Today is Monday, November 18th

Great Debate

STATS Monday

4:20pm MS Room 302

FRQ Test Tomorrow

Please Scroll to the End for Questions & Videos

Due Today:

- Death Packet

Assigned: Today, Nov 1st

CWK 4.2 FRQ Distributions

Assigned: November 15th

Assigned Today:

No Assignment Given

Due: November, Nov 18th

Due Tomorrow:

FRQ Test

Highly Recommended:

- You should be studying for Tuesday's FRQ Test by working all of the problems in the review videos. The Videos are hyperlinked in the review just click on the year and problem number. Scroll to the End

You are Welcome

Test Review: Probability Distributions

The Videos are hyper linked just click on the year and problem numbers

Normal Curves: 2004B Question 3 & 2006 Question 3 & 2014 Question 3 & 2010 Question 2

Understand the Empirical Rule and that the **greatest percent of the data is centered about the mean.**

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

Given actual values or Z-scores be able to calculate probabilities for the normal curve—**Normal CDF**
Given an area or a percent or a probability of a normal curve-**Inverse Norm**

- be able to calculate a Z-score and
- be able to use the calculated Z-score to find an X value, Mean or Standard Deviation.

Central Limit Theorem: 2007B Question 2 & 1998 Question 1

States that the sampling distribution of the expected values (means) of a population with a mean of μ and a standard deviation of σ will be approximately normal for any population regardless of the shape of the underlying population if the **sample is large enough**. **We say that the sample is large enough if the sample size is greater than or equal to 30. n ≥ 30**

Understand that Central Limit Theorem applies to the distribution of the **sample means**, distribution of the **sample averages** or **distribution of $\bar{x}$**. As the sample size increases the distribution of the $\bar{x}$'s becomes normal.

Understand that the distribution of individual values (the sample itself) follows the underlying distribution and that central limit theorem does not apply.

> **Note:** Increasing sample size decreases variability (the data moves closer to the mean). Changing sample size does not affect bias.

Distribution of Sample Means:

2004B Question 3 & 2006 Question 3 & 2014 Question 3 & 2010 Question 2

The distribution of the sample means is normal with $\mu = \bar{x}$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{or} \quad Z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \quad \text{where} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Be able to Calculate probabilities using a normal curve—Normal CDF

Geometric Distribution: 2011B Question 3 and 2016 Question 4

Has only 2 possible outcomes {success (p) and failure (q) q=1-p} Each trial is independent, the probabilities cannot change & the **number of trials not known**.

Equation: $q^{k-1}p$ **k is the number of trials until the 1st success Mean or E(X)** $\mu = \frac{1}{p}$

Or **P(First)** = $[P(A^C)]^{Power-1} \times P(A)$

- What is the probability that the **1st** success will occur on a given trial (Geometric PDF)
- What is the probability that the **1st** success will occur no later than or by (Geometric CDF)

Be able to recognize and calculate geometric probabilities and show work.
(Use the formula or blank method)

The Videos are hyper linked just click on the year and problem numbers

Binomial Distribution:

2011B Question 3 & 2006 Question 3 & 2016 Question 4 & 2014 Question 3 & 2007B Question 2

Has only 2 possible outcomes {**success (p) and failure (q) q=1-p**} Each trial is independent, the probabilities cannot change & the **number of trials is pre-determined/fixed**.

Equation: $\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$ **Mean or E(X)=np; Standard deviation** $\sigma = \sqrt{npq}$

- **n is the number of trials**
- **k is the number of successes**

You must show your work
You need to be able to recognize a graph created by using the binomial.
Remember: A Binomial CDF is just the sum of all of the PDF's

Extreme Values: 2004B Question 3

If you calculate a probability less than 5%, you can conclude that the claimed mean or claimed proportion is incorrect. Anything less than 5% we will consider unusual.

Combining Random Variables. 2008B Question 5 **This problem is super important**

You will need to remember that means can be added or subtracted. Standard deviations must be converted to variances and must be summed. Take the square root of the resulting sum.

Determining whether a distribution is Normal: 2011 Question 1

Inverse norm: 2008B Question 5 & 2010 Question 2

Contingency Table 2017 Question 3

You will need to be able to create a table and use it to calculate values

How do I determine which formula I should be using?

Geometric:

I would begin by looking for the word first. If first is included in the question, then we are probably working with a geometric distribution. Now we need to determine whether or not the distribution is a CDF or a PDF. If the first success occurs on a specific attempt then it is a PDF for example:

Example geometric pdf: What is the probability that a basketball player who makes 82% shots, misses their first basket on their fourth shot?

The fact that the first miss must occur on a singular value in this case the fourth shot, makes this a geometric pdf.

The work shown for this problem would be:

$$\frac{.82}{\text{Basket}} \times \frac{.82}{\text{Basket}} \times \frac{.82}{\text{Basket}} \times \frac{.18}{\text{Miss}} \quad \text{or } (.82)^3 \times (.18)^1 = .0992$$

Example geometric cdf: What is the probability that a basketball player who makes 82% shots, first miss occurs during their first four shots?

The fact that the first miss can occur on any of the four shots makes this a cdf.

The work shown for this problem would be:

$$(.82)^0 \times (.18)^1 + (.82)^1 \times (.18)^1 + (.82)^2 \times (.18)^1 + \ . \ . \ . + (.82)^3 \times (.18)^1 = .5479$$

Binomials:

While a geometric will continue until the first success a binomial has a fixed number of attempts or trials and there may be no successes.

Binomials are usually phrased as:

- What is the probability of some number of successes in a given number of trials? (Binomial PDF)
- What is the probability of at least some number of successes in a given number of trials? (Binomial CDF) usually **1-**Binomial CDF
- What is the probability of no more than some number of successes in a given number of trials? (Binomial CDF)
- What is the probability that the number of successes in a given number of trials are between 2 values? (Binomial CDF)--**(Binomial CDF of larger value) - (Binomial CDF smaller value)**

Example binomial pdf: What is the probability that a basketball player who makes 82% shots, shoots 25 times and misses 6?

The fact that there are a fixed number of trials (25) and a fixed probability (.18) lets me know that I am working with a binomial. The fact that I have an exact number of misses (6) lets me know that I am working with a binomial PDF.

0 1 2 3 4 5 6 7 8 ….25

Work shown: $\binom{25}{6}(.18)^6(.82)^{19} = .1388$

Example binomial cdf: What is the probability that a basketball player who makes 82% shots, shoots 25 times misses no more than 6?

The fact that there are a fixed number of trials (25) and a fixed probability (.18) lets me know that I am working with a binomial. The fact that the shooter doesn't have to miss a specific number of shots but can miss either 0, or 1 or 2 or 3 or 4 or 5 or 6 shots lets me know that I am working with a binomial CDF.

0 1 2 3 4 5 6 7 8 25

Work shown: $\binom{25}{0}(.18)^0(.82)^{25} + \ldots + \binom{25}{6}(.18)^6(.82)^{19}$ = .8512

Example binomial cdf: What is the probability that a basketball player who makes 82% shots, shoots 25 times and misses at least 6?

0 1 2 3 4 5 6 7 8 25

Work shown: $\binom{25}{6}(.18)^6(.82)^{19} + \ldots + \binom{25}{25}(.18)^{25}(.82)^0$ = .8512

Or

$1 - [\binom{25}{0}(.18)^0(.82)^{25} + \ldots + \binom{25}{5}(.18)^5(.82)^{20}]$ = .8512

For both cases the calculator is 1-Binomial CDF of X -s 5

Remember for at least problems subtract out the highest number you don't want. We wanted 6, so we subtracted the CDF of 5

Example binomial cdf: What is the probability that a basketball player who makes 82% shots, shoots 25 times and misses between 6 and 10 shots inclusive?

0 1 2 3 4 5 6 7 8 9 10 25

Work shown: $\binom{25}{6}(.18)^6(.82)^{19} + \ldots + \binom{25}{10}(.18)^{10}(.82)^{15}$ = .2852

The calculator is: Binomial CDF of X is 10 – Binomial CDF of X is 5

Normal Distribution: Look for the word normally distributed. If included we can expect to use the normal and should begin by writing the equation:

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

Normal CDF: If you are asked to find a percent, probability or area and a mean and standard deviation are given, you will be using a Normal CDF. Draw the graph and shade them plug in the values and calculate.

Example Normal CDF: The scores on an IQ test are normally distributed with a mean of 100 and a standard deviation of 15. What is the percentile rank of an individual with an IQ below 95?

Work shown: $P(Z < \frac{95-100}{15})$ = .3694

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

Inverse Norm: If you are asked to find a value or a z-score and a probability or percent is given along with a mean and standard deviation, you will be using an Inverse Norm. Draw the graph and shade and calculate your z-score using inverse norm with a mean of zero and a standard deviation of one. Now plug the z-score into the equation and then run an inverse norm using the actual mean and standard deviation to calculate the value of interest.

Example Inverse Norm: The scores on an IQ test are normally distributed with a mean of 100 and a standard deviation of 15. What is the IQ score of someone who is at the 83^{rd} percentile?

Work shown: **$P(Z < \frac{X-100}{15}) = .83$**

Calculator: 2^{nd} VARS Inverse norm for area = .83 mean of 0 & standard deviation of 1 → **Z =. 95**

$$(.95 < \frac{X-100}{15})$$ you may solve algebraically

Calculator: 2^{nd} VARS Inverse norm for area = .83 mean of 100 & standard deviation of 15 → **X = 114.3**

Normal by Central Limit Theorem: Look for the word normally distributed and the phrases sampling distribution of $\bar{x}$ or sampling distribution of the means or averages.

If included we can expect to use $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

If we see the sampling distribution of $\bar{x}$ or sampling distribution of the means or averages, but the population is unknown or is something other than normal we can still use $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ if the sample size (n) is greater than or equal to 30

$= \bar{x}$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{or} \quad Z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \quad \text{where} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Example Normal CDF for CLT: The scores on an IQ test are normally distributed with a mean of 100 and a standard deviation of 15. What is the probability that a group of 7 randomly selected adults have an average IQ below 95?

Work shown: **$P(Z = \frac{95-100}{\frac{15}{\sqrt{7}}}) = .1889$**

Hyperlinked Video Questions

1998 Question 1 Consider the sampling distribution of a sample mean obtained by a random sampling from an infinite population. This population has a distribution that is highly skewed toward the large values.

(a) How is the mean of the sampling distribution related to the mean of the population?
(b) How is the standard deviation of the sampling distribution related to the standard deviation of the population?
(c) How is the shape of the sampling distribution affected by the sample size?

2004B Question 3 Trains carry bauxite ore from a mine in Canada to an aluminum process plant in northern New York state in hopper cars. Filling equipment is used to load ore into the hopper cars. When functioning properly, the actual weights of ore loaded into each car by the filling equipment at the mine are approximately normally distributed with a mean of 70 tons and a standard deviation of 0.9 ton. If the mean is greater than 70 tons, the load mechanism is overfilling.

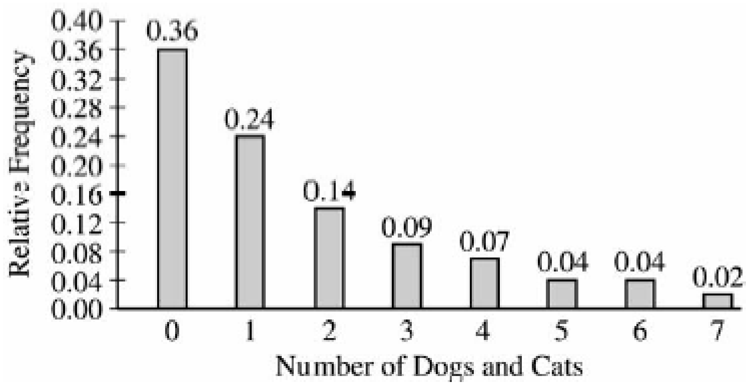
(a) If the filling equipment is functioning properly, what is the probability that the weight of the ore in a randomly selected car will be 70.7 tons or more? Show your work.
(b) Suppose that the weight of the ore in a randomly selected car is 70.7 tons. Would that fact make you suspect that the loading mechanism is overfilling the cars? Justify your answer.
(c) If the filling equipment is functioning properly, what is the probability that a random sample of 10 cars will have a mean ore weight of 70.7 tons or more? Show your work.
(d) Based on your answer in part (c), if a random sample of 10 cars had a mean ore weight of 70.7 tons, would you suspect that the loading mechanism was overfilling the cars? Justify your answer.

2006 Question 3 The depth from the surface of Earth to a refracting layer beneath the surface can be estimated using methods developed by seismologists. One method is based on the time required for vibrations to travel from a distance explosion to a receiving point. The depth measurement (M) is the sum of the true depth (D) and the random measurement error (E). This is, M = D + E. The measurement (E) is assumed to be normally distributed with mean 0 feet and standard deviation 1.5 feet.

(a) If the true depth at a certain point is 2 feet, what is the probability that the depth measurement will be negative?
(b) Suppose three independent depth measurements are taken at the point where the true depth is 2 feet. What is the probability that <u>at least one</u> of these measurements will be negative?

What is the probability that the <u>mean</u> of the three independent depth measurements taken at the point where the true depth is 2 feet will be negative?

2007B Question 2 The graph below display the relative frequency distribution for X, the total number of dogs and cats owned per household, for the households in a large suburban area. For instance, 14 percent of the households own 2 of these pets.

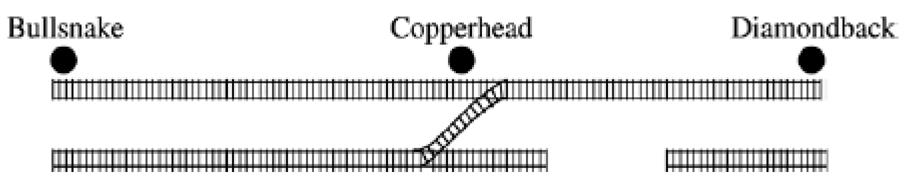


(a) According to a local law, each household in this area is prohibited from owning more than 3 of these pets. If a household in this area is selected at random, what is the probability that the selected household will be in violation of this law? Show your work.

(b) If 10 households in this area are selected at random, what is the probability that exactly 2 of them will be in violation of this law? Show your work.

(c) The mean and standard deviation of X are 1.65 and 1.851, respectively. Suppose 150 households in this area are to be selected at random and $\bar{X}$, the mean number of dogs and cats per household, is to be computed. Describe the sampling distribution of $\bar{X}$, including its shape, center, and spread.

2008B Question 5 Flooding has washing out one of the tracks of the Snake Gulch Railroad. The railroad has two parallel tracks from Bullsnake to Copperhead, but only one usable track from Copperhead to Diamondback, as shown in the figure below. Having only one usable track disrupts the usual schedule. Until it is repaired, the washed-out track will remain unusable. If the train leaving Bullsnake arrives at Copperhead first, it has to wait until the train leaving Diamondback arrives at Copperhead.



Every day at noon a train leaves Bullsnake heading for Diamondback and another leaves Diamonback heading for Bullsnake.

Assume that the length of time, X, it takes the train leaving Bullsnake to get to Copperhead is normally distributed with a mean of 170 minutes and a standard deviation of 20 minutes.
Assume that the length of time, Y, it takes the train leaving Diamondback to get to Copperhead is normally distributed with a mean of 200 minutes and a standard deviation of 10 minutes. These two travel times are independent.

(a) What is the distribution of Y – X?

(b) Over the long run, what proportion of the days will the train from Bullsnake have to wait at Copperhead for the train from Diamondback to arrive?

(c) How long should the Snake Gulch Railroad delay the departure of the train from Bullsnake so that the probability that it has to wait is only 0.01?

2010 Question 2: A local radio station plays 40 rock-and-roll songs during each 4-hour show. The program director at the station needs to know the total amount of airtime for the 40 songs so that time can also be programmed during the show for news and advertisements. The distribution of the lengths of rock-and-roll songs, in minutes, is roughly symmetric with a mean length of 3.9 minutes and a standard deviation of 1.1 minutes.

a) Describe the sampling distribution of the sample mean song lengths for random samples of 40 rock-and-roll songs.

b) If the program manager schedules 80 minutes of news and advertisements for the 4-hour (240 minute) show, only 160 minutes are available for music. Approximately what is the probability that the total amount of time needed to play 40 randomly selected rock-and-roll songs exceeds the available airtime?

2011 Question 1 A professional sports team evaluates potential players for a certain position based on two main characteristics, speed and strength.

(a) Speed is measured by the time required to run a distance of 40 yards, with smaller times indicating more desirable (faster) speeds. From previous speed data for all players in this position, the times to run 40 yards have a mean of 4.60 seconds and a standard deviation of 0.15 seconds, with a minimum time of 4.40 seconds, as shown in the table below.

	Mean	Standard deviation	Minimum
Time to run 40 yards	4.60 seconds	0.15 seconds	4.40 seconds

Based on the relationship between the mean, standard deviation, and minimum time, is it reasonable to believe that the distribution of 40-yard running times is approximately normal? Explain.

2011B Question 3 An airline claims that there is a 0.10 probability that a coach-class ticket holder who flies frequently will be upgraded to first class on any flight. This outcome is independent from flight to flight. Sam is a frequent flier who always purchases coach-class tickets.

(a) What is the probability that Sam's first upgrade will occur after the third flight?

(b) What is the probability that Sam will be upgraded exactly 2 times in his next 20 flights?

(c) Sam will take 104 flights next year. Would you be surprised if Sam receives more than 20 upgrades to first class during the year? Justify your answer.

2014 Question 3 Schools in a certain state receive funding based on the number of students who attend the school. To determine the number of students who attend a school, one school day is selected at random and the number of students in attendance that day is counted and used for funding purposes. The daily number of absences at High School A in the state is approximately normally distributed with mean of 120 students and standard deviation of 10.5 students.

(a) If more than 140 students are absent on the day the attendance count is taken for funding purposes, the school will lose some of its state funding in the subsequent year. Approximately what is the probability that High School A will lose some state funding?

(b) The principals' association in the state suggests that instead of choosing one day at random, the state should choose 3 days at random. With the suggested plan, High School A would lose some of its state funding in the subsequent year if the mean number of students absent for the 3 days is greater than 140. Would High School A be more likely, less likely, or equally likely to lose funding using the suggested plan compared to the plan described in part (a)? Justify your choice.

(c) A typical school week consists of the days Monday, Tuesday, Wednesday, Thursday, and Friday. The principal at High School A believes that the number of absences tends to be greater on Mondays and Fridays, and there is concern that the school will lose state funding if the attendance count occurs on a Monday or Friday. If one school day is chosen at random from each of 3 typical school weeks, what is the probability that none of the 3 days chosen is a Tuesday, Wednesday, or Thursday?

(d) The school board has decided to randomly select one day from each of five weeks. What is the likelihood that more than 2 of the days selected are either a Monday or Friday?

2016 Question 4 A Company manufactures model rockets that require igniters to launch. Once an igniter is used to launch a rocket, the igniter cannot be reused. Sometimes an igniter fails to operate correctly, and the rocket does not launch. The company estimates that the overall failure rate, defined as the percent of all igniters that fail to operate correctly, is 15 percent.
A company engineer develop a new igniter, called the super igniter, with the intent of lowering the failure rate. To test the performance of the super igniters, the engineer uses the following process.

Step 1: One super igniter is selected at random and used in a rocket.

Step 2: If the rocket launches, another super igniter is selected at random and used in a rocket.

Step 2 is repeated until the process stops. The process stops when a super igniter fails to operate correctly or 32 super igniters have successfully launched rockets, whichever comes first. Assume that super igniter failures are independent.

(a) If the failure rate of the super igniters is 15 percent, what is the probability that the first 30 super igniters selected using the testing process successfully launch rockets?
(b) Given that the first 30 super igniters successfully launch rockets, what is the probability that the first failure occurs on the thirty-first of thirty-second super igniter tested if the failure rate of the super igniters is 15 percent.
(c) Given that the first 30 super igniters successfully launch rockets, is it reasonable to believe that the failure rate of the super igniters is less than 15 percent? Explain.
(d) Assume the failure rate for the rocket igniters is 15% and the company decides to fire the rocket 30 times and then count the number of failures. What is the likelihood that there are 6 to 15 failures inclusive?

2017 Question 3 A grocery store purchases melons from only two distributors, J & K. Distributor J provides melons from organic farms. The distribution of diameters of the melons from Distributor J is approximately normal with mean 133 millimeters (mm) and standard deviation 5 mm.

(a) For a melon selected at random from Distributor J, what is the probability that the melon will have a diameter greater than 137 mm?

Distributor K provides melons from nonorganic farms. The probability is 0.8413 that a melon selected at random from Distributor K will have a diameter greater than 137mm. For all the melons at the grocery store, 70 percent of the melons are provided by Distributor J and 30 percent are provided by Distributor K.

(b) For a melon selected at random from the grocery store, what is the probability that the melon will have a diameter greater than 137mm?

(c) Given that a melon selected at random from the grocery store has a diameter greater than 137mm, what is the probability that the melon will be from distributor J?