList: L2)

#13.

Harvey found out that his z score on a college readiness test, compared to others who took the same test was 1.25. Which of the following best describes how you might interpret this value?

a. Harvey's score was 125.
b. Harvey's score was 1.25 standard deviations above the mean of all people taking the test.
c. Only 1.25% of the people taking the test had scores higher than Harvey.
d. Harvey scored 1.25 point above the mean of all people taking the test.
e. Harvey's score was 1.25 times the mean score of all people taking the test.

Z-score = number of standard deviations above (+) or below (−) the mean.

#14.

You want to compare the number of home runs hit in the American League to the number of home runs hit in the National League each year over the past 25 years. Which of the following is likely to be most useful in graphically demonstrating the differences between the two leagues?

a. Parallel boxplots
b. Scatterplot of American League vs. National League
c. Back-to-back stemplots
d. Side-by-side histograms
e. Cumulative frequency plots

back-to-back stemplots ....
(stem & leaf)
(for example)

2	7	1
21	6	012
7110	5	12345
321	4	2478
9442	3	213
2	2	1
3	1	

... preserve the original data values.
these are lost with boxplots, histograms, and cumulative frequency plots.
Scatterplots show individ data but not grouped into categories for comparison.

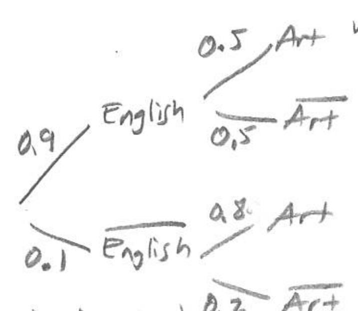
#15.

At a local community college, 90% of students take English. 80% of those who don't take English take art courses, while only 50% of those who do take English take art.

What is the probability that a student takes art?

a. .80
b. .53
c. .50
d. 1.3
e. .45

Tree is best way to visualize this;



$P(Art) = (0.9)(0.5) + (0.1)(0.8)$
$= \boxed{0.53}$

#16.

At a local community college, 90% of students take English. 80% of those who don't take English take art courses, while only 50% of those who do take English take art.

What is the probability that a student who takes art doesn't take English?

a. .08
b. .10
c. .8
d. .85
e. .15

$P(\overline{English} \mid Art) = \dfrac{(0.1)(0.8)}{(0.9)(0.5)+(0.1)(0.8)}$

$= \boxed{0.1509}$

#17.

1. You want to conduct a survey to determine the types of exercise equipment most used by people at your health club. You plan to base your results on a sample of 40 members. Which of the following methods will generate a simple random sample of 40 of the members?

a. Mail out surveys to every member and use the first 40 that are returned as your sample. ← No, nonresponse bias
b. Randomly pick a morning and survey the first 40 people who come in the door that day. ← No, convenience sample, undercoverage bias
c. Divide the number of members by 40 to get a value k. Choose one of the first kth names on the list using a random number generator. Then choose every kth name on the list after that name. ← No, systematic sample, not SRS because not all samples of 40 are equally likely to be selected.
d. Put each member's name on a slip of paper and randomly select 40 slips. ✓ very best way to sample
e. Get the sign-in lists for each day of the week, Monday through Friday. Randomly choose 8 names from each day for the survey. ← No, undercoverage from days not examined or those that don't sign in.

#18.

In a large population, 55% of the people get a physical examination at least once every two years. A SRS of 100 people are interviewed and the sample proportion is computed. The mean and standard deviation of the sampling distribution of the sample proportion are

a. 55, 4.97
b. .55, .002
c. 55, 2
d. .55, .0497
e. You cannot determine the standard deviation from the informa-

for sample proportions:

$\mu_{\hat{p}} = p \qquad \sigma_{\hat{p}} = \sqrt{\dfrac{pq}{n}}$

$\mu_{\hat{p}} = \boxed{0.55} \qquad \sigma_{\hat{p}} = \sqrt{\dfrac{(0.55)(0.45)}{100}}$

$= \boxed{0.0497}$

#19.

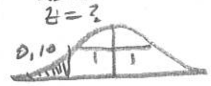
The weights of a large group of college football players is approximately normally distributed. It was determined that 10% of the players weigh less than 154 pounds and 5% weigh more than 213 pounds. What are the mean and standard deviation of the distribution of weights of football players?

a. 183.5, 19.44
b. 185.8, 22.36
c. 179.8, 20.17
d. 167.3, 18.66
e. 170.9, 19.85

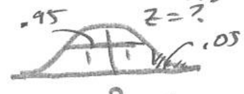
use $z = \frac{x-\mu}{\sigma}$

use 2 standardized distributions to find z-scores matched with data Xs and solve a system to find μ, σ!

x = 154 lbs
z = ?



x = 213 lbs
z = ?



$z = \text{invNorm}(.1, 0, 1) = -1.2816$

$-1.2816 = \frac{154-\mu}{\sigma}$

$z = \text{invNorm}(.95, 0, 1) = 1.6449$

$1.6449 = \frac{213-\mu}{\sigma}$

$-1.2816\sigma + \mu = 154$
$1.6449\sigma + \mu = 213$

$$\left[\begin{array}{cc|c} -1.2816 & 1 & 154 \\ 1.6449 & 1 & 213 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{cc|c} 1 & 0 & 20.1606 \\ 0 & 1 & 179.837 \end{array}\right] \begin{array}{l} \leftarrow \sigma \\ \leftarrow \mu \end{array}$$

#20.

An advice columnist asks readers to write in about how happy they are in their marriage. The results indicate that 79% of those responding would not marry the same partner if they had it to do all over again. Which of the following statements are true?

a. It's likely that this result is an accurate reflection of the population.
b. It's likely that this result is higher than the true population proportion because persons unhappy in their marriages are most likely to respond. ← an example of voluntary response bias
c. It's likely that this result is lower than the true population proportion because persons unhappy in their marriages are unlikely to respond.
d. It's likely that the results are not accurate because people tend to lie in voluntary response surveys.
e. There is really no way of predicting whether the results are biased or not.

#21.

A school survey of students concerning which band to hire for the next school dance shows 70% of students in favor of hiring The Greasy Slugs. What is the probability that, in a random sample of 200 students, at least 150 will favor hiring The Greasy Slugs?

a. $\binom{200}{150}(.7)^{150}(.3)^{50}$

b. $\binom{200}{150}(.3)^{150}(.7)^{50}$

c. $P\left(z > \frac{.75-.70}{\sqrt{\frac{.7(.3)}{200}}}\right)$

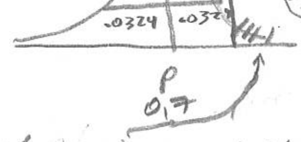
d. $P\left(z > \frac{.75-.70}{\sqrt{\frac{.7(.3)}{150}}}\right)$

e. $P\left(z > \frac{.70-.75}{\sqrt{\frac{.7(.3)}{200}}}\right)$

we would do the problem this way: this is a sampling distribution of sample proportion so:

$\mu_{\hat{p}} = p = 0.7 \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.7)(.3)}{200}} = .0324$

this $\hat{p} = \frac{150}{200} = .75$



$P(\hat{p} > .75) = \text{normalcdf}(.75, 999, 0.7, .0324) = \boxed{.0614}$
$\hat{p}$ (x), μ, σ

another way to write this is based on z-score formula: $z = \frac{x-\mu}{\sigma}$

here $z = \frac{.75-.70}{\sqrt{\frac{.7(.3)}{200}}}$

So you could write: $\boxed{P\left(z > \frac{.75-.70}{\sqrt{\frac{(.7)(.3)}{200}}}\right)}$

#22.

Which of the following describes an experiment but not an observational study?

a. A cause-and-effect relationship can be demonstrated.
b. The cost of conducting it is excessive.
c. More advanced statistics are needed for analysis after the data are gathered.
d. By law, the subjects must be informed that they are part of a study.
e. Possible confounding variables are more difficult to control.

experiments are the only thing which can establish a causal relationship

#23.

A least-squares regression line, $\hat{y} = a + bx$ is to be constructed for two variables x and y. As part of the process it is determined that $r = .77$, $\bar{x} = 3.5$, $s_x = .32$, $\bar{y} = 17.8$, and $s_y = 3.6$. What is the slope of the regression line?

a. 5.09
b. .068
c. 11.25
d. 8.66
e. 3.92

$b = r\frac{s_y}{s_x}$

$b = (.77)\frac{(3.6)}{(.32)}$

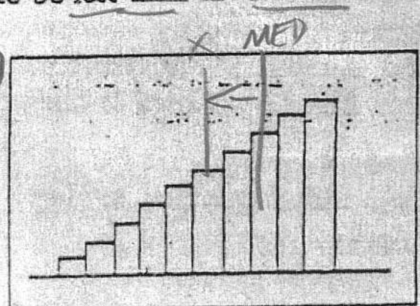
$b = \boxed{8.66}$

(if we need to intercept, a we could plug in the $(\bar{x}, \bar{y})$ point)

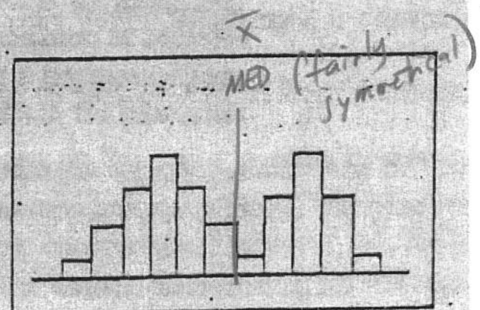
#24.

For which one of the following distributions is the mean most likely to be less than the median?

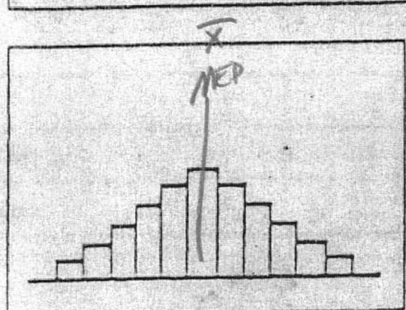
a.



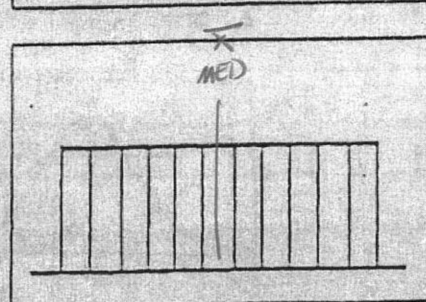
d.

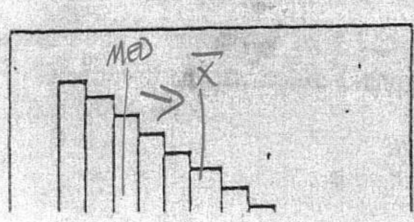


1.



a.





skew or outliers pull the mean away from the median toward the skew.

#25.

In an experiment, the purpose of randomization is to

a. equalize blocks in a block design
b. reduce variability by repeating the experiment on many subjects
c. control for variables not under study that might affect the response
d. control for common characteristics
e. make sure each subgroup is fairly represented

#26.

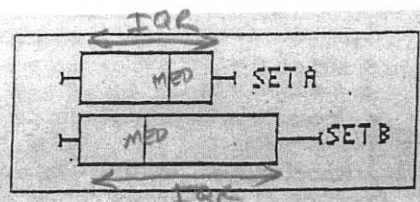
Which of the following statements is correct?

✓ I. The area under a probability density curve for a continuous random variable is 1. ← always true for any probability density function.
✓ II. A random variable is a numerical outcome of a random event. ← I would say 'the value of a random variable is a numerical outcome of an event' or 'a random variable assigns a numerical value to an event in a sample space' but this wording seems correct.
✓ III. The sum of the probabilities for a discrete random variable is 1.

this is saying the sum of all probabilities = 100%

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

#27.



Given the two boxplots above, which of the following statements is true?

X I. The boxplot for Set B has more terms above its median than the boxplot for Set A. ← doesn't make sense. Even if one was Set A the box size is related to distribution of values not the size of the dataset.
✓ II. The boxplot for Set B has a larger IQR than the boxplot for set A.
✓ III. The median for Set A is larger than the median for set B.

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

#28.

You are interested in determining which of two brands of tires (call them Brand G and Brand F) last longer under differing conditions of use. Fifty Toyota Camrys are fitted with Brand G tires and 50 Honda Accords are fitted with Brand F tires. Each tire is driven 20,000 miles, and tread wear is measured for each tire, and the average tread wear for the two brands is compared. What is wrong with this experimental design?

→ a. The type of car is a confounding variable.
b. Average tread wear is not a proper measure for comparison.
c. The experiment should have been conducted on more than two brands of cars.
d. Not enough of each type of tire was used in the study.
e. Nothing is wrong with this design—it will work quite well to compare the two bands of tires.

'Confounded variables' means an outcome could be produced by either variable but the experimental design doesn't allow for the effect of each variable to be determined separately. Would be better to have each tire brand with each type of car to see which is important (could be one or both variables).

#29.

The following are the probability distributions for two random variables, X, and Y:

X	$P(X=x)$	Y	$P(Y=y)$
3	$\frac{1}{3}$	1	$\frac{1}{8}$
5	$\frac{1}{2}$	3	$\frac{3}{8}$
7	$\frac{1}{6}$	4	? $\left(\frac{5}{16}\right)$
		5	$\frac{3}{16}$

If X and Y are independent, what is $P(X=5 \text{ and } Y=4)$?

a. 5/16
b. 13/16
c. 5/32
d. 3/32
e. 3/16

1st need to find missing ? probability, must sum to 1

$\frac{1}{8}+\frac{3}{8}+?+\frac{3}{16}=1$, $?=\frac{5}{16}$

$P(X=5 \text{ AND } Y=4)=P(X=5)\cdot P(X=4 \mid X=5)$

but if independent, then

$P(X=4 \mid X=5)=P(X=4)$

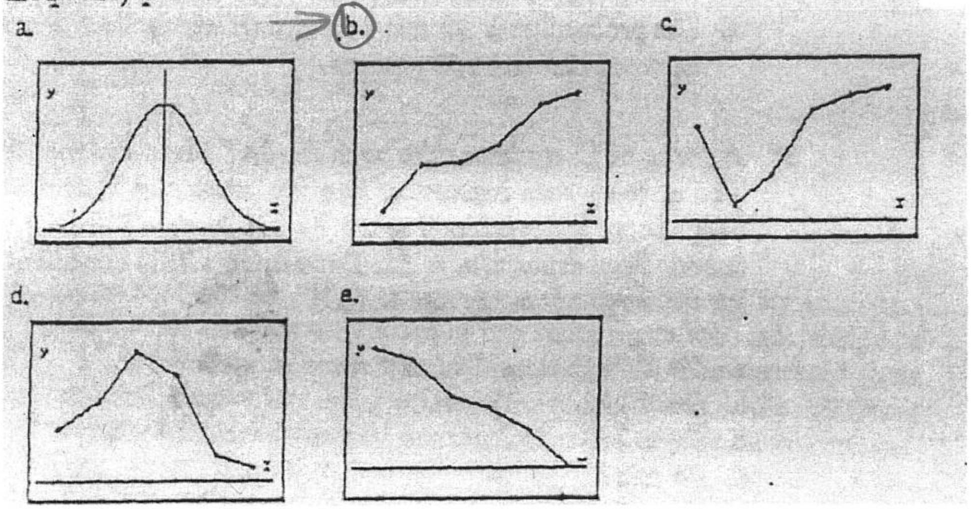
so

$P(X=5 \text{ AND } Y=4)=P(X=5)\cdot P(X=4)$

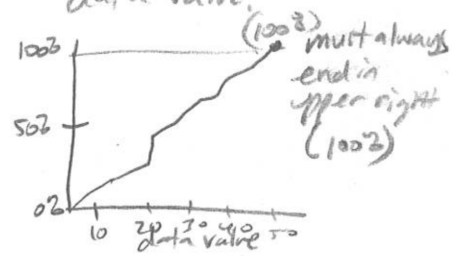
$=\left(\frac{1}{2}\right)\cdot\left(\frac{5}{16}\right)=\boxed{\frac{5}{32}}$

#30.

Which of the following graphs could be the graph of a cumulative frequency plot?



cumulative freq. plots show the cumulative percentage for each data value;



#31.

Consider the following set of data: {15, 17, 37, 45, 51, 52, 53, 56, 56, 57, 60, 63, 65, 67, 89}. Which of the following, using the 1.5(IQR rule) are outliers?

a. 89 only
b. 15 and 89 only
c. 15 only
d. 15, 17, and 89
e. 15 and 17 only

enter into L1 and use 1-var stats (scroll down for 5-number summary) (make sure FreqList is blank)

IQR = 63 − 45 = 18

MIN	Q1	MED	Q3	MAX
15 17 37	45	56	63	89

outliers: 15 17

1.5(18) = 27

lower fence 45 − 27 = 18

upper fence 63 + 27 = 90

#32.

At Midtown University, the average weight of freshmen boys is 170 lbs. with a standard deviation of 9 lbs. The average weight of freshmen girls is 115 lbs. with a standard deviation of 6 lbs. A new distribution is to be formed of the values obtained when the weights of the girls and the boys are added together. What are the mean and standard deviation of this new distribution? Assume that the weights of freshman boys and freshman girls are independent. ✓

a. 285, 15
b. 285, 117
c. 55, 10.8
d. 285, 10.8
e. The mean is 285 but, under the conditions stated in the problem, you cannot determine the standard deviation.

$\mu_B = 170$ $\mu_G = 115$

$\sigma_B = 9$ $\sigma_G = 6$

$Y = B + G$

so

$\mu_Y = \mu_G + \mu_B$

$\mu_Y = 170 + 115$

$\boxed{\mu_Y = 285}$

variances add

$\sigma_Y^2 = \sigma_B^2 + \sigma_G^2$

$\sigma_Y^2 = 9^2 + 6^2$

$\sigma_Y = \sqrt{81+36}$

$\boxed{\sigma_Y = 10.817}$

#33.

Fifty boys and 50 girls with ADHD (Attention Deficit Hyperactivity Disorder) were selected for an experiment to test a new drug for the treatment of ADHD. Half of the boys and half of the girls were selected at random to receive the new drug, and the other half of each group received a placebo. A reduction in symptoms of ADHD was measured for each subject. The basic design of this experiment is

a. completely randomized
b. completely randomized with one factor, gender
c. randomized block, blocked by drug and gender.
d. randomized block, blocked by gender
e. randomized block, blocked by drug

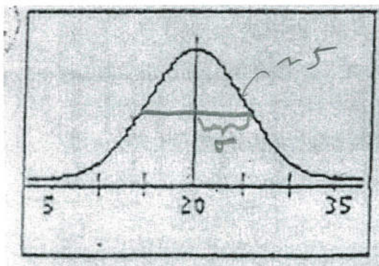
100 subjects

50 Girls
group 1 — drug 25 girls
group 2 — placebo 25 girls

50 boys
group 3 — drug 25 girls
group 4 — placebo 25 girls

this is how we know it was blocked by gender.
If it were completely randomized it could have uneven boy/girl mix in the treatment group.

#34.



Which of the following is the best estimate of the standard deviation for the approximately normal distribution pictured?

a. 10
b. 30
c. 5
d. 9
e. 15

#35.

You own an unusual die. Three faces are marked with the letter "X," two faces with the letter "Y," and one face with the letter "Z." What is the probability that at least one of the first two rolls is a "Y?"

a. 1/6
b. 2/3
c. 5/9
d. 1/3
e. 2/9

$P(Y) = \frac{2}{6} = \frac{1}{3}$

rolls are independent (Bernoulli setting)

X (# of rolls out of 2 are Y)	0	1	2

binompdf$(2, \frac{1}{3}, 0)$

$1 - $ binompdf$(2, \frac{1}{3}, 0)$

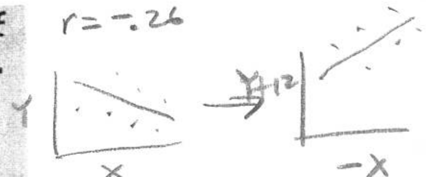
$.555\ldots = \boxed{\frac{5}{9}}$

#36.

The correlation between two variables X and Y is −.26. A new set of scores, X^* and Y^*, is constructed by letting $X^* = -X$ and $Y^* = Y + 12$. The correlation between X^* and Y^* is

a. −.26
b. .26
c. 0
d. .52
e. −.52

r = −.26



strength of association doesn't change but negating X reverses direction which mean slope and r must be negative of orig r.

r = −(−.26)

r = .26

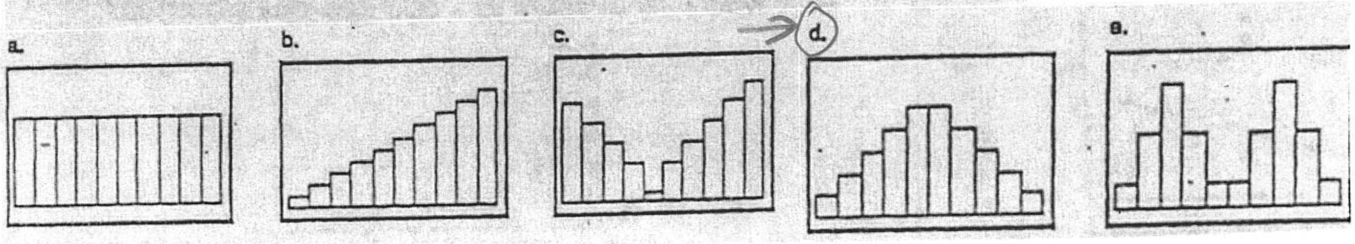
#37.

Which of the following best describes a "double-blind" study?.

a. The subjects are placed into treatment and control groups in such a manner that they do not know to which group they are assigned. — subjects blinded (single-blind)
b. The subjects are randomly assigned to treatment and control groups that controls for possible unknown biases that might be present in the study. — no blinding
c. Neither the subjects in the study nor the administrator of the study are aware of which subjects are in the treatment group and which are in the control group. — subjects & evaluators (administrators) both blinded
d. a technique for placing subjects in groups so as to protect against the placebo effect — grouping is not blinding (no blinding) placebo effect = what happens when you don't use a placebo
e. Volunteers are assigned to groups in such a way that they do not know into which groups the other volunteers have been placed. — single blind

#38.

Which of the five histograms pictured below has the smallest standard deviation?



small standard deviation when more values are clustered close to the mean

#39.

. One technique of drawing a sample for a survey is to select a stratified random sample. The purpose of this type of sample is to

a. Ensure that the sample is a simple random sample of the population. ← this would be an SRS.
b. Ensure that the sample is representative of the groups of interest in the population.
c. Economize by not having to sample from the entire population. ← this would be cluster sampling.
d. Control for lurking variables. ← done by randomizing (in any way).
e. Reduce the numbers you need to sample to arrive at valid conclusions. ← the samples needed for a particular confidence level don't change with sampling technique.

#40.

You roll two dice. What is the probability that the sum is six given that one die shows a 4?

a. 2/12
b. 2/36
c. 11/36
d. 2/11
e. 12/36

sum	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$P(6 \text{ sum} \mid \text{one is } 4) = \dfrac{\#\text{ of these with sum } 6}{\#\text{ways 1 is 4}} = \boxed{\dfrac{2}{11}}$

or:

$P(6 \cap 4) = P(4) \cdot P(6|4)$

$\frac{2}{36} = \frac{11}{36} \cdot P(6|4)$

$P(6|4) = \dfrac{2/36}{11/36} = \boxed{\dfrac{2}{11}}$

#41.

You play a game that involves rolling a die. You either win or lose \$1 depending on what number comes up on the die. If the number is even, you lose \$1, and if it is odd, you win \$1. However, the die is weighted and has the following probability distribution for the various faces:

Face	1	2	3	4	5	6
Win (x)	+1	−1	+1	−1	+1	−1
p(x)	.15	.20	.20	.25	.1	.1

Given that you win rather than lose, what is the probability that you rolled a "5?"

a. .50
b. .10
c. .45
d. .22

$P(5 \mid \text{win}) = \dfrac{.1}{.15 + .20 + .1} = \boxed{.22}$

#42.

A psychiatrist is studying the effects of regular exercise on stress reduction. She identifies 40 people who exercise regularly and 40 who do not. Each of the 80 people is given a questionnaire designed to determine stress levels. None of the 80 people who participated in the study knew that they were part of a study. Which of the following statements is true?

a. This is an observational study.
b. This is a randomized comparative experiment. ← no treatment, not an experiment.
c. This is a double-blind study. ← single-blind (participants)
d. This is a matched-pairs design. ← no treatment group / no 'pairing'
e. This is an experiment in which exercise level is a blocking variable. ← no treatment group, not an experiment.

#43.

Shanelle got the same score, 51, on two consecutive calculus quizzes. The mean for the class on the first quiz was 43 with a standard deviation of 6. The mean on the second quiz was 38 with a standard deviation of 9. Relative to the class, on which quiz did Shanelle do better?

a. She did better on quiz #1.
b. She did better on quiz #2.
c. She did equally well on each quiz because she got the same score.
d. She did better on quiz #2 because she was further above the mean than she was on quiz #1.
e. You cannot answer this question without knowing how many students actually took each quiz.

Compare z-scores:

$z_{1st\,quiz} = \frac{51-43}{6} = 1.33\sigma$ above mean

$z_{2nd\,quiz} = \frac{51-38}{9} = 1.44\sigma$ above mean

$\boxed{\text{better on 2nd quiz}}$

#44.

The distribution of a set of scores has mean of 35 and standard deviation of 12. Five is subtracted from each term in the distribution, and the result is multiplied by three. The new mean and standard deviation are

a. $\mu = 105, \sigma = 36$
b. $\mu = 90, \sigma = 12$
c. $\mu = 30, \sigma = 12$
d. $\mu = 90, \sigma = 36$

$Y = (X-5)3$

$\mu_Y = (\mu_X - 5)3$

$= (35-5)3 = \boxed{90}$

shifting doesn't affect measures of spread:

$\sigma_Y = (\sigma_X)3$

$= (12)3 = \boxed{36}$

#45.

Baxter is a 60% free-throw shooter who gets fouled during a game and gets to shoot what is called a "one-and-one" (that is, he gets to take a second shot—a bonus—if and only if he makes his first shot; each free throw, if made, is worth one point). Baxter can make 0 points (because he misses his first shot), 1 point (he makes the first shot, but misses the bonus), or 2 points (he makes his first shot and the bonus).

Assuming that each shot is independent, how many points is Baxter *most likely* to make in a one-and-one situation?

a. 2
b. 1
c. 0
d. .96
e. none of these is correct

X	0	1	2	
P	.4	.24	.36	(= 1.0)
	N	YN	YY	
	.4	(.6)(.4)	(.6)(.6)	

~~$p = .6$~~

~~$q = .4$~~

~~$EX = \mu = (0)(.4) + (1)(.24) + (2)(.36)$~~

~~$= .96$~~

wrong! this question is not asking for expected number of points it is asking how many points is Baxter most likely to make

$\boxed{\text{answer is 0}}$ because this has the highest probability

#46.

Baxter is a 60% free-throw shooter who gets fouled during a game and gets to shoot what is called a "one-and-one" (that is, he gets to take a second shot—a bonus—if and only if he makes his first shot; each free throw, if made, is worth one point). Baxter can make 0 points (because he misses his first shot), 1 point (he makes the first shot, but misses the bonus), or 2 points (he makes his first shot and the bonus).

Assuming that each shot is independent, how many points will Baxter make *on average* in a one-and-one situation?

a. 2
b. .96
c. 0
d. 1
e. .36

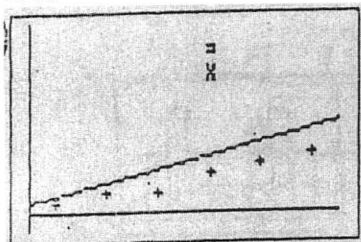
X	0	1	2
P	.4	.24	.36
	N	YN	YY
	(.4)	(.6)(.4)	(.6)(.6)

This one is expected value:

$EV = \mu = (0)(.4) + (1)(.24) + (2)(.36)$

$= \boxed{.96}$

#47.



For the graph given above, which of the following statements are true?

✓ I. The point marked with the "X" is better described as an outlier than as an influential point.
✓ II. Removing the point "X" would cause the correlation to increase.
✗ III. Removing the point "X" would have a significant effect on the slope of the regression line.

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

Yes, not influential because doesn't have much leverage (distance from mean horiz.)

yes, because this point is increasing average distance from LSRL.

no, because point is not influential

#48.

An electronics firm wants to survey its employees to determine their attitudes toward employee compensation. They obtain the sample for the survey by randomly selecting one of the first 20 names on an alphabetical list of employees and then select each 20th name on the list from then on. This is an example of which of the following?

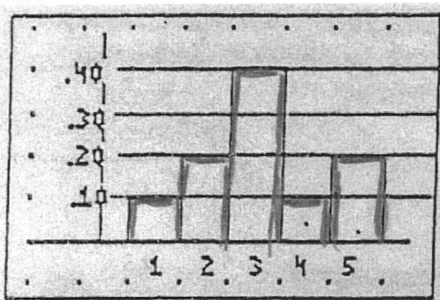
a. simple random sample
b. cluster sample
c. stratified random sample
d. convenience sample
e. systematic sample

#49.

Results of an experiment or survey are said to be *biased* if

a. Subjects are not assigned randomly to treatment and control groups.
b. Some outcomes are systematically favored over others.
c. There was no control group.
d. A double-blind procedure was not used.
e. The sample size was too small to control for sampling variability.

#50.



Given the probability histogram pictured for a discrete random variable X, what is μ_X?

a. 3.0
b. 25
c. 2.5
d. 3.1
e. 2.8

L1	L2
1	.1
2	.2
3	.4
4	.1
5	.2

1-Var Stats L1, L2

(don't forget to set FreqList: L2)

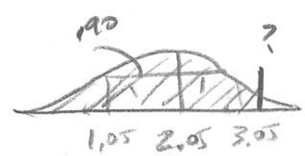
$\mu_X = \bar{x} = 3.1$

#51.

Free response questions on the AP Statistics Exam are graded on 4, 3, 2, 1, or 0 basis. Question #2 on the exam was of moderate difficulty. The average score on question #2 was 2.05 with a standard deviation of 1. To the nearest tenth, what score was achieved by a student who was at the 90th percentile of all students on the test? You may assume that the scores on the question were approximately normally distributed.

a. 3.5
b. 2.5
c. 2.9
d. 2.7
e. 3.3

$\mu = 2.05$
$\sigma = 1$



invNorm(0.90, 2.05, 1)

$= \boxed{3.33}$

#52.

40% of the staff in a local school district have master's degrees. One of the schools in the district has only 4 teachers out of 15. You are asked to design a simulation to determine the probability of getting this few teachers with master's degrees in a group this size. Which of the following assignments of the digits 0 through 9 would be appropriate for modeling this situation?

$\left(p = \frac{4}{15} = .266\right)$

modeling the district so 40% of the digits = masters

a. Assign "0,1,2" as having a master's degree and "4,5,6,7,8,9" as not having a degree. ← $\frac{3}{9}$ masters ≠ 40%
b. Assign "1,2,3,4,5" as having a master's degree and "0,6,7,8,9" as not having a degree. ← $\frac{5}{10}$ masters ≠ 40%
c. Assign "0,1" as having a master's degree and "2,3,4,5,6,7,8,9" as not having a degree ← $\frac{2}{10}$ masters ≠ 40%
→ d. Assign "0,1,2,3" as having a master's degree and "4,5,6,7,8,9" as not having a degree ← $\frac{4}{10}$ masters = 40% ✓
e. Assign "7,8,9" as having a master's degree and "0,1,2,3,4,5,6," as not having a degree. $\frac{3}{10}$ masters ≠ 40%

#53.

A study showed that persons who ate two carrots a day have significantly better eyesight than those who eat less than one carrot a week. Which of the following statements are correct?

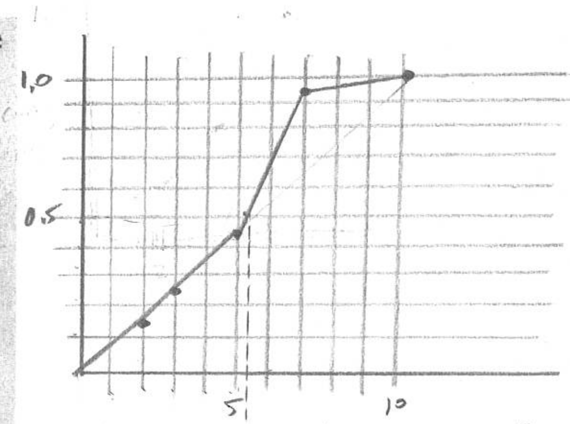
✗ I. This study provides evidence that eating carrots contributes to better eyesight. ← not an experiment, no causal relationship can be established.
✓ II. The general health consciousness of people who eat carrots could be a confounding variable. ← true, this is not controlled
✓ III. This is an observational study and not an experiment. ← true, no treatment is given.

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

#54.

Given the cumulative frequency table shown below, what are the mean and median of the distribution?

Value	Cumulative Frequency
2	.15
3	.25
5	.45
7	.95
10	1.00



This distribution doesn't get to 50% until after 5, so it is skewed left

a. Mean = 5.6, median = 7
b. Mean = 5.6, median = 5
c. Mean = 5.4, median = 7
d. Mean = 5.4, median = 5
→ e. Mean = 4.8, median = 6

so... mean is lower than median ←
and median must be >5
(but median can't be 7 because)