Math 166 - Exam 2 Review

NOTE: For reviews of the other sections on Exam 2, refer to the first page of WIR #4 and #5.

Section 3.2 - Measures of Central Tendency

- **Population vs. Sample Statistics** - In statistics, a **sample** is a group of items (or people) chosen to represent a larger group. This "larger group" is called a population.

    - When you have data on every single object in a group, then any statistics calculated using ALL of the data values are called **population statistics**.

    - If statistics are calculated using only a subset or a sample of the data values in a group, then these statistics are called **sample statistics**.

        * To get a sample that is representative of the entire population, the sample must be chosen randomly.

        * As long as the sample is random, the statistics of the sample can be used to approximate the statistics of the population.

- **Average, or Mean** - The average, or mean, of the n numbers $x_1, x_2, \ldots, x_n$ is calculated as

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

and is represented as μ if these data values are from a population or $\bar{x}$ if these data values are from a sample.

- **Expected Value of a Random Variable** X - Let X denote a random variable that assumes the values $x_1, x_2, \ldots, x_n$ with associated probabilities $p_1, p_2, \ldots, p_n$, respectively. Then the *expected value* of X, written $E(X)$, is given by

$$E(x) = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n$$

- **Mean of a Binomial Random Variable** X - If X is a binomial random variable associated with a binomial experiment consisting of n independent trials with probability of success p, then the mean, or expected value, of X is $E(X) = np$.

- **Median** - The median is the middle value in a set of data arranged in increasing or decreasing order (when there is an odd number of entries). If there is an even number of entries, the median is the average of the two middle numbers.

- **Mode** - The mode is the value that occurs most frequently in the set of data. When a data set has two values that occur an equal number of times and this frequency is larger than the frequency of any other data value, then we say the data set is **bimodal**. If no data value occurs more frequently than any other, then we say that the data set has no mode.

- Histograms and Measures of Central Tendency

    - The expected value of a random variable X can be thought of as the center of balance of a histogram.

    - The median of a data set is the place where the area of the histogram is cut in half.

    - The mode of a data set corresponds to the tallest rectangle of the histogram.

Section 3.3 - Measures of Spread

- **Population Variance** - a measure of dispersion (or spread) that is calculated by finding the average of the squares of the deviations from the mean. Population variance is denoted by $Var(X)$ or σ^2, and the units of variance are the square of the units of the original data.

- **Population Standard Deviation** - a measure of dispersion (spread) that is calculated by taking the square root of the population variance. The units of standard deviation are the same as the units of the random variable, and population standard deviation is denoted by σ (on the calculator, σ_X).

- **Sample Variance** - The variance of a *sample* with n data values is calculated by dividing by $n-1$ instead of n when taking the average of the squares of the deviations from the mean. Sample variance is denoted by s^2.

- **Sample Standard Deviation** - The standard deviation of a *sample* is found by taking the square root of the sample variance. Sample standard deviation is denoted by s (on the calculator, S_X).

- **Variance and Standard Deviation of a Binomial Random Variable** X - If X is a binomial random variable associated with a binomial experiment consisting of n independent trials with probability of success p and probability of failure q, then the variance of X is $Var(X) = npq$, and the standard deviation of X is $\sigma = \sqrt{npq}$.

- **Chebychev's Inequality** - Let X be a random variable with expected value μ and standard deviation σ. Then the probability that a randomly chosen outcome of the experiment lies between $\mu - k\sigma$ and $\mu + k\sigma$ is at least $1 - \dfrac{1}{k^2}$; that is,

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Section 3.4 - The Normal Distribution

- Properties of the Normal Curve

  1. The normal curve is completely determined by μ and σ. (σ determines the sharpness or flatness of the curve.)
  2. The curve has a peak at $x = \mu$.
  3. The curve is symmetric with respect to the vertical line $x = \mu$.
  4. The curve always lies above the x-axis but approaches the x-axis as x extends indefinitely in either direction.
  5. The area under the curve and above the x-axis is 1.
  6. For any normal curve, 68.27% of the area under the curve lies within 1 standard deviation from the mean, 95.45% of the area lies within 2 standard deviations of the mean, and 99.73% of the area lies within 3 standard deviations of the mean.

- The *standard* normal random variable Z has mean 0 and standard deviation 1.

---

1. Five cards are chosen without replacement from a standard deck of 52 playing cards. In how many ways can this be done if

   (a) all five cards must be of the same suit?

   T_1 - Choose suit $n_1 = 4 = C(4,1)$

   T_2 - Choose 5 cards of that suit $n_2 = C(13,5)$

   $C(4,1)C(13,5) = \boxed{5,148}$

   (b) exactly four of the cards must be of the same rank?

   T_1 - Choose a rank $n_1 = 13 = C(13,1)$

   T_2 - Choose 4 of that rank $n_2 = C(4,4)$

   T_3 - Choose a card of a different rank $n_3 = C(48,1)$

   $C(13,1)C(4,4)C(48,1) = \boxed{624}$

2. A student takes a 10 question multiple choice exam, each question of which has 5 answer choices (1 correct, 4 incorrect). Being unprepared for the exam, the student randomly guesses at each question.

   (a) What is the probability that the student gets exactly 6 questions correct?

   $n = 10$
   "success" — correct

   $P(X = 6) = binompdf(10, \frac{1}{5}, 6)$

   $p = \frac{1}{5}$
   Indep✓
   $X = \#$ of successes

   $= \boxed{0.0055}$

   (b) What is the probability that the student gets at least 60% of the questions correct?

   $0.6(10) = 6$ correct

   $P(\text{at least 6 correct}) = 1 - binomcdf(10, \frac{1}{5}, 5)$

   $= \boxed{0.0064}$

   (c) What is the probability that the student gets the first 4 correct and the last 6 incorrect?

   $P(\text{CCCCWWWWWW})$ C - correct   W - wrong

   $= P(C)P(C)P(C)P(C)P(W)\cdots P(W)$
   $= (\frac{1}{5})^4 (\frac{4}{5})^6 = \boxed{0.0004 \text{ or } 4.1943 \times 10^{-4}}$

   (d) How many questions should the student expect to get correct?

   $E(X) = 10(\frac{1}{5}) = \boxed{2}$

   (e) Find the variance and standard deviation for the number of questions answered correctly.

   $Var(X) = npq = 10(\frac{1}{5})(\frac{4}{5}) = \boxed{1.6}$

   $\sigma = \sqrt{npq} = \sqrt{1.6} = \boxed{1.2649}$

3. How many different social security numbers are there if

   (a) there are no restrictions?

   $\underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} \cdot \underline{10} = 10^9 = 1,000,000,000$

   (b) only the numbers 2, 3, 5, 7, and 8 are used, the first two numbers are not the same, and the last digit is odd?

   $\underline{5} \cdot \underline{4} \cdot \underline{5} \cdot \underline{5} \cdot \underline{5} \cdot \underline{5} \cdot \underline{5} \cdot \underline{5} \cdot \underline{3} = 937,500$

   Dif.   ↑
                                                            odd

4. Classify each of the following random variables and give the possible values they each may assume.

 (a) X = the number of times a coin is flipped until tails appears.

$$X = 1, 2, 3, 4, \ldots$$
$$\text{Infinite Discrete}$$

 (b) X = the number of cards drawn <u>without</u> replacement from a standard deck of 52 playing cards until a red card is drawn.

$$X = 1, 2, 3, 4, \ldots, 26, 27 \quad \text{Finite Discrete}$$

 (c) X = the weight of a newborn baby.

$$X > 0 \quad \text{Continuous}$$

 (d) X = the number of hours my cat Mouse sleeps in one day.

$$0 \le X \le 24 \quad \text{Continuous}$$

 (e) X = the number cards drawn <u>with</u> replacement from a standard deck of 52 playing cards until a club is drawn.

$$X = 1, 2, 3, \ldots \quad \text{Infinite Discrete}$$

5. The following game costs $3 per play: A bag contains 2 gold coins and 28 silver coins. The player grabs two coins from the bag without replacement. A player wins $20 for getting the two gold coins and $5 for getting one gold and one silver coin. There is no prize for selecting two silver coins.

             $X = $ net winnings

 (a) Give the probability distribution for the net winnings of a person who plays this game once.

Outcome	Win	X
2 gold	20-3	17
1 gold 1 silver	5-3	2
2 silver	0-3	-3

Value of X	17	2	-3
$P(X = x)$	$\dfrac{1}{435}$	$\dfrac{56}{435}$	$\dfrac{126}{145}$

$$P(X=17) = \frac{C(2,2)}{C(30,2)} \qquad \frac{C(2,1)C(28,1)}{C(30,2)} \qquad \frac{C(28,2)}{C(30,2)}$$

 (b) Find the (expected) net winnings of a person playing this game once. Round to 2 decimal places.

$$E(X) = 17\left(\frac{1}{435}\right) + 2\left(\frac{56}{435}\right) + (-3)\left(\frac{126}{145}\right)$$

$$= \boxed{-\$2.31}$$

In the long run, a person can expect to lose an average of $2.31 per play.

6. Random Elementary did a survey of its 55 fourth graders to find out how many siblings each of them have. The following table summarizes this information:

Number of Siblings	0	1	2	3	4	5
Number of Students	10	21	15	3	5	1 ← Total: 55

(a) Define an appropriate random variable X for this data, and then find $P(X > 2)$.

$$X = \text{Number of Siblings} \qquad P(X>2) = \frac{3+5+1}{55} = \boxed{\frac{9}{55}}$$

(b) Find $P(1 \le X < 4)$.

$$P(1 \le X < 4) = \frac{21+15+3}{55} = \boxed{\frac{39}{55}}$$

(c) Find $E(X)$ and interpret its meaning.

$E(X) \leftarrow$ use $\bar{X}$ in 1-var stats L_1, L_2

$\boxed{E(X) = 1.5455}$ you can expect a randomly selected 4th grader to have 2 siblings.

(d) Compute the mean, median, mode, standard deviation, and variance for the data. Be sure to label all answers.

$\mu = $ mean $= 1.5455$ pop. standard deviation $= \sigma_x = 1.2183$

median $= 1$ population variance $= \sigma_x^2 = 1.4843$

mode $= 1$

(e) All of the students in the fifth grade class were surveyed in the same way, and it was found that mean number of siblings was 1.7845 with a standard deviation of 1.0043. Which data set has a greater amount of spread (or dispersion) about its mean?

4th graders   5th graders   compare σ's
$\sigma = 1.2183$ $\sigma = 1.0043$ or σ^2's .

Since the standard deviation is larger for the 4th graders, the data set for the 4th graders has more dispersion about its mean.

7. Bob has 7 markers, 8 pens, 3 pencils, and 5 black crayons. If writing instruments of the same type are identical, in how many ways can Bob arrange these items in a single row on his desk?

$$\frac{23!}{7! \, 8! \, 3! \, 5!}$$

8. Rosa has 3 different red dresses, 9 different black dresses, 7 different blue dresses, and 2 different green dresses.

(a) In how many ways can Rosa arrange these dresses in her closet if there are no restrictions?

$$21!$$ or $$P(21,21)$$ (these are equal)

(b) If Rosa randomly arranges these dresses in her closet, what is the probability that dresses of the same color are kept together? E - dresses of the same color are kept together.

$$P(E) = \frac{n(E)}{n(S)} = \frac{4! 3! 9! 7! 2!}{21!}$$

$$= 1.0310 \times 10^{-8}$$

$n(E)$: T_1 - arrange colors $n_1 = 4!$
T_2 - arrange red dresses $n_2 = 3!$
T_3 - arrange black dresses $n_3 = 9!$
T_4 - arrange blue dresses $n_4 = 7!$
T_5 - arrange green dresses $n_5 = 2!$

(c) Rosa has decided to pack 4 dresses for a trip. How many ways can she choose 4 dresses if

i. there are no restrictions?

$$C(21,4) = 5985$$

ii. at most 1 of the dresses is blue?

1 or less   Exactly 1 Blue or   0 Blue

$$C(7,1)C(14,3) + C(14,4) = 3549$$

iii. exactly 1 is blue and exactly 2 are red?

T_1 - choose 1 blue $n_1 = C(7,1)$
T_2 - choose 2 red $n_2 = C(3,2)$
T_3 - choose 1 nonblue, nonred $n_3 = C(11,1)$

$$C(7,1)C(3,2)C(11,1) = 231$$

(d) If Rosa randomly selects 8 dresses from her closet, what is the probability that

i. exactly 4 are blue or exactly 2 are green?
      B   U   G

$$P(B \cup G) = P(B) + P(G) - P(B \cap G)$$

$$= \frac{C(7,4)C(14,4)}{C(21,8)} + \frac{C(2,2)C(19,6)}{C(21,8)} - \frac{C(7,4)C(2,2)C(12,2)}{C(21,8)} = \frac{503}{1710}$$

ii. at least 2 are black?
   2 or more

E - 2 or more black
E^c - 1 or less black

$$P(E) = 1 - P(E^c) = 1 - \frac{C(9,1)C(12,7) + C(12,8)}{C(21,8)} = \frac{3109}{3230}$$

Exactly 1 or 0 black

9. Suppose you roll two fair 6-sided dice and take the sum of the numbers landing up. You will win twice what you paid if the sum is 7 or 11. You win nothing if the sum is 2, 3, or 12. For any other sum, you win $5. The game costs $10 to play. Let X denote the net winnings of someone who plays once. Find and interpret $E(X)$.

Outcome	Win	X
7 or 11	20 -10	10
2,3, or 12	0 -10	-10
all others	5 -10	-5

Value of X	10	-10	-5
$P(X=x)$	$\frac{8}{36}$	$\frac{4}{36}$	$\frac{24}{36}$

$1 - \frac{8}{36} - \frac{4}{36}$

6 outcomes add to 7
2 outcomes add to 11
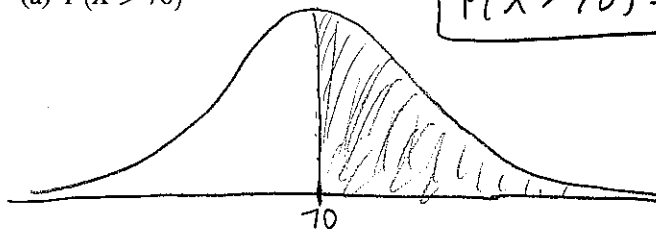
1 outcome adds to 2
2 outcomes add to 3
1 outcome adds to 12

$E(X) = 10\left(\frac{8}{36}\right) + (-10)\left(\frac{4}{36}\right) + (-5)\left(\frac{24}{36}\right)$

$\boxed{E(X) = -2.22}$ A person can expect to lose an average of $2.22 per play.
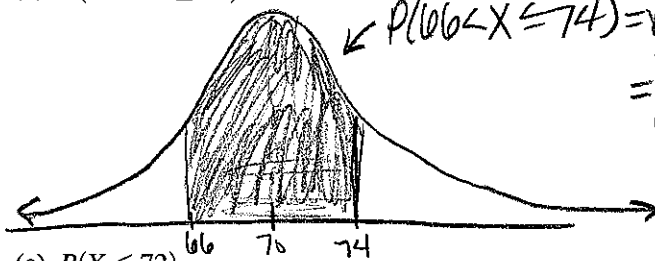
10. Let X be a normal random variable with $\mu = 70$ and $\sigma = 4$. By first sketching a normal curve and shading an appropriate area under the curve, find each of the following probabilities.
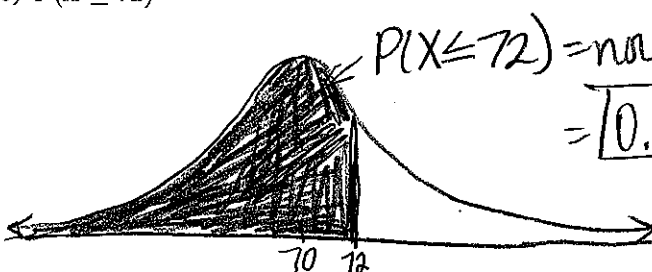
(a) $P(X > 70)$

$\boxed{P(X > 70) = 0.5}$

(b) $P(66 < X \leq 74)$

$P(66 < X \leq 74) = \text{normalcdf}(66, 74, 70, 4)$
$= \boxed{0.6827}$

(c) $P(X \leq 72)$

$P(X \leq 72) = \text{normalcdf}(-1E99, 72, 70, 4)$
$= \boxed{0.6915}$

11. The police department of a certain town estimates that 23% of all drivers in their town do (not) wear their seatbelts. If 60 cars are stopped at random, what is the probability that more than 90% of these drivers are wearing their seatbelts?

$n = 60$

"success" - wearing seatbelt

$p = 1 - 0.23 = 0.77$

Indep ✓

$X = \#$ of successes

90% of 60 = 54

$P(\text{more than } 54) = P(X \geq 55)$
$= 1 - \text{binomcdf}(60, 0.77, 54)$
$= \boxed{0.0028}$

12. A probability distribution has a mean of 100 and a standard deviation of 4. Use Chebychev's Theorem to estimate the probability that an outcome of the experiment lies

   (a) between 90 and 110.

$$\begin{array}{cc} 100 & 110 \\ -90 & -100 \\ \hline 10 & 10 \end{array}$$

$$10 = k \cdot 4$$
$$\frac{10}{4} = k$$
$$2.5 = k$$

$$P(90 \leq X \leq 110) \geq 1 - \frac{1}{(2.5)^2} = \boxed{0.84} \quad \overset{\text{at least}}{}$$

   (b) between 91.56 and 108.44.

$$\begin{array}{cc} 100 & 108.44 \\ -91.56 & -100 \\ \hline 8.44 & 8.44 \end{array}$$

$$8.44 = k \cdot 4$$
$$2.11 = k$$
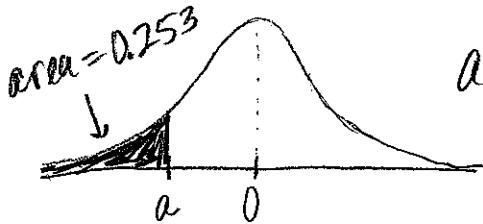
$$P(91.56 \leq X \leq 108.44) \geq 1 - \frac{1}{(2.11)^2} =$$

at least $\boxed{0.7754}$

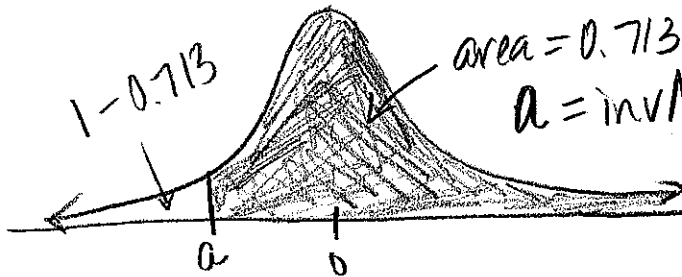13. Let Z be the standard normal random variable. Find the value of a if $\quad \mu = 0 \quad \sigma = 1$

   (a) $P(Z < a) = 0.253$

area $= 0.253$

$$a = \text{invnorm}(0.253)$$
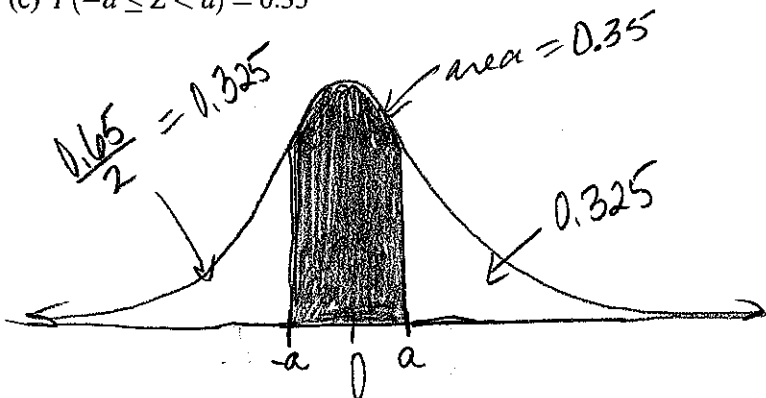$$\boxed{a = -0.6651}$$

   (b) $P(Z \geq a) = 0.713$

$1 - 0.713$

area $= 0.713$

$$a = \text{invNorm}(1 - 0.713) = \boxed{-0.5622}$$

must be area to the left of a.

   (c) $P(-a \leq Z < a) = 0.35$

$\frac{0.65}{2} = 0.325$

area $= 0.35$

0.325

$1 - 0.35 = 0.65 \leftarrow$ unshaded area

$$a = \text{invnorm}(1 - 0.325)$$
$$\boxed{a = 0.4538}$$

14. How many distinguishable arrangements are there of the letters in the word CONOCIMIENTO?

$$\frac{12!}{2! \cdot 3! \cdot 2! \cdot 2! \cdot 1! \cdot 1! \cdot 1!} = 9,979,200$$

15. At a certain hospital, the weights of babies at birth are normally distributed with a mean of 7.5 pounds and a standard deviation of 1.1 pounds.
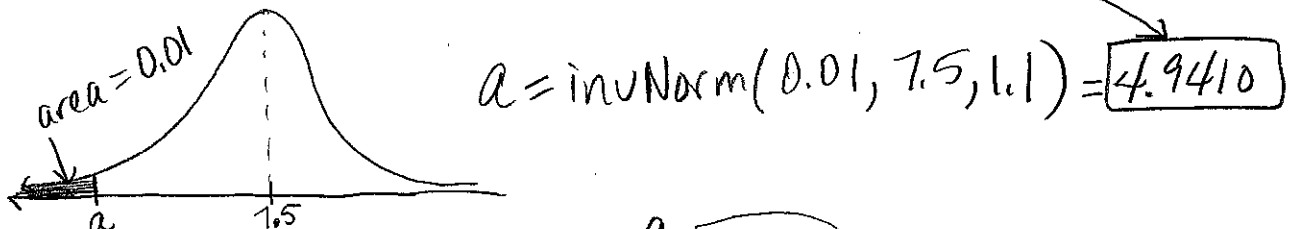
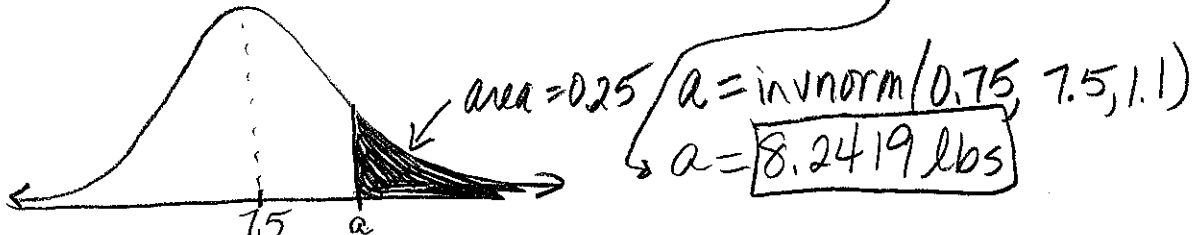(a) What is the probability that a randomly selected newborn at this hospital weighs more than 8 pounds?

$P(X > 8) = \text{normalcdf}(8, 1E99, 7.5, 1.1)$

$= \boxed{0.3247}$
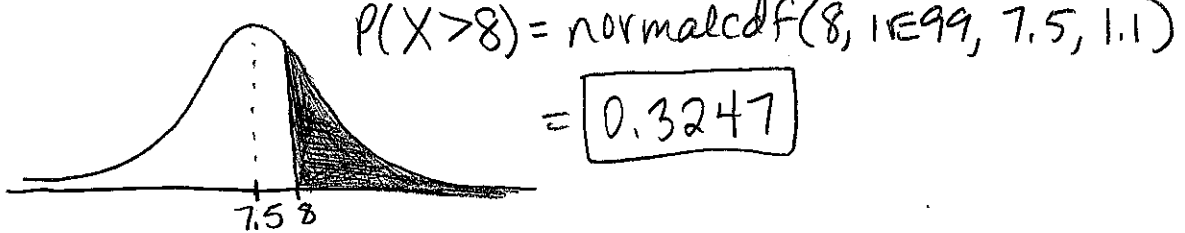
(b) What is the probability that a randomly selected newborn at this hospital weighs exactly 7.5 pounds?

$\boxed{P(X = 7.5) = 0}$

(c) Only 1% of all babies born at this hospital weigh less than a pounds.

area = 0.01

$a = \text{invNorm}(0.01, 7.5, 1.1) = \boxed{4.9410}$

(d) 25% of all babies born at this hospital weigh more than a pounds.

area = 0.25 $a = \text{invnorm}(0.75, 7.5, 1.1)$
$a = \boxed{8.2419 \text{ lbs}}$

(e) If you randomly access records of 1,000 newborns born at this hospital, how many of those babies would you expect to weigh more than 9 pounds at birth?

Find $P(X > 9)$:

$P(X > 9) = \text{normalcdf}(9, 1E99, 7.5, 1.1)$
0.0863
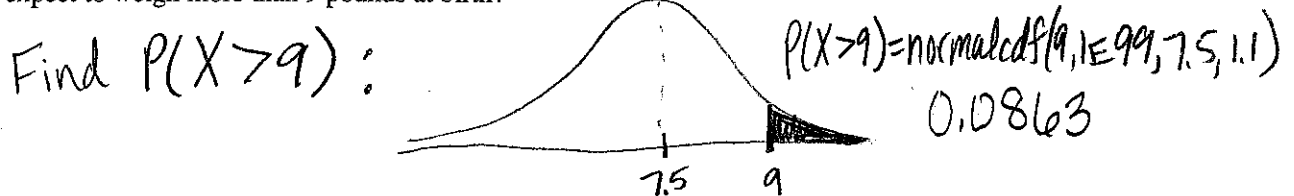
Find $E(x)$: $1000 * 0.0863 = 86.3$

you would expect 86 of the babies to weigh more than 9 lbs.

16. A physics instructor gave an exam to her class that had an average of 65 and standard deviation of 13. She decided to assign grades as follows: the top 6% and the bottom %6 will receive A's and F's, respectively. The next 16% in either direction will be given B's and D's, and the remaining students will receive C's. Assuming that the grades on the exam are normally distributed, find the cutoffs for each grade level.
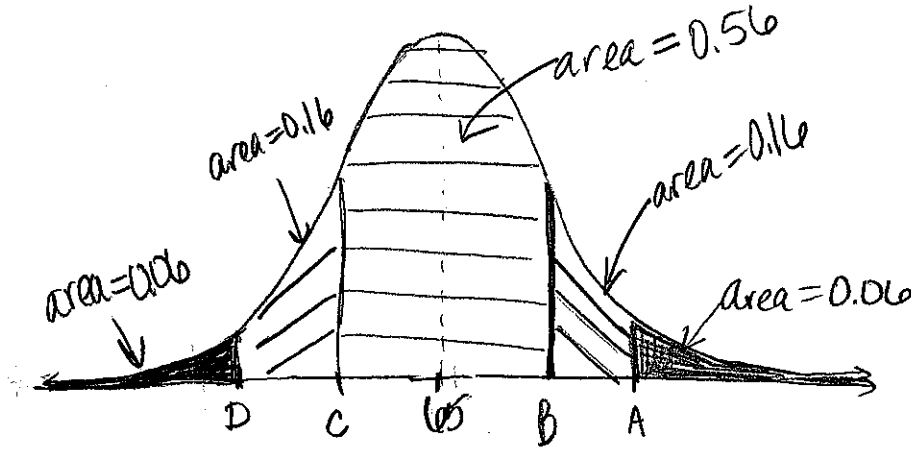
minimum for A: invnorm$(1-0.06, 65, 13) = 85.2121$

minimum for B: invnorm$(0.56+0.16+0.06, 65, 13) = 75.0385$

minimum for C: invnorm$(0.06+0.16, 65, 13) = 54.9615$

minimum for D: invnorm$(0.06, 65, 13) = 44.7879$

17. Katie has purchased a one-year insurance policy on her Stradivarius violin for $250. In the event of a minor damage to her violin, the insurance company will pay Katie $750 for damages. In the event that Katie's violin is damaged beyond repair or stolen, the insurance company will pay her $3,100. The insurance company estimates that the probability that Katie's violin will suffer minor damage in the next year is 0.005, and the probability that her violin will need to be replaced in the next year is 0.0001. What is the insurance company's (expected) gain from selling this policy?

$E(X), X = gain$

X = the insurance company's gain

Value of X	$P(X=x)$
minor damage $250-750$	0.005
replace $250-3100$	0.0001
Neither 250	0.9949

$E(X) = (-500)(0.005) + (-2850)(0.0001)$
$+ (250)(0.9949)$

$\boxed{E(X) = \$245.94}$

18. How many 7-symbol license plates with 3 letters and 4 digits are possible if

    (a) there are no restrictions?

$$\underline{10}\; \underline{(26)}\; \underline{10}\; \underline{10}\; \underline{(26)}\; \underline{10}\; \underline{(26)}$$
$$\;\;L\;\;L\;\;L$$

T_1 - Choose 3 places for letters $n_1 = C(7,3)$

T_2 - Pick letters $n_2 = 26^3$

T_3 - Pick digits $n_3 = 10^4$

$$\boxed{C(7,3)\cdot 26^3 \cdot 10^4} =$$
$$6,151,600,000$$

    (b) no symbol is repeated?

T_1 - choose 3 places for letters $n_1 = C(7,3)$

T_2 - Arrange letters $n_2 = 26\cdot 25\cdot 24$

T_3 - Arrange digits $n_3 = 10\cdot 9\cdot 8\cdot 7$

$$\underline{(26)}\; \underline{10}\; \underline{9}\; \underline{(25)}\; \underline{8}\; \underline{7}\; \underline{(24)}$$

$$\boxed{C(7,3)P(26,3)P(10,4)} =$$
$$2,751,849,000$$

    (c) letters must be kept together, digits must be kept together, and no digit is repeated?

letters first : $\underset{L}{\underline{26}}\cdot \underset{L}{\underline{26}}\; \underset{L}{\underline{26}}\cdot \underset{D}{\underline{10}}\cdot \underset{D}{\underline{9}}\cdot \underset{D}{\underline{8}}\cdot \underset{D}{\underline{7}}$

or $+$ $= \boxed{177,164,080}$

digits first : $\underset{D}{\underline{10}}\; \underset{D}{\underline{9}}\; \underset{D}{\underline{8}}\; \underset{D}{\underline{7}}\; \underset{L}{\underline{26}}\; \underset{L}{\underline{26}}\; \underset{L}{\underline{26}}$

19. The following histogram gives the probability distribution for a random variable X which takes on the values 0, 1, 2, 3, 4, and 5.

    (a) Find $P(X = 3)$.

$$P(X=3) = 1 - 0.2 - 0.15 - 0.3 - 0.1 - 0.05 = \boxed{0.2}$$

    (b) Find $E(X)$.

$$E(X) = 0(0.2) + 1(0.15) + 2(0.3) + 3(0.2) + 4(0.1) + 5(0.05)$$

$$\boxed{E(X) = 2}$$