AP STATS TEST REVIEW TEST #1 Spring Semester

Chapters 8 and 9 (w/ Chapter 7 topics)

1. You wish to survey the people who have brought in their cars for service during the past month. You decide to pick a random sample of gas stations in the city and then **survey all customers from those stations** who had work done during the past month. This procedure is an example of which type of sampling?

   A. Cluster
   B. Convenience
   C. Simple random
   D. Stratified
   E. Systematic

Use the following scenario to answer questions 5 to 7: A psychologist studied the number of puzzles that subjects were able to solve in a five-minute period while listening to soothing music. Let *X* be the number of puzzles completed successfully by a random chosen subject. The following probability distribution for *X* was found:

X	1	2	3	4
P(X)	0.3	0.4	0.2	0.1

2) What is the probability that a randomly chosen subject completes **more than** the expected number $E(X) = \mu_X$ of puzzles in the five-minute period?

$$P\ (X > \mu_X) = P(X > 2.1) = P(X = 3) + P(X = 4)$$

$$P(X > 2.1) = 0.3$$

The probability that a randomly chosen subject completes more than the expected number of puzzles in the five-minute period is 0.3 or 30%

3) What are the values for σ_X and σ_X^2 or $(\sigma_X)^2$ for the random variable *X*?

$$\sigma_X = 0.9434 \quad \text{and} \quad (\sigma_X)^2 = 0.890$$

4) Let *D* be the difference in the number of puzzles solved by two randomly selected subjects in a five-minute period. What is the standard deviation of *D*?
   a) 2.3 b) 0.81 c) 1.27 **d) 1.334** e) 1.8

$$\text{If } D = X_1 - X_2, \text{ then } what\ is\ \sigma_D = ?$$

To find the result of the sum or difference of for more than one random variable, you must ***add the variances*** of each random variable and then take the square root.

$$\sigma_D = \sqrt{(\sigma_{X_1})^2 + (\sigma_{X_2})^2}$$

$$\sigma_D = \sqrt{(0.9434)^2 + (0.9434)^2}$$

$$\sigma_D = \sqrt{0.89 + 0.89} = 1.334$$

5) Suppose the outstanding loan for all college graduates forms a Normal distribution with $\mu = \$23{,}000$, and $\sigma = \$7{,}200$. If a college student is randomly selected, what is the probability that their mean outstanding loan is under \$21,000?

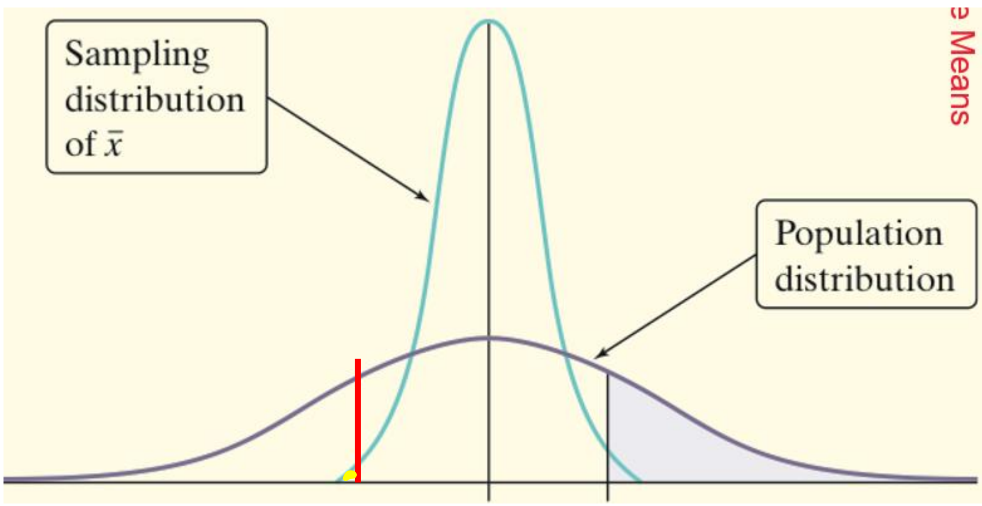
z-score = $= \dfrac{observ. - mean}{standard\ deviation} = \dfrac{x_i - \mu}{\sigma} = \dfrac{21000 - 23000}{7200} = -0.274$

Approx. 0.392 or 39.2% probability of randomly selecting a college student whose mean outstanding college debt is under \$21,0000 (debt $\leq \$21K$)

6) Suppose the average outstanding loan for college graduates is \$23,500 with a standard deviation of \$7,200. In an **SRS of 50** graduating college students, what is the probability that their mean outstanding loan is under \$21,000?

Since we have SRS $n = 50$, *then sampling distribution is approx. Normal*

If $n \geq 30, then\ CLT\ applies$



$$\mu_{\bar{X}} = \mu = \$23{,}500$$

$$\sigma = \$7200$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{7200}{\sqrt{50}} = 1018.23$$

$$P(Z < 21000) = normalcdf(0, 21000, 23500, 1018) = 0.0070$$

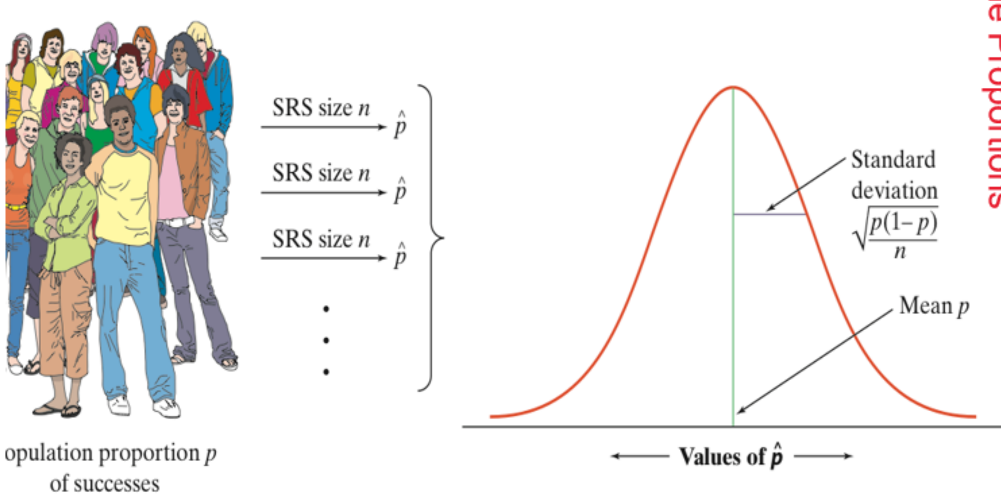
Therefore the probability of a SRS of 50 graduating college students having a mean outstanding loan under \$21,000 would be 0.007 or less than a 1% chance.

7) Suppose 54% of fourteen to nineteen year olds expect to have “a great life.” In an **SRS of 125** fourteen to nineteen year olds, what is the probability that *between 50% and 60%* will say they expect to have “a great life”?

The Sampling Distribution of $\hat{p}$

Sample Proportions

We can summarize the facts about the sampling distribution of $\hat{p}$ as follows:



$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.54)(0.46)}{125}} = 0.04458$$

Since $np = 125(0.54) = 67.5$ and $n(1 - p) = 125(0.46) = 57.5$ are both **greater than 10**, we can assume that the sampling distribution of sample proportion will be *approximately normal*, and can use standardized Normal (z-distribution) to find the desired probability

$$P(0.5 < \hat{p} < 0.6) = ?$$

$$z = \frac{0.5 - 0.54}{0.04458} \text{ and } z = \frac{0.6 - 0.54}{0.04458}$$

$$z = -0.8973 \quad \text{and} \quad z = 1.3459$$

There is a **72.6%** chance that this

8) The sampling distribution of the sample means is close to the normal distribution
   A. Only if the population is unimodal, not badly skewed, and does not have outliers.
   B. No matter what the distribution of the population or what the value of *n*.
   C. If *n* is large ($n \geq 30$), no matter ***any shape*** of the distribution in the population.
   D. If the standard deviation of the population is known.
   E. Only if both *n* is large and the population has a normal distribution.

9) What is the probability of rolling at least one six on three regular 6-sided dice?

$$P(at\ least\ 1\ six) = 1 - P(no\ sixes)$$

$$= 1 - \left(\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}\right)$$

$$= 1 - 0.5787 = \mathbf{0.4213}$$

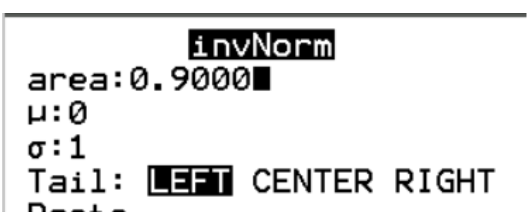
The probability of rolling at least one six on a three regular 6-sided dice is about 0.4213, or slightly more than 42%

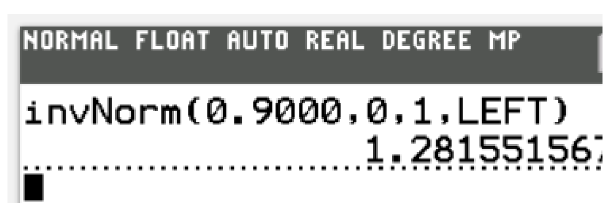
10) A child is 40 inches tall, which places her in the **90th percentile** of all children of similar age. The heights for children of this age form an approximately Normal distribution with a mean of 38 inches. Based on this information, what is the standard deviation of the heights of children this age?

$$z - score = \frac{obser - mean}{S.D.}$$

What is the *z-score* that corresponds to the 0.9000?

About 1.28





$$1.282 = \frac{40 - 38}{x}$$

$$\text{S. D. is} \rightarrow x = \frac{2}{1.282} = \mathbf{1.56}$$

11) In a congressional district, 55% of the registered voters are Democrats. Which of the following is equivalent to the probability of getting less than 50% Democrats in a *random sample of size 100*? (you must consider the S.D. of a sampling distribution)

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.55(0.45)}{100}}$$

$$Z - score = \frac{statistic - parameter}{S.D. of\ the\ statistic}$$

A. $P\left(Z < \frac{0.50-0.55}{100}\right)$

B. $P\left(Z < \frac{0.55-0.50}{\sqrt{100}}\right)$

C. $P\left(Z < \frac{0.50-0.55}{\sqrt{\frac{0.55(0.45)}{100}}}\right)$

D. $P\left(Z < \frac{0.55-0.50}{\sqrt{\frac{0.55(0.45)}{100}}}\right)$

E. Cannot be determined without standard deviation of population

AP Statistics 2020 Formulas and Tables Sheet

III. Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: statistic $\pm$ (critical value)(standard error of statistic)

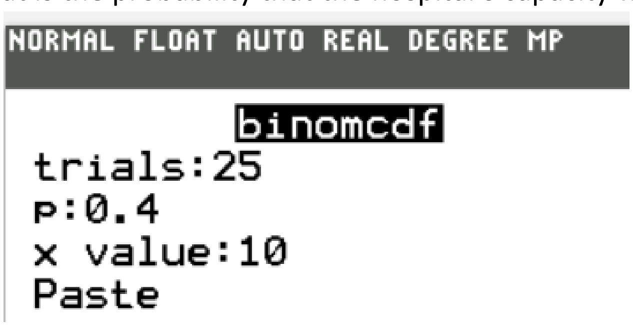
The *standard deviation of the statistic* is ________________________________

While the standard error of the statistic is an ______________________________

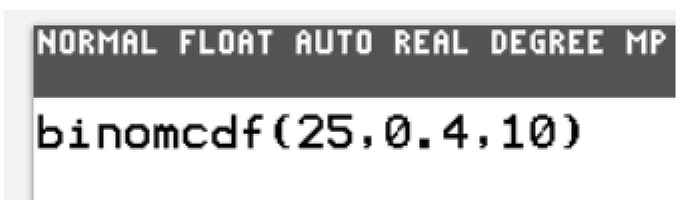
12) A study of voting chose **663 registered voters (sample)** at random shortly after an election. Of these, 72% (statistic) said they had voted in the election. Election records show that only 56% of registered voters *(parameter)* voted in the election. Which of the following states is true about this situation?

a) 72% is a sample, 56% is a population
b) 72% and 56% are both statistics
c) 72% is a statistic and 56% is a parameter
d) 72% is a parameter and 56% is a statistic
e) 72% and 56% are both parameters

13) Twenty-five people have been exposed to a particular disease. Each one independently has a 40% chance of contracting the disease. A hospital has the capacity to handle 10 cases of the disease. Binomial setting (B.I.N.S.)
What is the probability that the hospital's capacity will be exceeded?



OR



You should list the following on your paper for a Free Response question:
Define the random variable:
Let *X* = the number of people who contract the disease, having been exposed to it

$$P(hospital's\ capacity\ exceeded) = P(X > 10) or\ P(X \geq 11)$$

$$P(X > 10) = 1 - P(X \leq 10)$$

$$binomcdf\ (n = 25, p = 0.4, X \leq 10)$$

The probability that the hospital's capacity will be exceeded is **approximately 0.414** or a little greater than **41% chance.**

14) What is the name of a characteristic or variable of a sample?

A ***statistic***, usually a sample mean $(\overline{x})$,or a sample proportion $(\hat{p})$

15) A television network conducts a weekly survey to determine the proportion of viewers who watch various programs. For the coming year, decide to double the sample size. The main benefit of this is to:

a) Reduce undercoverage bias
b) Reduce nonresponse bias
c) Eliminate *sampling error*
d) Decrease the population variability
e) Decrease the standard deviation of the sampling distribution

16) Decreasing the sample size from 750 to 375 would multiply the standard deviation within a sampling distribution of $\bar{x}$ by:

Consider some known population S.D.: *Let* $\sigma = 10$

S.D. of a sampling distribution of sample means is $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

So compare what happens when we change the sample size *n*

$$\sigma_{\bar{X}} = \frac{10}{\sqrt{750}} = 0.3651 \quad \text{vs.} \quad \sigma_{\bar{X}} = \frac{10}{\sqrt{375}} = 0.5164$$

$$\text{OR} \quad \sigma_{\bar{X}} = \frac{10}{\sqrt{(0.5)750}} = \frac{1}{\sqrt{(0.5)}} \cdot \frac{10}{\sqrt{750}}$$

$$\text{OR} \quad \frac{1}{\sqrt{(0.5)}} = \frac{1}{\sqrt{\frac{1}{2}}} = \frac{1}{\frac{1}{\sqrt{2}}} = 1 \cdot \frac{\sqrt{2}}{1} = \sqrt{2}$$

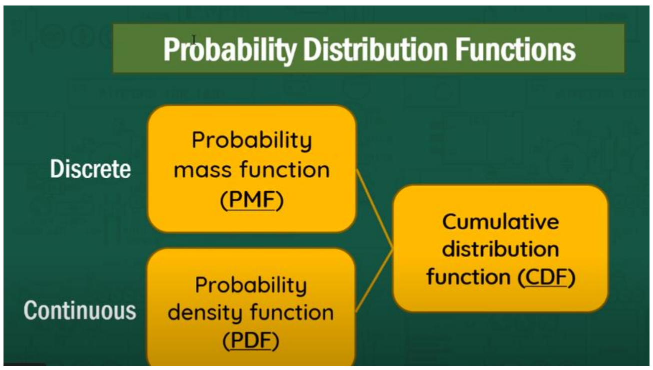
So, decreasing the sample size by HALF would multiply the standard deviation within a sampling distribution of $\bar{x}$ by $\sqrt{2}$

Typically, we want to decrease the standard deviation, so *by what factor* would you need to increase the sample size if you wanted to decrease the standard deviation of the statistic in half?

You will need to increase your sample size by a **factor of four** (four times larger, going from $n \text{ to } 4n$) to decrease the standard deviation to half of its original size.

17) During a bad economy, a graduating college student goes to career fair booths in the technology sector (e.g., Google, Apple, Qualcomm, Texas Instruments, Motorola, etc) - and his/her likelihood of receiving an off-campus interview invitation after a career fair booth visit depends on how well he/she did in AP statistics. Specifically, an **A** in AP statistics results in a probability $\boldsymbol{p} = \mathbf{0.95}$ of obtaining an invitation, whereas a C in the class results in a probability of $p = 0.15$ of an invitation.

(a) Give the *pmf* for the random variable *Y* that denotes **the number of career fair booth visits** a student must make **before his/her first invitation** including the visit that results in the invitation. Express your answer in terms of ***p*** *as a* ***p****robability function*.



Y = # of career booth visits until 1st *success* (*invitation for interview*)

$P\,(Y = x) = (1 - p)^{x-1}p$, for $x \geq 1$

(b) On average, how many booth visits must an A student make before getting an off-campus interview invitation? How about a C student? Y_A = "A" student, Y_C = "C" student

$$E(Y_A) = \frac{1}{p} = \frac{1}{0.95} = \mathbf{1.053}\ ;\ E(Y_C) = \frac{1}{p} = \frac{1}{0.15} = \mathbf{6.667}$$

(c) Assuming that each student visits 5 booths during a typical career fair, find the probability that an A student in stats *will not get* an off-campus interview invitation. Similarly, find the probability that a C student in stats will get an invitation during a typical career fair.

$$P(an\ A\ student\ will\ get\ an\ interview\ visiting\ up\ to\ 5\ booths) = P(Y_A \leq 5)$$

$$P(Y_A \leq 5) = P(Y_A = 1) + P(Y_A = 2) + \cdots P(Y_A = 5) = 0.999999688$$

$$P(an\ A\ student\ will\ \mathbf{not}\ get\ an\ interview\ if\ visit\ 5\ booths) = 1 - P(Y_A \leq 5)$$

Probability that an "A" student will **not** get an interview when they visit 5 career fair booths is approx. 3.125×10^{-7} or 0.000000312, or about 3 chances out of ten-million.

$$P(an\ C\ student\ will\ get\ an\ interview\ if\ visit\ 5\ booths) = P(Y_C \leq 5)$$

$$P(Y_C \leq 5) = P(Y_C = 1) + P(Y_C = 2) + \cdots P(Y_C = 5) =$$

$$P(Y_C \leq 5) = 0.15 + 0.85(.15) + 0.85^2(.15) + 0.85^3(.15) + 0.85^4(.15)$$

$$P(Y_C \leq 5) = 0.5563$$

Probability that an "C" student ***will get*** an interview when they visit 5 career fair booths is approximately 0.5563 or a little less than **56% chance**.

1. A statistic tends to produce an accurate estimate of a population characteristic when these aspects of the statistic are true: a) _unbiased estimator(s)___ b) _small S.D. about the statistic_____

2. Suppose a 95% confidence interval is computed for μ resulting in the interval (112.4, 121.6). Circle all of the following that are TRUE:

a. 95% of the population fall within the interval (112.4, 121.6)

b. There is a 95% chance that μ falls within the interval (112.4, 121.6)

c. 95% of all of the possible values of μ fall within the interval (112.4, 121.6)

d. The point estimate is $\bar{x} = 117$

e. 95% of all of the possible samples of the same size will produce intervals that capture the true μ

3. If σ = 10 for a normally distributed population, then the sample size required to estimate a population mean μ to within .5 with 95% confidence is what size?

We want a margin of error within 0.5, so $M.E. \leq 0.5$

$$M.E. = z^* \frac{\sigma}{\sqrt{n}}$$

$$1.96 \cdot \frac{10}{\sqrt{n}} \leq 0.5$$

So $\sqrt{n} \geq 39.2$, so $n \geq 1536.64$

$$\therefore n \geq \mathbf{1537}$$

Spring TEST #1 – Test **Review FRQ**

To increase morale among employees, a company began a program in which one employee is randomly selected each week to receive a gift card. Each of the company's 200 employees is equally likely to be selected each week, and the same employee could be selected more than once. Each week's selection is independent from every other week.

(a) Consider the probability that a particular employee receives at least one gift card in a 52-week year. $P(winning\ 1\ gift\ card) = ?$